PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 15 (1995), pp. 29–36

NEYMAN'S STATISTICAL PHILOSOPHY

BY

E. L. LEHMANN (BERKELEY, CALIFORNIA)

1. Inductive behavior vs. inductive inference: Neyman and Fisher. A dramatics meeting of the Royal Statistical Society occurred on December 18, 1934, when R. A. Fisher – already at the height of his fame although not universally popular – presented a paper on *The logic of inference*. He stated that it was concerned with the problem of drawing inferences from the particular to the general or, in statistical language, from the sample to the population. He pointed out that deductive arguments based on probability were not adequate for such inferences, "for reasoning of a genuinely inductive kind," unless one used the theory of inverse probability [i.e. Bayesian inference] thereby "making it a deduction from the general to the particular." He then advanced the ambitious claim that truly inductive reasoning could be based on his concept of likelihood, which appeared "to take its place as a measure of rational belief when we are reasoning from the sample to the population," and proceeded to develop some of its properties.

The paper was attacked ferociously in the discussion following Fisher's presentation, and further objections were raised in later written contributions to the discussion. The last of these was by Neyman who started out in a highly complimentary vein. He then went on: "The psychology of some readers lead them to develop a theory already started along the way indicated. They probably think: 'What an interesting problem is raised! how could I develop it further?'... I personally seem to have another kind of psychology and can't help thinking: 'What an interesting way of asking and answering questions, but can't I do it differently?'"

Neyman points out that the properties Fisher establishes in his paper such as the asymptotic normality and efficiency of the maximum likelihood estimator are obtained on the basis of probability calculations and "not because we believe blindly in some magic property of the [likelihood] function."

He then turns to his principal theme, "whether it is possible to construct a theory of mathematical statistics independent of the conception of likelihood," and after some discussion gives his answer: "the basic conception is the conception of frequency of errors in judgement..." "I grant," he says, "it is only a principle. If the principle is accepted, then we have to deal with mathematical problems of finding a criterion... such as to minimize the actual probability of an error in judgement. The complex of results in this direction may be considered as a system of mathematical statistics alternative to that of Professor Fisher, and entirely based on the classical theory of probability."

The idea that a satisfactory theory of statistical inference can be based on the frequency of errors resulting from the use of an inference procedure had already been advanced in Neyman and Pearson (1933). There they developed a theory of hypothesis testing involving two possible errors: false acceptance and false rejection of the hypothesis. This led them to a more general proposal: "Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behavior with regard to them, in following which we insure that, in the long run of experience we shall not be too often wrong." And in subsequent passages they point out that a test constitutes such a 'rule of behavior' namely by telling us when to reject the hypothesis and when to accept it.

Neyman later developed this idea further, labeling his deductive approach 'inductive behavior' in distinction to Fisher's 'inductive inference.' Pearson did not follow his collaborator into this philosophical arena. Responding to an attack by Fisher (1955) on the point of view "expressed in numerous papers by Neyman, Pearson, Wald and Bartlett," Pearson concluded his paper (E. S. Pearson, 1955) with: "Professor Fisher's final criticism concerns the use of the term 'inductive behavior,' this is Professor Neyman's field rather than mine."

For Neyman the behavioristic point of view is central to his thinking about the foundations of statistics and his efforts to make statistics into a well-defined independent subject. It appears in the title of several of his papers devoted to this project: 'raisonnement inductif ou comportement inductif' (1947); 'the problem of inductive inference' (1955); 'inductive behavior' as a basic concept of philosophy of science (1957); 'behavioristic points of view on mathematical statistics' (1966); 'foundations of behavioral statistics' (1971). It is also discussed in some of Neyman's other writings, for example in his papers: *Foundations of the general theory of statistical estimation* (1951) and *Frequentist probability and frequentist statistics* (1977). Finally, the topic is discussed in his *Lectures and Conferences on Mathematical Statistics* of 1952, in which he states (p. 210) that "the concept of 'inductive behavior' is discussed in some detail in a book [Neyman (1950)], in which it is treated as the motivational basis of the whole theory of statistics."

