Some answers to the 'cryptic issues for discussion' posed by Prof. D.R. Cox on occasion of his Bernoulli Lecture at Groningen University

May 12, 1997

by

Otto J.W.F. Kardaun

While being written down without specific discussion, these answers to the questions posed in [1] may differ from and tend to be more personally biased than those formulated by what may be called the 'Groninger School of Statistics', which has had a profound influence on the author in the period 1979-1986.

- *How is ill-conditioning by overfitting to be avoided?*

  In an experimental environment, ill-conditioning is provoked by the fact that in some directions of the space of the regression variables, the range of the available data is not large with respect to the measurement errors in those directions. This leads to unstable behaviour of the corresponding directional derivatives of the regression plane, i.e. of certain linear combinations of the regression coefficients. Notably ordinary least squares regression yields in that situation biased point estimates and inaccurate (often too narrow) interval estimates, the latter impairing the diagnostic capability of the OLS procedure for ill-conditioning that is not very strong.

  A more parsimonious model formulation or regularisation by introducing 'stiffness' in the direction(s) where the data variation is scarce (and with respect to some predetermined value of the directional derivative) can damp out the effects of ill-conditioning from overfitting. This does not help, however, if the dependence of the response variable in those ill-conditioned directions is of interest for the investigation.

  Amelioration is in such a situation possible by using estimation techniques based on errors-in-variable models [2,3,4], and from the experimental side by either reducing the measurement errors in the regression variables or by extending the range of the data in the direction(s) where it is most needed.

- *How convincing is A. Birnbaum's argument that likelihood functions from different experiments that happen to be proportional should be treated identically (the so-called strong likelihood principle)?*

  In objectivistic statistics (especially in the context of the Neyman-Pearson interpretation) the likelihood ratio is the quantity of actual ('physical') interest. In the extended discourse of subjectivistic statistics (Jeffreys as well as de Finetti) the absolute value of the posterior density (and hence of the likelihood) is of intrinsic interest. In Bayes Theorem, multiplying the likelihood with a constant, does not influence the transition from prior to posterior probabilities. From these two points of view, `the weak form of the` likelihood principle is reasonable.

- *What is the role of probability in formulating models for systems, such as economic time series, where even hypothetical repetition is hard to envisage?*

  In my view, it constitutes, in estimation theory, a convenient way of expressing the sensitivity of the model with respect to (certain types of) disturbances, and in prediction theory a way of expressing prediction ranges. It seems sound to keep the artificial nature of such type of models clearly in mind. By considering a number of ('similar') time series or a number of one-step ahead predictions per series, such models are in principle open to systematic confrontation with what has actually happened, see e.g. the Grecon activities, though, as always, one can argue about the specific reasons when a mismatch occurs in particular cases.

- *Should nonparametric and semiparametric formulations be forced into a likelihood-based framework?*

  I am inclined to answer this question in the negative. The likelihood approach enjoys the qualities of good asymptotic properties (which come often into effect already for reasonable sample sizes) as well as practical computability for large classes of (parametric) problems. These advantages are severely dimmed if the number of parameters increases proportional to the number of observations. Inconsistency results and identifiability problems (among others in the context of errors-in-variable models) are known since the fifties. Some version of partial likelihood (as orginated by Cox, 1972-1975) seems a good way to proceed for semi-parametric models. On the other hand, functionals of which the gradient yields estimating equations that produce consistent and efficient estimates are useful anyhow, whether or not they live in an infinite dimensional space (and whether or not they are called likelihood).

- *Is it fruitful to treat inference and decision analysis somewhat separately?*

  The answer is definitely yes. The fact that statistical inference has been put into a neat decision theoretic framework by Wald et al. is a fascinating intellectual construction, but should not become an 'idola theatri' (in the sense of Bacon) by too absolute an interpretation, because it has its artificial aspects.

  Statistical inference is largely oriented towards the scientific goal (for instance through estimation of regression parameters) of describing or predicting the behaviour of nature as expressed by (and compatible with) the dataset at hand and regularised by the class of models applied by the investigator.

  In decision analysis, the objective is to reach decisions based on data, models, and (perceived) loss functions for the various actions. Hence, the scope of application of decision analysis is broader than of statistical inference, which entails –in principle at least– a difference in methodological aspects in concrete situations, even though *formally* statistical inference can be viewed as a sub-area of decision analysis.

  Treating these two areas of scientific endeaviour somewhat distinctly is, among others, useful to describe the conflict of interest that may arise between the goals of practical decision analysis (e.g. how to treat a particular patient or to optimise the performance of a particular nuclear fusion device) and of statistical inference (how to estimate accurately, and reliably, the physical effect of a certain treatment). See e.g. [5], pp.482, which was inspired by [6].

- *How possible and fruitful is it to treat quantitatively uncertainty not derived from statistical variability?*

  Quite a few initiations have been made in fuzzy set theory, see the didactical monograph [7], and, for a concise description, [8], pp.4. [9] gives a recent discussion. These seem interesting in the context of (generalised) probability theory. A basic problem is how to assign, in a statistical context, values corresponding to these uncertainty measures. Subjectivistically, one could think of some extension of the elicitation methods presented by Prof. Molenaar during the Bernoulli mini-symposium. Objectivistically, I do not know how to collect empirically instances of necessity and possibility relations.

- *Are all sensible probabilities ultimately frequency based?*

  Personally, I would be in favour of using consistently another word than 'probability' for each type of uncertainty that has not a frequency based underpinning, e.g. 'degree of conviction' for a 'personalistic probability' (de Finetti, Savage), 'verisimilitude' for a personal probability of an ideally coherent and objective person (Jeffreys), and 'credence' for what might be called an intersubjective (non-frequentist) probability, a 'reasoned common degree of belief' formed by scientific investigation, discussion and evaluation, see e.g. [10].

