The Foundations of Statistics Are There Any?
Author(s): J. Kiefer
Source: *Synthese*, Vol. 36, No. 1, Foundations of Probability and Statistics, Part I (Sep., 1977), pp. 161-176
Published by: Springer
Stable URL: http://www.jstor.org/stable/20115219
Accessed: 25-05-2016 23:55 UTC

J. KIEFER

THE FOUNDATIONS OF STATISTICS—ARE THERE ANY?

1. BEGINNINGS

The fact that an author asks such a question as that of the title no doubt makes his answer easy for most readers to guess. Still, some explanation of the answer may be called for. Perhaps it is best to begin by explaining the question, and that will require some definitions. Many contributors to this volume, experienced in philosophical and foundational writing, may find my interpretation of some words and ideas shallow, biased, or incomplete. (Also, the paper, like almost all on foundations, will no doubt be viewed as somewhat polemical!) These remarks will nevertheless outline some of the misgivings I share with a number of statisticians, at the seriousness with which certain foundational developments are viewed by their proponents, as the salvation of statistics.

To narrow the discussion, my meaning of *foundations* will be restricted to the context of a *predictive science*, and in such a science a *successful foundational scheme* will be a set of undefined objects and axioms from which useful predictive rules can be derived. (This need not mean that the statistical problems referred to later are necessarily phrased as problems of prediction of the value of a random variable to be observed in the future.) We are not concerned merely with a collection of observations about some aspect of the real world, but about a collection of rules of sufficient simplicity and generality in their applicability that they can be useful in predicting what will be observed in other settings involving the same type of phenomena. The obvious example is physics, and I note only that we allow for more than one successful foundational scheme and need not even assume there is any God-given 'correct' one; on the other hand, we are not interested here in most of the foundational schemes that appear as answers to the proverbial final examination question in

Synthese **36** (1977) 161–176. *All Rights Reserved.*

cosmology, 'define 'universe' and give three examples'; if such foundational schemes are included, the question of the title has a positive but valueless answer.

Thus, also, we are not talking about areas of knowledge such as political history, and we may argue about whether we include economics. There are some economic principles that are usefully applied, but we have all seen the limited applicability of much of macro-economics in a real world that rarely satisfies the necessary conditions for proposed theories to be used. Humans, and collections of them, are less regular than inanimate objects and plants, at least in terms of our current ability to predict their behavior.

How do statistics and foundational systems for statistics fit into this picture? First, some more definitions.

A *probability structure* here means a classical (Kolmogorov) one, in terms of which the *frequentist* viewpoint of statistics, as emphasized by Neyman, is based; *subjective probability* will be labeled as such, and *Bayesian* will mean *subjective Bayesian* here unless the existence of a physical prior law is mentioned. (Thus, notwithstanding the opinion of a colleague of mine that, for statistics, the important ideas of the meaning of 'probability' are those *not* to be found in Feller or Loève, the term will be used here as it is by such authors as these.) Probability theory studies the behavior of complex chance systems made up of simpler chance entities whose laws are given. It provides a useful foundational system in a number of scientific disciplines, in which its conclusions have been seen repeatedly to correspond to what is observed in many real world situations; such settings exhibit outcomes that conform closely to what we would guess from such results as the law of large numbers. If insurance companies and gamblers couldn't make long run predictions, they wouldn't be in business.

A *setting for statistical inference* is a collection of *possible* probability structures, exactly one of which (called the *true state of nature*) is the mechanism actually operating in the setting at hand; which one, is unknown. *Statistical inference* in the broad sense is the collection of procedures for reaching a conclusion about which possible state of nature is the true one in such a setting, together with the probabilistic properties (operating characteristics) of such procedures.

This last definition is purposely quite broad, and the controversial possible choices of narrower definitions will be discussed in the next section.

