Discussion by Professor Lucien LeCam

(University of California at Berkeley)

Professors Berger and Wolpert are to be thanked and congratulated for giving us a closely argued view on the foundations of statistics. Their arguments in favor of the Likelihood Principle are very persuasive indeed. One may suspect, however, that some readers will be convinced and converted while some others will hold fast to their misguided beliefs, in spite of all the evidence.

I shall try here to indicate why the present writer belongs to the latter category.

There is a body of statistical theory, call it "type 1", that deals with the following kind of systems. When contemplating a particular unresolved question, one devises experiments to ascertain what the facts are. The mathematician will abstract the idea of "experiment", using an object formed by a family of probability measures on a suitable field. The consequences of using particular procedures to analyse the "experiment" are then describable in probabilistic language. One can attempt to single out procedures that have a reasonable performance in this probabilistic world. That is a bit like selecting tools: wrenches are often, but not always, successful at unscrewing bolts; paint brushes often fail in the same activity.

This kind of endeavor has given us the Neyman-Pearson theory and Wald's theory of "statistical decision functions". One can readily claim that the whole enterprise is misguided, but it does seem to have a role to play in certain endeavors, like planning experiments, settling arguments that involve several scientists and odd questions such as "is methotrexate effective in the treatment of colon cancer."

There is another body of theory, call it "type 2", that deals with axioms of coherent behavior and principles of evaluation of evidence. Some of it, and perhaps most of it, has to do with what "one" should "think" after the results of the experiments have become known. Comparatively little has been written on how "one" can transmit the "evidence" to another person, even in the Berger-Wolpert text, this communication problem takes second place to the " one should think" question.

Berger and Wolpert see evidence of contradiction between the "type 1" and "type 2" approaches. In a strictly mathematical view of the problem, there is no overlap between the two approaches because "type 1" does not have any probabilities to play with once the dice have been cast.

Consider for instance an experiment involving two containers, one with 50% red objects, the other with 25%. A coin is tossed to select a container. Then one extracts a ball from that container. It turns out to be blue. When all of that has been properly carried out there are no probabilities left since the container has been selected. It is either the first or the second and not a probabilistic mixture of both. Any assignment of probabilities at that stage requires amplification of the model, with thinking about possible repetitions of the experiments or degrees of belief, or betting strategies or whatever.

Berger and Wolpert try to convince us that in such a situation one should follow the likelihood principle. The argument is thorough. L. J. Savage's argument was also very thorough, but I have yet to find a *scientist* who would be convinced by a posterior distribution on the methotrexate and colon cancer question if the prior has been supplied by a pharmaceutical company. The point is that one can easily argue oneself into a corner.

In the present case, however, I think that the argument has one major flaw. It is based on the assumption that given an experiment E, and the result x of that experiment, there is a well defined object Ev (E,x). The nature of the object Ev (E,x) is not described explicitly. This is not the problem. What matters is the assumption that there is such an object, or more specifically a function $(E,x) \to Ev\,(E,x)$. Starting from such an assumption, and adding a few other "principles", one can prove that the function Ev must have certain properties.

(I am reminded here of the standard "proof" that $1 = (-1)$, assuming that $\sqrt{}$ is a function: certainly $\sqrt{1/(-1)} = \sqrt{(-1)/1}$ and by multiplication $(\sqrt{1})^2 = (\sqrt{-1})^2$. I am also reminded of Spinoza's "Theologia more geometrico demonstrata").

The very existence of the function Ev is not clear to this writer. Even if it exists in a strictly mathematical abstraction of "experiments" and results, the relevance to practical applications is not directly evident.

Several years ago a problem of this nature was raised during the conference on the Future of Statistics held at Madison. Someone asked the panel how they would report the evidence in a clinical trial of a drug intended to suppress renal calculi. The answer, given by G. Barnard, was "report the likelihood function". That may be, but one should also report the age, ancestry, health status of the participants, the presumed mode of action of the drug, its manufacturer, ideas about whether calculi occur in clusters or bunches, their size distribution, whether their formation may be spurred or hindered by nutritional factors, etc., etc., including whether the randomization used (or unused) led to apparent disbalance.

There is no shortcut to reporting what was actually done and observed. In situations involving games of chance with definite rules, one might simplify the evidence report. It is also true that Savage could argue that anyone playing games according to Savage's rules need only report (to himself) the resulting posterior distributions.