- *Was R.A. Fisher right to deride axiomatic formulations (in statistics)?*

  Not entirely, even though over-emphasis of axiomatisation tends to lead to mental paralysis. Axiomatisation is useful (a) as an intellectual game, (b) as a didactically efficient way of summarising a number of results, while, at the same time, (c) imposing a certain mathematical structure as 'the most appropriate one' for a domain of discourse. Of course, any 'good' axiomatisation induces (c) only with due scientific reason. To my knowlegde, there exists little systematic theory how to evaluate formal correctness versus empirical relevance of axiomatic systems.

- *How can randomization theory of experimental design and survey sampling best be accomodated within broader statistical context?*

  —

- *Is the formulation of personalistic probability theory by de Finetti and Savage the wrong way round? (It puts betting behaviour first and belief to be determined from that.)*

  Betting behaviour seems to me a reasonable operational way of eliciting, i.e. making explicit to other people (possibly including the person himself) personalistic probabilities, without any assessment of which comes first, the egg or the chicken. The last question is often difficult to answer anyhow. (Let me recall in this context that some evolution theorists have defined a chicken as the way an egg produces the next egg.)

- *How useful is a personalistic theory for public discussion?*

  —

- *In a Bayesian formulation, should priors constructed retrospectively after seeing the data be treated distinctively?*

  To reasonable approximation: no. The posterior probability based on one set of data (and on an 'original' prior distribution) can be viewed as a prior distribution (and used as an 'input quantity') for analysing another set of data, and so on. If one looks closer, there is a fine distinction between the original prior distribution (not based on any data) and all subsequent prior distributions (partially based on data). Of course, care should be taken not to use the same data for choosing a prior distribution and testing an hypothesis about the actual distribution compatible with the data.

- *Is the only sound justification of much current Bayesian work using rather flat priors the generation of (approximate) confidence limits? Or do the various forms of reference priors have some other viable justification?*

  Let me restrict attention to a particular example. In a physical context, it has been occasionally proposed to use Jeffreys' prior $f(\theta) = c/\theta$ for the distribution of a positive, dimensional physical parameter (such as length), based on the argumentation that this is the only density invariant under the choice of units. This argumentation is not compelling since it seems natural to consider, for each choice of units, coherently, one member of a one-parametric scale family of priors distributions.

- *What is the role in theory and in practice of upper and lower probabilities?*

  If with upper and lower probabilities is meant the end-points of a confidence interval (or an interval estimate) of a 'success' probability for an individual, it seems to have some use in practice, as illustrated in [11,5,12], among others. In nuclear fusion, it is an interesting question to transform an interval estimate for the confinement time of the ITER tokamak into an (accurate) lower bound for the probability to achieve a sustained burning plasma that produces at least a certain amount of heating power (e.g. 1 GW), as a function of the plasma density, the auxiliary heating power, and some other plasma characteristics that can be controllled within a certain operating region.

— *References* —

1 Cox, D.R. (1997), The Nature of Statistical Inference (The 1997 Johann Bernoulli Lecture) Nieuw Archief voor Wiskunde, IV, 15, 233-242.

2 Anderson, T.W. (1984), Estimating Linear Statistical Relationships (The 1982 Wald Memorial Lecture), Annals of Statistics 12, 1-45.

3 Hillegers, L.T.M.E. (1986), *The Estimation of Parameters in Functional Relationship Models*, Ph.D. Thesis, University of Eindhoven.

4 Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995), *Measurement Error in Non-linear Models*, Chapman and Hall, London.

5 Kardaun, O.J.W.F. (1986), *On Statistical Survival Analysis*, Ph.D. Thesis, University of Groningen, Chapters 2 and 3 also in: *Handbook of Statistics* VIII, Elsevier (1991).

6 Schwartz, D., Flamant, R. and Lellouch, J. (1980), *Clinical Trials*, Academic Press, London.

7 Klir, G.J. and Yuan, B. (1995), *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Englewood Cliffs, New Jersey.

8 Kardaun, O.J.W.F. and Kus, A. (1996), Basic Probability Theory and Statistics for Experimental Plasma Physics, IPP-Report 5/68, Garching.

9 Laviolette M., Seaman, J.W., Jr, Barrett, J.D., Woodall, W.H., et al. (1995), A Probabilistic and Statistical View of Fuzzy Methods (with Discussion), Technometrics 37, 249-292.

10 Kardaun, O.J.W.F. (1995), Uncertainties in the Confinement Prediction for ITER, in: Scientific Minutes 3^{rd} ITER Confinement Database and Modelling Workshop, Naka, Japan, Oct. 16-18, IPP-IR-96/5 1.1 (http://www.ipp.mpg.de/ipp/netreports).

11 Schaafsa, W. (1984), Me and the Antropologist, in: Proc. of the Seventh Conference on Probability Theory 1982, 333-343, Editura Academiei Republicii Socialiste Romania, Bukarest.

12 Kardaun, J.W.P.F. and Kardaun, O.J.W.F. (1990), Comparative Diagnostic Performance of Three Radiological Procedures, Methods of Information in Medicine, 29, 12-22.

Original text: Utrecht 16.5.1997, Garching 25.5.1997

Revision (text in this font): 25.11.1997, (references): 28.12.97, 22.3.98.

Acknowledgement: Discussion with Prof. Dr. W. Schaafsma et al. at Kyushu University.