2. WHAT IS STATISTICAL INFERENCE?

Which real world settings, where data are collected, should be regarded as settings of statistical inference? We shall take a broad view here, not for the sake of generality, but because the distinctions that are often made seem to me to be either illusory or else due to practical difficulties in defining the setting precisely. Thus, a number of authors have distinguished between problems of *decision* and of *inference* or reaching *conclusions* (e.g., Tukey, 1960); and Birnbaum (for example, in the present volume) has made a similar distinction between the *behavioral* and *evidential* interpretations of decision. Examples are often given of the difference between typical statistical problems of technology and of science: a manufacturer decides (with resulting *action*) whether a lot is defective or not; but a scientist (let us call him 'A') might only state that a particular hypothesis (= probability structure) is to be rejected without asserting which state is guessed to be true, or that a particular subset of the possible states of nature seems to him to contain the true one (in which case, if the set is too large to be useful, or if it carries little confidence, we shall call it a *weak decision*, and, otherwise, a *strong decision* – a simplistic dichotomy for the sake of brevity), or that a likelihood function (or perhaps a fiducial distribution or subjective posterior law) he presents reflects the plausibility of various possible states being true. Compare this scientist with scientist B who analyzes a series of observations and decides that they verify (as a useful approximation to truth) the law of gravity for rocks thrown near the earth's surface, including the value of the gravitational constant. Certainly B has used statistics as a tool of predictive science, the gravitational law being useful for predicting what will occur in some future settings; moreover, his results may serve as building blocks for further development of the subject. His results are not 'actions', but his decision implies future actions by him and others, and might be made with greater certainty than that of the manufacturer (who *must* act, whether or not confident).

Almost every A would like to be a B. But A's setting is generally 'dirtier' in one of several respects: (i) he does not have enough observations to feel confident about making a precise assertion about what is true (a strong decision); (ii) he could not, before taking (or analyzing) the observations, make a complete list of the possible true laws and assertions (or decisions) for his setting. Because of (ii), the framework of formal decision theory may be useless to A; and, even if he were not faced with (ii), and hence could list the possible states of nature and corresponding assertions, the presence of (i) would deter him from making anything but a weak decision. Still, he would like to discover (approximate) truth and to build upon it, or have others do so, in further investigations. One may regard his experiment as a first stage in a complex sequential undertaking in which *the possible decisions* (other than, perhaps, weak ones) *are themselves not delimited at the start.* While A cannot now list all possible outcomes of the various stages of this undertaking, or the subsequent stages they will suggest, he would not be carrying out the present stage without the conviction that those subsequent stages will develop in such a way as to lead to the discovery and/or verification of a truth about nature (or to the strong decision that he should quit). It seems more fruitful, in assessing the foundations of statistics, to regard A and B as having the same ultimate goal in the use of statistics, even though the settings they encounter may differ in complexity by orders of magnitude, rather than to regard their problems (decision or inference, etc.) as being fundamentally different because this *first stage* is so different.

From this point of view, the presentation of the likelihood function as illustrative of inference rather than decision is unsatisfactory. For, a next stage of experimentation, and perhaps a more precise listing of possible future terminal decisions, must be *decided upon* from that function. The same is true for the listing of all data and details of the experiment, that some of us would prefer to give in poorly defined settings with inconclusive observations. The choice may be more complex and less well-based than in simple technological settings; but it is dodging the issue and ignoring what happens next, to view the likelihood function (with perhaps different modifications corresponding to the prior laws of different Bayesians) as a different *ending* to a statistical problem, to be contrasted with a decision. (With this in mind, we shall aim our criticism in most subsequent discussion, at Bayesian decision theory rather than at axioms that supposedly lead to inference rather than decision. We will not take the space here to discuss other criticisms that have been raised about the likelihood principle, regarding identical interpretation of two identical likelihood functions in settings in which the two experiments and sets of possible states of nature differ.)

The separation of descriptive statistics from a more narrowly defined inference or decision theory often seems artificial, in this light. The informative summarization and presentation of data is frequently not only for the purpose of recording history, but for the purpose of assessing the possible mechanisms and deciding on the next stage of experimentation or a future action. (Comprehensive studies of pollution, such as that of Box and Tiao, or of income and wealth distribution, are best viewed in this way.) Even when the presentation is primarily of historical interest, the choice of statistical method (width of histogram intervals, method of smoothing, etc.) is presumably motivated by the same desire to approximate the truth well as is the choice of the physical or biological scientist who is trying to answer a more specific question. In another methodological domain, Tukey's extensive collection of methods of exploratory data analysis is aimed at treating settings in which the possible states of nature and decisions are typically not easy to formulate in advance of the experiment; but the methods are surely also to be viewed as being used in one stage of a larger undertaking in which there lurks the aim of an ultimate strong decision or an accurate assertion about nature.

