7

Summary

Kei Takeuchi (1933–) was born in Tokyo, Japan, and graduated in 1956 from the University of Tokyo. He received a Ph.D. in economics in 1966 (Keizaigaku Hakushi) and his research interests include mathematical statistics, econometrics, global environmental problems, history of civilization, and Japanese economy. He is the author of many books on mathematics, statistics, and the impacts of science and technology on society. His 1976 paper, although obscure and in Japanese, is important as it gives the general result from Kullback–Leibler information, now called TIC in honor of his name. He is currently a professor on the Faculty of International Studies at MeijiGakuin University and Professor Emeritus, University of Tokyo.

I will provide a brief summary of some of the main issues. The remarks below are written from a science perspective because that is what model based inference is about. I wrote this text for others interested in good science strategies and effective methods and the important concept of *evidence*. Application of the information-theoretic approaches are very broad and potentially useful over a very wide range of science and nonscience applications.

7.1 The Science Question

The central science question is of critical importance and one must always ask if the question is worthy of study and well focused. The emphasis of this textbook is on a science philosophy that encourages hard thinking to derive a small set of plausible science hypotheses, H_i. I think this issue might often take a good person a substantial amount of time and mental effort over several weeks or months. Here one must work hard to define a set of good alternative hypothesis concerning the overall science question. Study of the literature is often a starting place; here one is encouraged to read broadly and not just on the very specific species or process of interest. One should confer with others, attend relevant meetings, use e-mail to correspond with others, ask questions, and try to gain new insights. The emphasis should be on thinking of the various alternatives.

This hard thinking process must go far beyond notions of a null hypothesis. The derivation of a small set of plausible, alternative hypotheses is both difficult and rewarding. This is not something that can be done in an afternoon or a few days; one should be prepared to put their mind to this critical matter. Chamberlin wrote of "…the thoroughness, the completeness, the all-sidedness, and the impartiality of the investigation." He stated, "There is no nobler aspiration of the human intellect that the desire to compass the cause of things." Finally, he believed, "The vitality of the study quickly disappears when the object sought is a mere collection of dead, unmeaning facts." Akaike (Kyoto Award ceremony in 2007) advised, "Select one problem and continue to pursue it until you find the perfect solution."

Before the investigation can move ahead, the alternative hypotheses must be in place. Ideally, data collection (i.e., study design) would be somewhat optimized to try to separate the support for these alternatives and lessen model selection uncertainty. From there one wants to provide measures of quantitative evidence for members of this set and gain a comprehension and understanding of the results. Finally, the set evolves as implausible hypotheses are identified and deleted, remaining hypotheses are refined and strengthened, and new hypotheses are suggested. Some higher dimensioned models with low support might be kept if the sample size is to increase substantially for the next data set. It is this notion of evolving sets that can allow very rapid progress in a field of science. This evolution can provide fast learning if used effectively. This process does not always prevent mistakes or occasionally taking the wrong path; but science has a way of backing up and correcting these setbacks.

I feel there is often far too much descriptive work done in many of the life sciences; this seems particularly true in my fields of ecology and natural resource issues. Some *a priori* thinking can lead to a more confirmatory approach and this has a variety of rewards. We all need to think more about our strategy for doing good science. Issues such as random sampling and scope of inference by defining the population to which inductive inferences are to be made seem so fundamental; I think it is a disservice to continue to accept research papers where such basic things are clearly lacking. A culture needs to be developed to enforce and expect higher standards in our science.

7.2 Collection of Relevant Data

The collection of relevant data should deserve special attention. Utmost care should be exercised and this is not the place for volunteers unless carefully trained and closely supervised. A great deal is known about the proper design of experiments and valid sampling protocols. There are dozens of good books on both of these important topics and there is no excuse for collecting data that are fundamentally flawed. Still, I see data collected from convenience sampling where any valid inductive inference from the sample to the population is precluded. In some cases the population of interest is not even defined. I see obvious confounding in experiments and a lax attitude where many variables are measured just because they are easy to measure. Many fields in the life sciences could benefit from more coursework in experimental design and sampling theory. The information-theoretic approaches are not meant to fix bad data, we must accept these challenges as a way to make progress in our science.

Large sample size conveys many advantages in the empirical sciences as does the use of many replicates. Estimators have better performance, precision is enhanced, and evidence for the alternative hypotheses is sharpened; all of these allow better understanding as a result. I see many papers that ask good science questions but they have only 20–50 samples and the need to estimate at least, say, 6–8 parameters. In such cases, there is relatively little information in the data and valid inferences may tend to be shallow and somewhat uninteresting.

7.3 Mathematical Models

Information is buried within the data and much of this information can often be extracted by using a mathematical model. Good models of hypotheses are essential in empirical science and are the basis for rigor in the investigation. Soule (1987) suggests, "Models are tools for thinkers, not crutches for the thoughtless." Importantly, the inductive inference is model based. Modeling is both an art and a science and this is a place where consultation with a statistician might be very helpful. Ideally, one hopes that there is a one-to-one correspondence between the jth hypothesis and its model.

Modeling is done to evaluate alternative hypotheses, gain insight into model structure, allow predictions, aid in variable selection in regression, and provide objective means of smoothing to identify patterns in the data. Modeling is an essential aspect of empirical science.

