Short Presentations

A Proposal for Future Statistics Education

Makio Ishiguro - Tokyo, Japan

Recent developments in technology have remarkably increased the need for the quantitative analysis of data, often in ways which are difficult to satisfy by standard software. What has to be done in such cases is to build new software, and this kind of task usually leads to the problem of statistical modelling. Thus, statistical modelling is the task not only of expert statisticians, but of all those concerned with data analysis.

Hence, the increasing need for the data analyst who is capable of developing new models. Hence, some part of teaching time has to be spared to teach the art of model building.

Typical steps for statistical problem solving are: model building; model fitting; model evaluation/selection; interpretation of the results. Here, we restrict ourselves to cases where the models are expressed in the form of probability distributions for observed data. We shall discuss the software and hardware which are required to accomplish the above process in a lecture-room. Note that the availability of a computer is assumed as a matter of course.

(i) *Software*

(a) Editor: A screen editor is indispensable to build up a model, or a routine which computes the log likelihood of the model.

(b) Optimisation procedure: Maximisation of the log likelihood is the most popular procedure for fitting models to the data. Numerical procedures should be used for the maximisation. Traditional statistical procedures seem to be designed so that the computing cost will be as low as possible. However, taking into account the recent development of computing technology, we need not be afraid of some computation. We may relax our attachment to the analytical methods so as to concentrate our attention more on the physical interpretation of the data.

The optimisation procedure should be one that works without a user supplied gradient-evaluation algorithm. The run-time cost could be reduced by the gradient evaluation algorithm (see Powell, 1981; Ishiguro and Akaike, 1989). However, it is usual that the necessary effort to write down the algorithm is substantial and it is not rare that by mischance the algorithm obtained contradicts the algorithm for calculating

the function value. When the analyst is groping for a good model to describe a given phenomenon, the expectation of possible trouble often suppresses the desire to develop better ideas.

Though skill in numerical processing is important, it can wait.

(c) Model evaluation criterion: To demonstrate the model building process, we have to prepare a way to evaluate models. If we are to rely on statistical testing, we have to invent some new statistic, determine its distribution, compute percent points, then choose a suitable significant level, every time we think of a new model. It is impossible to do this process in a classroom.

On the other hand, the calculation of AIC (Akaike, 1973; Sakamoto et al., 1986) defined by

$$\text{AIC} = -2 \times (\text{max log likelihood}) + 2 \times (\text{number of free parameter})$$

is easy enough for use in a classroom.

(d) Graphical routines: It is important to see the data and the result of analysis. Human eyes are not made to read digits.

(ii) *Hardware*

(a) Video projector: It is important to show the data and results. It will be fine if the projector is of the bitmap display class.

(b) Work-stations: There should be work-stations on which the students can practice what they have learned and try their own ideas.

(iii) *Database*

(a) Standard statistical software: There should be one set of standard statistical software to do standard analysis.

(b) Subroutine bank: It is necessary to provide building blocks of models.

(c) Data bank: It is necessary to provide materials to work on. Teachers can not afford a failure. They have to succeed in the model building process to impress their students. Hence, a well-tested, interesting, fail-safe set of data has to be at hand. Watching it, students will be encouraged to develop their own models, and if they come up with some better model, the lecture will have been an impressive one indeed.

References

Akaike, H (1973) Information theory and an extension of the maximum likelihood principle. In: B N Petyov and F Csai (eds) *2nd International Symposium on Information Theory*. Akademiai Kiado, Budapest, 267-281.

Ishiguro, M and Akaike, H (1989) DALL : Davidson's algorithm for log likelihood maximisation - a FORTRAN subroutine for statistical model builders. *Computer Science Monographs* **25**. The Institute of Statistical Mathematics, Tokyo.

Powell, M J D (1981) *Nonlinear Optimization 1981*. Academic Press, London/New York.

Sakamoto, Y, Ishiguro, M and Kitagawa, G (1986) *Akaike Information Criterion Statistics*. D Reidel Publishing Company, Dordrecht/Tokyo.

Teaching Statistical Models - An Algorithm and Some Results

Afonso Varzea Tavares - Lisbon, Portugal

The items covered in the paper are as follows:

(i) teaching as an action leading the student towards a better understanding and interpretation of reality;
(ii) statistics as an area of knowledge that supplies scientific models closer to reality than classical deterministic ones;
(iii) the appeal of simplified models versus the risk of lack of rigour and of credibility;
(iv) the role of algorithmic computations versus laboratory simulation results.

The items are exemplified in an application to truncated geometric distribution compared with a classical unbounded model.

The quality of future contributions to society by present day students depends directly on their ability to understand and interpret the complex reality around them. The importance of such a point derives naturally from the fact that society's survival depends on its problem-solving oriented people.