The *foundations* of statistics should not consider decision and inferential measure such as likelihood as *final steps* of two kinds of problems and distinguish between them in that form. A more sensible distinction would be one that accepted the similar ultimate purpose of statistics in complex problems of varying completeness in their formulation, and which would regard decision theory (as the term is now understood) as a simple but useful case of a more general incompletely formulated setting of the nature mentioned above. (The latter is what most scientific investigators are faced with, but an appropriate foundational system that yields useful consequences about such complex settings, has not yet appeared.) It would be pretentious to view these as analogues of, say, Newtonian and relativistic mechanics, but there is at least a partial analogy.

In summary, the distinction between decision and inference (as final products of the axioms) seems questionable at the foundational level;

however, even if the distinction is accepted, in settings of inference we are left with the unanswered question of what to do next on the basis of something like the likelihood function; we now look further at the question of the title in settings of decision.

3. FOUNDATIONS OF WHAT?

Successful theories in the physical or biological sciences are able to predict how entities (or collections of them) behave under various circumstances; questions of how they *should* behave (e.g., occasionally violating gravity or the laws of thermodynamics, to aid mankind during the energy shortage; or to achieve some more heavenly purpose) do not any longer arise in the minds of many scientists. Social sciences often do not stop short of such questions; economists recommend some action to their government (or perhaps revolution against it) to achieve the 'better' of several possible economic (or economic-political) states. This is a concern with how people and other economic entities *should* behave. Genetic engineering, as distinguished from an understanding of the DNA mechanisms, may be viewed similarly.

Are the foundations of statistics concerned with how statisticians behave or how they *should* behave? To start with, we accept the unfortunate reality that human behavior is necessarily involved in planning and analyzing experiments. Scientific knowledge and technological invention are human perceptions and developments, and they have proceeded from the knowledge at previous stages only with the intervention of human ingenuity. No computerized knowledge producer is available to avoid this; moreover, the programming of the rules by which such a computer would make choices would itself ultimately be a human choice. These obvious comments are made because many foundational developments are evidently attempts to find a single simple prescription which is meant to provide a satisfactory 'program' for a large collection of settings of widely differing character. A step in these developments is the confusion between the consequences of systems of axioms of rational (or coherent, etc.) behavior, which presumably tell us how (rational) statisticians behave, and the question of how one *should* behave (choose a terminal decision or the next stage of experimentation) in an actual setting.

These are not the same because the axioms, which lead to 'acting like a Bayesian', are not complete in the sense of mathematical logic. (We are discussing here the axioms for decision making; as indicated earlier, axioms that lead to the presentation of, say, the likelihood function, do not come to grips with the experimental or terminal decision that must still be made.) That is, anyone who acts relative to any subjective prior law whatsoever is 'rational'. My own observation, on questioning a number of practicing Bayesians, is that many do not appreciate that there is no absolute (as distinguished from subjective and individualistic) basis for asserting that this *should* be their behavior; nor even that, if a physical prior law exists, their rational behavior might not minimize the actual expected loss. In a setting where we have a precise listing of all components of a suitably simple and regular decision-theoretic model, Wald's work shows that this rational subjective Bayesian behavior is equivalent to picking an admissible procedure. (*Variation* in settings is discussed in Section 5.) Although the picking out of an admissible procedure is equivalent to acting like a Bayesian relative to *some* prior law, many of us do not find it helpful in practice to know this conclusion of the mathematics of behavior; we are not comforted by the unthinking self-satisfaction each of a number of subjectivists may attain upon using his own prior law, all in the same setting. As Le Cam (1968) put it, "Bayesian statistics is intended for personal views, it does not provide any way of scientific discourse."

To summarize, these foundations for a theory of how rational individuals *do* act, at best cannot provide the desired program for automatic decision making, and at worst have deluded some practitioners into believing that the foundations provide such a program for how they *should* act.