7.4 Data Analysis

Data analysis begins with the estimation of the unknown parameters and their covariance matrix for each model (these important issues are not the subject of this book; however, Appendix A provides a brief overview of likelihood methods). Other statistics also need to be provided (e.g., adjR^2, goodness-of-fit assessments, residual analyses) as these help in the critique of model assumptions. These procedures provide assurances that at least some of the models in the set are worthwhile. Then, one must have the value of the residual sum of squares or the value of the maximized log-likelihood for least squares or likelihood approaches, respectively. These values are the basis for the evidential approaches.

Several things can go awry here: using "all possible models," mixing response variables, counting estimable parameters incorrectly, doing data dredging in the middle of attempting an *a priori* analysis, failure of algorithms to converge (Appendix A.7), etc. Over-fitting and spurious effects should be avoided (see Appendix F). Advice and review by a person in the statistical sciences might be carefully considered at this stage.

7.5 Information and Entropy

The ability to quantify information has opened many important doors in science and technology. Boltzmann's entropy is the negative of Kullback-Leibler information and these are fundamental to deep science. Akaike found a link between expected K–L information and the maximized log-likelihood function and this was a pivotal breakthrough. The log-likelihood is a natural estimate of entropy. Akaike's AIC exploited this link and provided an asymptotic correction of bias. A second order bias correction was soon found and this is important to use in applications. While probabilities are multiplicative, information and entropy are additive. These fundamental quantities lead to ways to obtain a formal "strength of evidence" for alternative science hypotheses.

7.6 Quantitative Measures of Evidence

I ask graduate students, "what justifies a conclusion." This is a vexing problem for some students as well as professionals in the field. I think an answer relates primarily to "valid methodology." It is the methodology that must be assessed to judge an inference or conclusion: it is the rigor of the *process* that is important.

Hypotheses can be easily ranked using the Δ_i values. These values are pivotal in various measures of evidence as they are on the scale of information. Being able to quantify information loss is very important in applied science.

Plausible hypotheses exist only within a fairly narrow band; perhaps 0–8 or 12 on a scale of information loss if the independence assumption can be met.

It is simple to obtain the (relative) likelihood of each model i, given the data: $\mathcal{L}(g_i|x)$. These are useful measures of the strength of evidence for science hypotheses and do not depend on other models in or out of the set.

It is equally simple to compute the discrete probability of each model i, given the data: $\text{Prob}\{g_i|x\}$. These measures of strength of evidence are conditional on the set of hypotheses and their models.

Finally, an *evidence ratio* is just the quotient of 2 model probabilities (or 2 model likelihoods) and is another way to effectively quantify the evidence for any two hypotheses, as represented by their models. Only the two models being compared are relevant here, regardless of other models in or out of the set.

The hard science stops with the provision of various quantitative measures of the evidence. Following this, value judgments can be made to qualify the evidence. The investigator is certainly able to make their value judgments as are others. In many cases, honest observers will reach the same qualitative conclusions about the strength of evidence, while in other cases there may be honest differences in this interpretation.

This distinction helps scientists with the contentious issue of "advocacy." Scientists certainly have the right to clearly state and stand behind the objective, quantitative result; this is not advocacy. The qualification of the result can sometimes push the issue into an advocacy position – these become value judgments.

7.7 Inferences

Most inference methods in the life sciences are inductive and statistical. Properly done, both allow rigorous inference from the sample data to the population sampled. Initially, there is interest in the estimates of model parameters and measures of the uncertainty but one must determine which model to use as a basis for these estimates.

It now seems clear that final inferences should routinely be based on all the models in the set – multimodel inference. This important extension allows information in the data from models other than the best to be used in making inferences. The main tools here at the moment are model averaging and the use of unconditional variances to incorporate model selection uncertainty into estimates of precision. Both approaches are easy to compute and to understand. While there may be cases where inference is sensibly confined to a single model; however, the use of all the models will be commonplace. Multimodel inference is most often the effective path to reliable evidence.

There needs to be increased awareness of conditions that lead to spurious effects (e.g., Freedman's paradox). If one has little background science to guide the hypothesizing and modeling, small sample size, many predictor variables, and many models, the results will likely be largely spurious. People often fear they will miss an effect that is contained in the data; however, they should have an equal fear that they will find something that is not there at all (i.e., spurious). A type of model averaging is useful in reducing the important issue of model selection bias.

7.8 *Post Hoc* Issues

I encourage some *post hoc* examination of the data. This might include the addition of new hypotheses and models to represent them or slight changes to several of the better models. Such examination can include residual analyses and goodness-of-fit results leading to additional models. Because such examination and subsequent modeling are based on the same data, the conclusions from such activities must be recognized as being weaker than the more confirmatory inferences.

7.9 Final Comment

Given some background science and philosophy (Chaps. 1 and 2), it can be helpful to view the information-theoretic approaches at three different levels. The first level is conceptual and entails the Principle of Parsimony and Occam's razor (Chap. 2). The second level is the rigorous target of model selection – expected Kullback-Leibler information (Chap. 3). The third level provides a simple approach to application – various forms of Akaike's information criterion and quantitative measures of strength of evidence (Chaps. 3 and 4). These approaches are simple to compute and seem compelling. The entire approach seems to encourage people to be good scientists and allow fast learning and understanding.

The cutting edge in model based empirical science is the concept of multimodel inference (Chap. 5). There are substantial advantages to be realized in basing inferences on all the models in the set. Doing so is computationally trivial and easy to understand and interpret.