Most important, however, are the skills of the qualified people in tasks related to guiding the evolution and initiating the progress. Here the key words are forecasting, risk, and confidence.

Teaching is a catalyst of such a reaction and so society's strength reflects its teaching quality.

The distance from reality to the classical rigorous deterministic models increases in proportion to the rigour that is required.

It is now accepted in all fields of science that reality is better described through statistical models. This is not only for understanding present events but also for process forecasting. Thus, teaching statistics is a corner-stone to most university studies, from agriculture to anthropology, management to sociology, mathematics to biology, economics to informatics.

To develop a statistical course around conceptual structures and simplified models is quite appealing; they are elegant and easy to deal with. To begin at this point seems to be strongly advisable in a pedagogical sense. However, to end here involves serious risks, considering that most of the time the assumptions of the model clearly over-simplify the reality they are supposed to model.

The first points to emphasise are the lack of rigour of the model and the departure of the model results from the experimental observations. This makes the teacher's task difficult, if he or she wants to produce real support to the course he or she conducts. The second point is the lack of credibility to the student in face of such deviation. This can even diminish the student's motivation for statistical studies.

(i) *A negative binomial experiment:* A typical situation occurs if we want to model the behaviour of an industrial machine in connection with "the time of first defect".

This situation is clearly probabilistic and binomial, and it is a classical result that the best description is the geometric probability distribution, as a special case of the negative binomial. So

$$P(x) = (1-p)^{x-1}p; \quad X = (1,2,...); \quad \sum_{x \in X} P(x) = 1$$

where x = number of time units to the first defect; P(x) = probability of x; p = probability of one defect occurring during a time unit and X = space of x (space of results).

The expected value of a r.v. having such a distribution is E(x) = 1/p. If the parameter p takes the value 0.2, then the expectation is 5 time units.

Consider the assumptions this model is based on. Amongst these assumptions we can identify: the constancy of parameter p; the upper limit infinity of the results space.

The first assumption is not far from reality if we consider a given epoch of machine life: new machine, used machine, etc. The second assumption must be viewed differently, given that an industrial machine is not allowed to run continuously up to its collapse. A periodic maintenance routine is in general established, and after maintenance the machine restarts working in original conditions. So, the second assumption is not so close to reality, and the deviation of the model is then more significant as the period of the maintenance routine becomes relatively shorter.

Consider now this last situation. The probability distribution changes to:

$$P(x) = (1/K)(1-p)^{x-1}p; \quad X = [1,L]; \quad \sum_{x \in X} P(x) = 1$$

$$\text{where } K = \sum_{i=1}^{L} (1-p)^{i}p; \quad L = \text{period of the maintenance routine.}$$

This last expression is obviously not so elegant as the first one. In this case, to compute the expected value $E_1(x)$ of r.v. (time of first defect) we proceed from the definition:

$$E_1(x) = \sum_{x \in X} xP(x) = \sum_{x=1}^{L} x(1-p)^{x-1} / \sum_{x=1}^{L} (1-p)^{x-1}.$$

The use of a calculator or computer is naturally very helpful in finding numerical results to such an expression. An algorithmic form to compute the expected value $E_1(x)$ is interesting in view of a program-oriented approach:

L ; p

N ← 0 ; D ← 0

{C_i ← (1-p) ↑ (i-1) ; N ← N+i.C_i ; D ← D+C_i} i = 1,...,L

E_1 ← N/D

Using the same value for p (p = 0.2), and L valued 2 to 50, the computational results with this algorithm are:

p = 0.2 E(x) = 5	L	$E_1(x)$
	2	1.444
	5	2.561
	10	3.797
	25	4.905
	50	4.999

The convergence of the model to the unbounded situation is apparent.

(ii) *The agreement of the model with simulation results:* To complete this line of reasoning it is convenient to collect data and compare the results with the model ones.

We have used simulation because we can characterise the process well. We used a random number generator after having tested it by a Kolmogorov test. The number of simulation runs for each result was selected to be 300 for reasons of stability.

Considering p = 0.2, L = 2 to 25, simulation results are:

p = 0.2	L	$E_1(x)$
	2	1.465
	5	2.6
	10	3.585
	25	4.705

The agreement between the model and the experimental behaviour is very clear. It is also clear that the simplified unbounded model and the simulation diverge. This divergence is noticeable for values of L lower than 3(1/p) (5 versus 4.5) and increases rapidly so that for L equal to (1/p) the unbounded model gives a result about two times the experimental one (5 versus 2.6).

(iii) *Conclusions:* Teaching statistics is important to society and so is considered in most university curricula.

It is advisable to include in statistical courses not only the study of ideal classical models, but also more complex and realistic models able to be checked against experimental results.

Our work above is but one elementary example of this.