It may be objected that (in the simple, well-defined model discussed shortly above) the non-Bayesian decision theorist's presentation of the minimum complete class of all admissible procedures is also not a satisfactory end. I agree. One must still pick the admissible procedure actually to be used. In a given setting, two non-Bayesians might pick different procedures. A propos of this, a theoretical Bayesian who is among the deeper thinkers about foundations has written me, in criticizing some non-Bayesian, frequentist conditional confidence procedures, that "telling [the experimenter] that his confidence should depend on the procedure he was using is lame, lame, lame." To me it seems just as lame that a scientist announces a conclusion and his measure of belief in it, both dependent upon his subjective prior law.

This degree of arbitrariness non-Bayesians perceive in the choice of procedure is an inescapable feature of most statistical settings. Moreover, the subject matter of statistics – the collection of settings of statistical inference – is so complex that many of us believe it impossible to codify it in terms of a simple program that will pick a suitable procedure in each setting. Extant systems of statistical foundations are aimed at achieving such simplistic codification.

What remains of our original question is, ignoring the objections thus far stated, and being pragmatic, has such a codification proved *useful* in predictive sciences, in practice?

4. DOESN'T IT WORK?

Since subjective Bayesian analysis is used by so many more people than it was twenty years ago, must its foundational system not be a useful one? In answering this, we look at fairly strict adherents to the principle of making decisions in terms of one's own subjective prior law. We do this because the choice of a decision is not primarily a consequence of the subjective Bayesian foundations if, for example one uses a number of possible subjective prior laws to obtain operating characteristics of Bayes procedures that are compared further before choosing a procedure (which therefore, from our previous discussion, amounts in a fixed setting to choosing among that number of admissible procedures); or if the prior law used is obtained as a good approximation to a genuine physical prior law (which, if it exists and is known, is what any of us would use, with some precautions as described in connection with the procedures T_0 and T_1, below).

Some Bayesians have let it be known that, by this stage of development of the methodology, there is more than one Bayesian school. The present discussion cannot be trusted to reflect understanding of all nuances of difference, but one belief that seems no longer to be as widely held as it once was in some Bayesian quarters, is the view that a scientist or business man can always find a prior law that has a strong *physical* basis, by sufficiently thorough examination of the background of knowledge related to his present setting. There is a philosophical question here, of whether it is meaningful that such a physical prior law exists (does Nature throw dice in choosing the true natural law?), let alone whether we can come close to knowing it if it does. When I once asked a leader of Bayesian thought what prior law he would use for the distance to a new astronomical body the like of which no one had seen before (e.g., the first pulsar), he responded with a collection of betting assertions such as 'surely you wouldn't bet more than a million to one that the body is more than 5×10^9 light years away.' He then went on to point out that, for sufficiently large sample sizes, the exact nature of the prior law tends to disappear in the posterior law and in the numerical value of our estimate of the true distance. This line of argument indicates no belief in the availability of a good approximation to a physical prior law, and the reliance on asymptotic theory (mathematically astronomical sample sizes rather than actual ones of astronomy!) does not support the usefulness of Bayesian foundations in this setting. A more satisfactory answer would have been, 'in such settings where we are dealing with new phenomena and have few observations, the Bayesian approach will not accomplish any more than others.' Of course some problems are hard, why not admit it? Incidentally, this asymptotic theory that Bayesians rely upon in taking comfort in the use of diffuse prior laws, was initiated by v. Mises (1919) in the Bernoulli case almost sixty years ago, was later extended by Kolmogorov, and in more recent years has been treated in great generality by Wolfowitz, Le Cam, and others. A principal conclusion is that, for sufficiently large sample sizes, *many* procedures have good properties; but how large a sample is needed for a Bayes procedure to have such properties *depends on the prior law employed*, so that we can take little comfort in what we would get as an estimate from using a prior law that reflects the mentioned million-to-one odds, if the true distance were in fact near to or less than 5×10^9 light years.

Why is the Bayesian approach so prevalent in business schools (at least in the U.S.)? Businessmen must often make immediate terminal decisions (actions). Sometimes they do encounter settings that fit into a background that produces a sensible physical prior law (or approximation thereof). Where they don't, they must still act, and the Bayesian approach provides quick and simple recipes for action. It is an old political proverb, that American businessmen mind an indecisive government more than they mind one whose policies are harsh but known. Presumably this attitude is reflected inwardly in the importance businessmen give to decisiveness even at the expense of making some incorrect decisions. We, and they, *will never know how much more or less they lose* by being subjective Bayesians than they would by using another approach. Also, whatever success the introduction of the Bayesian approach has led to in the business world has in part been due to the coincident introduction of quantitative methods of assessing various possible prospects, that would also benefit other methodologies.

