Discussion of Papers on the Foundation of Statistical Science

Steffen L. Lauritzen
Aalborg University, Department of Mathematical Sciences
Fredrik Bajers Vej 7E
DK-9220 Aalborg, Denmark
steffen@math.auc.dk

1. Introduction

First let me congratulate the authors for writing such stimulating papers on the foundations of our subject. Since the mid 70s there has been remarkably silent concerning this vital topic, partly because discussions in earlier times seemed to lead to more dispute than clarification and partly because statisticians more or less have placed themselves in opposing camps with no communication possible.

It is therefore particularly welcomed to see these three papers of which the first two in one way or another are concerned with an attempt to bridge the gap between the camps, whereas the third paper presents a fascinating alternative foundation for a part of our science. My comments will largely be directed to the first two of today's speakers, as I my other colleague has concentrated on the latter.

2. A main example

To focus the discussion I will use a variant of a paradox of Armitage (1961,1963), see also Stein (1962).

We are considering two experiments. Each of these have as their basis an infinite sequence $X_1, \ldots, X_n, \ldots$ of i.i.d. normal random variables with mean θ and variance 1. The object of inference is θ. The experiments are as follows:

Experiment I: The sample size is determined as $n = 100$. The outcome of the experiment has $\bar{x}_n = 1.1$.

Experiment II: The sample size is determined as $n = \inf\{m \mid |\sqrt{m}\bar{x}_m| > 10\}$. The outcome of the experiment has $n = 100$ and $\bar{x}_n = 1.1$.

The law of the iterated logarithm ensures that the second experiment will terminate, even if $\theta = 0$, and with a likelihood function that displays overwhelming evidence agains this hypothesis. Variants of the paradox bear the name of *sampling to a foregone conclusion*.

The paradox is that the two experiments have identical likelihood functions, whereas most statisticians would agree that the second experiment is manipulated and the evidence is not to be trusted. One of the issues addressed in the first two papers is precisely concerned with this and similar paradoxes.

I will refer to the second experiment as the *experiment of the manipulating scientist*. To design an experiment as the second, the main purpose of the scientist must be obtaining a seemingly strong conclusion and get a paper published rather than searching for truth.

3. Asymptotic inference

Professor Pierce is emphasizing that asymptotic considerations lead to a distinction between paradoxes due to the application of stopping rules, such as above, and effects of sample procedures due to censoring mechanisms. The latter are more benign, and if their effect is not asymptotically negligible, they are at least controllable within the accuracy that inference is at all possible.

But what is the formal distinction between manipulating and non-manipulating experimental designs? Both censoring and sequential procedures lead to 'uninformative' designs in the sense that the likelihood function is not affected. Is there a precise formulation that enables an innocent statistician to distinguish?

Can a martingale or game-theoretic approach help us out here? In the decision theoretic framework for statistical inference, statistics is seen as a game played between nature and the statistician. Could we introduce the possibly manipulating scientist as a third player? In a very fundamental sense there is nothing asymptotic about these problems. They lie at the heart of the challenges of our science.

4. Likelihood

Professor Royall is bridging the gap by emphasizing that evidence is evidence, even though it could be misleading. I have sympathy with this point of view, but dislike that the concepts of likelihood and evidence are merged completely into one and the same.

The concept of likelihood is clear, well defined and fundamental to statistical science. Most uses of the English language, including scientific uses, give a meaning to the word 'evidence' which has many more facets, depending on the context in which it is used.

Nevertheless I found the likelihood based approach to design of experiments described by Professor Royall most illuminating and interesting. The classic Neyman–Pearson based approached works well in a quality control setting where it seems to fit well with the main purpose of testing, but considerations often seem absurd in more scientific contexts.

However, also Professor Royall's approach seem to fail to deal with the manipulating scientist. In the second of the experiments described, the probability of obtaining misleading evidence seem to be equal to 1 if $\theta = 0$, despite the claim that this probability in general is bounded above by a small constant. How would Professor Royall deal with the manipulating scientist? Is there any other way than calling his cards?

REFERENCES

Armitage, P. (1961). Contribution to the discussion of Smith (1961). Journal of the Royal Statistical Society, Series B, 23, 30–31.

Armitage, P. (1963). Sequential medical trials: some comments on F. J. Anscombe's paper. Journal of the American Statistical Association, 58, 384–387.

Smith, C. A. B. (1961). Consistency in statistical inference and decision (with discussion). Journal of the Royal Statistical Society, Series B, 23, 31–37.

Stein, C. (1962). A remark on the likelihood principle. Journal of the Royal Statistical Society, Series A, 125, 565–568.

RESUMÉ

On discute quelques problêmes d'inference de la vraisemblence asymptotique avec un exemple d'expériment sequentiel decrit par P. Armitage.