It does not follow from such mathematical theorems that one must necessarily frame practical questions in terms of Savage's games or in terms of the Berger-Wolpert rules of evidence, even if these authors eventually argue themselves into a Bayesian framework.

Here the situation is complex because "type 1" theories have given proofs that "experiments" are characterized by the distributions of their likelihood functions. Also it is a standard result of "type 1" theories that Bayes procedures, or their limits form complete classes. A main difference is that the "type 1" theories insist that they are about risk functions, not possible interpretations of single posterior distributions.

The passage from advocation of the likelihood principle to Bayesian theory is described by Berger and Wolpert but not as a strictly logical consequence of the L. P. principle and other explicitly stated axioms. It is weak compared to the rest. However, in the process, they also demonstrate that they do not abide by their own L. P. prescriptions.

This occurs in the discussion of an example of C. Stein. The authors say

"note that it was assumed that $x = y = \sigma d$ in the above conditional analysis, and *since it can be shown that* Y *is almost certain to be enormous....*" (emphasis added).

That seems to be a very direct appeal to a frequency evaluation of the situation, and not even a conditional one at that. Such an appeal does not fit with the logic of the rest of the paper.

There are other matters that should be discussed, but it would take too much space. One of them has to do with approximation. Assuming that the function Ev exists and that if (E_1, x_1) and (E_2, x_2) give the same likelihood function, then the evidences are the same, is one entitled to presume that if the likelihood functions are 'approximately' the same then the evidences are also "approximately" the same?

Here we have two "approximately" with undefined, but perhaps definable meaning in the first instance and an apparently undefinable meaning in the second occurrence since Ev (E,x) itself is an undefined object.

For instance it is a classical result that if one takes a very large sample $(x_1,\ldots,x_n)$ from the standard Cauchy $\{\pi[1+(x-\theta)^2]\}^{-1}$, for "most" samples the likelihood function will be "close to" one obtainable from a single observation y from $\mathcal{N}(\theta, \frac{2}{n})$. Does that have any "evidential" meaning for the L.P.? Must one necessarily interpret it only through a computation of posterior distributions? If so, for what priors?

To summarize Berger and Wolpert have given us a valiant defense of the L.P. However it does depend on a basic assumption of existence of the evidence function Ev. This function, if it exists, does not conform to the tradition of reports in scientific journals. The theory does not actually conflict with the so-called "classical" one because their domains of existence are separate and their aims different.

This author presumes that there is some value in some of "classical statistics" and also in the likelihood principle, but feels that one cannot support the practical application of either (or of other theories) on purely mathematical grounds. One should keep an open mind and be a bit "unprincipled".

DISCUSSION OF THE SECOND EDITION BY PROFESSOR LE CAM

In order to avoid any misunderstanding, let me repeat two of the criticisms made above: Using "evidence" as a function of a pair (E, x) and using "approximate" likelihood functions. The two points are highly interconnected.

For simple experiments Professors Berger and Wolpert use a mathematical entity E that consists of a set Θ and a family $\{P_\theta, \theta \in \Theta\}$ of probability measures on a given σ-field.

"Evidence" is undefined, but it is supposed to be a function of the pair (E, x) where x is the observed value. This may seem innocuous, but it is definitely not. It rejects a part of the data usually considered as part of the evidence in the common language use of the term, namely the thinking that went into the selection of the mathematical entity called E.

Except perhaps in those cases where the randomization is man-made on purpose and perhaps also in the Poisson formulas for radioactive decay, the selection of the "model" E is based on rather loose arguments that are not themselves representable by pairs (E', x'). Here, by "model" I do not mean something like "linear models" but more a mathematical construct that attempts to catch the important features of a physical or biological phenomenon. I have, of course, no objection to a theory of "evidence" based on a function of pairs (E, x). It just fails to connect properly with my own intuitive notion of what evidence is. Therefore I do not feel bound in practice by the theorems derived from such a theory.

Even if one tries very hard to put the information in the form (E, x) one will almost always put in E certain formulas for the sake of convenience, simplicity or plain laziness. Thus (E, x) will only be our approximation to a "better" (E', x'). It is my feeling that, if one wants to take into account the fact that such approximations are the rule, one must also explain what differences they may make in the use of the undefined "evidence" Ev (E,x). I am not too sure that this can be accomplished without introducing in the system a variety of concepts that go beyond pairs (E, x).

In summary I remain opposed to the apparent normative aspect of a theory that says that I *must* abide by the LP when I am unable to put my emotions and various bits of knowledge, or lack of knowledge, into it.