Another occasional contributor to Bayesian success is actually a departure from the strict Bayesian scheme. It is astounding how many practicing Bayesians who argue articulately about the wisdom of incorporating into a prior law one's feelings about the relative importance of different parts of the risk function (perhaps not an outrageous *first step* in trying to select a procedure), then act blindly in using the Bayes procedure that results, without ever examining the risk function (operating characteristic) of that procedure and comparing it with those of other procedures. It is easy to give examples in which the Bayes procedure T_0 relative to a given prior law P_0 has large (even unbounded) risk on a sizable set B_0 of small P_0 probability, whereas another procedure T_1 has much smaller risk than T_0 on B_0 and only slightly larger risk on the complement of B_0; T_1 is not Bayes, having slightly larger integrated risk with respect to P_0 than T_0 does. (This, somewhat loosely, is described as a *subadmissibility* phenomenon; the precise definition is given in terms of sequences of procedures and settings.) You may feel that I don't know the right practicing Bayesians, but the overwhelming majority of those I have questioned had never thought of this; many are thereafter quick to realize that P_0 was only an approximate input and T_1 is more sensible to use than T_0. I am told by John Pratt that sensitivity analysis is the name of the game some Bayesians (largely complementary to my sample) go through in concluding that they shouldn't use T_0. It is comforting that this occurs sometimes; but it is disturbing to learn, continually, of much-heralded Bayesian interactive computing packages that query and reexamine the customer's prior law, but that typically then grind out the posterior law or some form of decision without ever having looked at the operating characteristic of the Bayes procedure. This last is but one shortcoming of routine, increasingly mechanized Bayesian analysis; at least as serious is the fact that there are so many customers who do not understand fully the consequences of what they are doing in using such routines, to which they have been attracted by the oversimplified promise of such ease and comfort in reaching a quick decision to which such beautiful names are attached. Such mechanical routines also discourage the possibility of deserved distrust in the model and of insight the data might yield in the form of a possible new natural law to be verified in subsequent experiments.

The modification of strict Bayesian analysis, that compares the risk function of T_0 with that of other procedures, is not in conformity with Bayesian foundations. The scheme of picking a single subjective P_0 and looking only at the risk functions of nearby competitors of T_0 to see whether one of them is more sensible, also seems too rigid; but, in any event, it bears as much relationship to a routine that might be followed by a non-Bayesian decision theorist, as it does to the strict dictates of the axiom systems for rational behavior. (The spirit of this modification of the Bayes routine has been formalized in various criteria; for example, one of these chooses a minimax procedure among the procedures that are Bayes relative to some law in a specified set of prior laws that replaces P_0. This also seems too rigid; for example, there is still the possibility of occurrence of the subadmissibility phenomenon. Again, it is a matter of statistics being too complex a subject for any simple prescription – even one that requires the space of a book or two – to yield a satisfactory choice of procedure in all settings.)

Thus, strict subjective Bayesianism does not seem useful, and the sensible modifications some Bayesians have suggested remove the resulting methodology from being an illustration of the usefulness of Bayes *foundations*.

5. OTHER COMMENTS

5.1. *More on Bayes*

The subjective Bayesian approach has other difficulties than those mentioned earlier. For example, as Wolfowitz (1962) pointed out, the rational behavior axiom that assumes comparability of all pairs of choices is more severe than one that is much more straightforward (although phrased in terms of less primitive choices), that the statistician must be able to select a procedure he prefers from the class of all those available. Moreover, the crucial Savage-Lindley conclusion that rational statisticians behave like Bayesians is of limited impact for several reasons. On the one hand, if we restrict consideration to a fixed setting, the conclusion amounts only to that mentioned earlier, that the admissible and Bayes procedures coincide in suitably regular settings – not very startling, once understood. On the other hand, as Birnbaum (1977) discusses elsewhere in this volume, the basic *assumption* of preservation of indifference sets under probabilistic mixing is highly debatable, as is therefore the conclusion that, as the experiment varies but the states of nature remain fixed, a rational person should use the same prior law. This means that, theoretically, it is not as conclusive as we have been told by Bayesian theorists, that we should be able to select our procedure on the basis only of a subjective prior law and without looking at the set of available risk functions in the setting at hand except through their integrated risk relative to that prior law; and, practically, an examination of the variety of possible geometric configurations of risk functions in different settings, which may exhibit such phenomena as that of subadmissibility alluded to earlier, adds to one's misgivings about the practical virtues of the Bayesian prescription.

In assessing the practical impact of the foundations of rational behavior, we must also face the reality that the set of possible states of nature a statistician encounters generally *does* change as he goes from setting to setting involving different phenomena, so that, even if we accept the axioms and ignore Birnbaum's criticism, the conclusion of rational behavior can then only be that we act like Bayesians in each setting *with a different subjective prior law in each setting*. Except when a relationship among those prior laws on different sets in different settings can be established, the rational behavior axioms imply only that (under regularity conditions) the statistician uses *some* admissible procedure in each setting.

Among the Bayesians I know there are some who are followers of Fisher in using randomization of the experiment but followers of Lindley in conditioning (after the randomization is performed) on the experiment actually used. This is inconsistent with the axioms, as many Bayesian theorists may agree: a Bayesian can only think of randomizing among the *best* (Bayes) experiments, but then has no reason to randomize at all. Perhaps this is a reflection on my sampling of Bayesians. Perhaps, too, it is another illustration of the tendency mentioned earlier, of the Bayesian rationale, to invite an unwarranted satisfaction with a simplistic approach that yields quick answers, often without any understanding of the consequences in terms of actual probabilities of error or expected loss. The increasing use of the approach in applications by unthinking practitioners is a more serious worry than any axiomatic inadequacy *per se*.

5.2. *Probability and Betting*

A Bayesian colleague told me he cannot understand how I can look at risk-functions and choose among them without thinking in terms of 'probabilities' of the various states of nature. This reflects not only his understandable tendency to see the choices of others through his own choice mechanism, but also a logical fallacy in his perception of psychology: even assuming Bayesian axioms, the fact that one can be viewed as acting *just as though he is a Bayesian* relative to some prior law (a description of mathematical psychology), does not mean he *does* choose a decision in that way with a conscious injection of his prior law.

A related misunderstanding occurs in connection with the odds and bets that are often a hallmark of Bayesian foundational discussions. These discussions arise, for example, in the Bayesian dissatisfaction with the Neyman-Pearson confidence coefficient, that non-Bayesian frequentists refuse to regard (after the experiment) as a 'probability' of correctly covering the true parameter value, instead interpreting its meaning in terms of the expectation viewpoint before the experiment is conducted, or in terms of the actual frequency of successes in many (not necessarily identical) experiments as justified by the law of large numbers. To make the controversy clearer, let me describe it in termsof the single flip of a fair coin – not a problem of inference about the true probability model, which is known, but, rather, one of guessing the outcome of the flip. Before the coin is flipped we would all speak of probability $\frac{1}{2}$ of the coin coming up 'heads'. Some subjectivists have asked, 'if the coin is flipped but the outcome is hidden from you, (a) wouldn't you still bet as though heads is a 50–50 chance, and, if so, (b) would you still insist on using the familiar (non-Bayesian frequentist) language that, once the coin is flipped, the probability is no longer $\frac{1}{2}$ that the coin will come up "heads", but is either 0 or 1 (you know not which) that it now *is* heads?' My answer to both questions is yes. My betting habits may be the same as they were *before* the coin was flipped, but that does not mean I must think the situations are the same or use the same language for them. There is no chance mechanism still to be used after the coin has been flipped, so it is not a (classical) probability situation. The fact that di Finetti has integrated personal and classical physical probabilities into one system, the fact that Bayesians therefore use 'probability' in the resulting broader sense, does not force me to do so, or even to believe in personal probability, even though my betting can be described by Bayesians in such terms. (The language of probability is also used in other contexts where there is no probability mechanism, such as in deterministic settings of mechanics with many particles; but there, in statistical mechanics, the language and probability calculus are being used, knowingly, to describe a probability model that turns out to yield good approximations of certain properties of the deterministic setting which are much more difficult to compute exactly. This is quite unlike the spirit of the Bayesian discussion.)

Incidentally, another common misunderstanding is present in the interpretation of frequentist probability (e.g., that a confidence interval will by chance cover the true parameter value) and posterior probability relative to a physical prior law and given that the observation $X = 2.9$. The former refers to a chance experiment still to be conducted, the latter does not, and this is not altered by one's similarity of betting behavior in the two instances. For a frequentist, the posterior law relative to the yet-to-be-observed X is the random quantity described in terms of an experiment yet to be conducted, and whose behavior in terms of the law of large numbers justifies the way we use it when $X = 2.9$. Of course, subjectivists require no such distinction.

We shall not enter here into the question of whether it is reasonable to develop a theory of inference about scientific truth that rests on willingness to make certain bets; faced with such objections, some Bayesians have attempted to rework some of their developments in other terms. See, e.g., Buehler (1976).

5.3. *What to do instead*

My only 'foundations', other than my reliance on Neyman's frequentist point of view, are qualitative: that, as stated more than once earlier, statistics is too complex to be codified in terms of a simple prescription that is a panacea for all settings, and that one must look as carefully as possible at a variety of possible procedures, both in the difficult settings in which one cannot easily exhibit all operating characteristics, and also in ones that are simple superficially, in which there is an approximate physical prior law with which to make a preliminary selection. Few non-Bayesian frequentists (if any) now have agreement with the straw men exhibited by the Bayesians, and who supposedly use tests of level 0.05 or 0.01 on every possible occasion without further examination of the possible power functions. (No doubt these Bayesians have met some such flesh-and-blood statisticians, as have I, and just as I have met the unthinking Bayesians they will regard as *my* straw men!)

There are a few simple settings, no doubt assumed to be valid more often than is warranted, in which some of the 'standard' procedures of statistics have been justified theoretically (for example, as being almost minimax, close to Bayes relative to a variety of prior laws, and without a sizable subadmissibility defect). Although theoretical statisticians work on such simple, often unrealistic models, it is a mistake to regard their conclusions, as some data analysts have, as practical recommendations; rather, they are guides to understanding and clues, obtained in tractable models, as to what one might investigate as possible good procedures in the more complex real-world settings where exact calculations are impossible. Whatever may be the eventual value of the current expanding trend to make simulation robustness studies of various procedures, it is doubtful that this development, initiated by Tukey and other data analysts, would have taken its present direction without the theoretical studies of Huber and others in simple cases; and it is difficult to imagine that it could have come out of elegant Bayesian computations with convenient conjugate prior laws.

Perhaps some day someone will prove me wrong and provide the satisfactory codification we do not now have. Until that new Copernicus arrives, I can only say to the personal probability advocates that the world of scientific inference is no more egocentric than it is geocentric.

Cornell University

ACKNOWLEDGMENT

Alan Birnbaum suggested the writing of this paper and contributed stimulation to its views in his criticism of Kiefer (1977), as did the criticism of George Barnard, Bob Buehler, and

John Pratt. Thanks are also due to Bob Hogg for inviting the talk that produced an earlier version, and to Mel Novick, who, in trying to convince me that my views were wrong, had the opposite effect.

REFERENCES

Birnbaum, A.: 1977, 'The Neyman-Pearson Theory as Decision Theory, and as Inference Theory; with a Criticism of the Lindley-Savage Argument for Bayesian Theory', this volume, pp. 19–49.

Buehler, R.: 1976, 'Coherent Preferences', *Ann. Math. Statist.* **4**, 1051–1064.

Kiefer, J.: 1977 (discussion), 'Conditional Confidence Statements and Estimated Confidence Coefficients', *J. Amer. Statist. Assn.*

LeCam, L.: 1968, in D. G. Watts (ed.), *The Future of Statistics*, Academic Press, New York, p. 143.

Savage, L. J.: 1954, *The Foundations of Statistics*, John Wiley, New York.

Tukey, J.: 1960, 'Conclusions versus Decision', *Technometrics* **2**, 423–433.

v. Mises, R.: 1919, 'Fundamentalsätze der Wahrscheinlichkeitsrechnung', *Math. Zeit.* **4**. 1–97.

Wolfowitz, J.: 1962, 'Bayesian Inference and Axioms of Consistent Decision', *Econometrica* **30**, 470–479.