The term 'inductive behavior,' or rather its French version 'comportement inductif' occurs for the first time in Neyman (1938), in which he distinguishes inductive behavior from inductive reasoning. Although there is no direct reference to Fisher in the relevant passage, it is clearly a reply to Fisher's presentation, and an elaboration of Neyman's earlier discussion of it. The difference between these two points of view was thought by both Fisher and Neyman to be at the heart of their lifelong dispute.

The possibility and nature of induction is a classical problem of philosophy, and it seems of interest to inquire into the reception of the ideas of Neyman and Fisher by philosophers of science. Of course, it takes time for one discipline to absorb, acknowledge, or even become aware of the concepts and ideas from another. A bridge from the Neyman–Pearson theory of hypothesis testing and Neyman's theory of confidence intervals to the philosophy of science was built by two highly influential books, which examined these theories from a philosophical point of view: Braithwaite's *Scientific Explanation* (1953) and Hacking's *Logic of Statistical Inference* (1965). These were followed by a number of other monographs and conference proceedings devoted entirely or at least in part to philosophical interpretations of various theories of statistical inference, among them Gillies (1973), Stegmüller (1973), Kyburg (1974), Harper and Hooker, Eds. (1974), Maxwell and Anderson, Eds. (1975), Rosenkrantz (1977), Asquith and Kyburg, Eds. (1979), Seidenfeld (1979), Fetzer (1981), and Howson and Urbach (1989).

In its treatment of induction, this literature mirrors the Fisher–Neyman conflict, dividing the authors into 'inductionists,' who discuss how inductive inference is or should be carried out, and 'deductionists,' who believe that only deductive arguments are legitimate.

The former group subdivides further according to the medium through which induction is to be effected: Fisher's likelihood, his fiducial method, the posterior Bayesian probabilities of Jeffreys, De Finetti or Savage, or possibly some new approach of their own. For these philosophers, Neyman's ideas are either irrelevant or serve only as a foil against which to set forth their preferred alternative approach.

2. Deductivism: Neyman and Popper. The deductivists also speak with more than one voice. Here Neyman has a rival in the person of Karl Popper. The views of these two men with respect to induction have important similarities, as can be seen from Popper's first book of 1934, which appeared in an English translation in 1959 under the title *The Logic of Scientific Discovery*.

After a brief introductory paragraph, Chapter 1 begins with a discussion of induction: "According to a widely accepted view – to be opposed in this book – the empirical sciences can be characterized by the fact that they use 'inductive methods,' as they are called." A few pages later, Popper contrasts this inductivist view with his own approach: "The theory to be developed in the following pages stands directly opposed to all attempts to operate with the ideas of inductive logic. It might be described as the *deductive method of testing*, or as the view that a hypothesis can only be empirically tested – and only after it has been advanced." Actually, at the time Popper was concerned only with deterministic situations. It was only in later editions that he extended his considerations to a stochastic setting. And in his autobiography of 1976 he refers (p. 79) to a "nonexistent 'method of induction,' of which nobody has ever given a sensible description." One could think that one hears Neyman speaking. Not only the terms they use but also the times at which they first put forth their ideas are very close. The appearance of Popper's book in 1934([1]) is sandwiched between the first mention in Neyman–Pearson (1933) of statistical tests as 'rules of behavior,' and Neyman's proposal to replace Fisher's inductive approach by a purely deductive one in his 1935 discussion of Fisher's paper to the Royal Statistical Society.

Since the difficulties with induction had been known for a long time, it is remarkable that independently and nearly simultaneously Neyman and Popper found a revolutionary way to finesse the issue by replacing inductive reasoning with a deductive process of hypothesis testing. They then proceeded to develop this shared central idea in different directions, with Popper pursuing it philosophically while Neyman (in his joint work with Pearson) showed how to implement it in scientific practice.

The difference in these two developments is illustrated by their respective attitudes toward the outcome of a test. For Popper, the meaning of this outcome is central to his approach. Rejection of hypotheses, which he calls *falsification*, is for him the principal tool of the scientific method. On the other hand, he never permits hypotheses to be accepted. At best, they receive some support from the test, which he calls *corroboration*.

In contrast, Neyman and Pearson use the terms *acceptance* and *rejection* rather casually, with Pearson later (1955) admitting that some of their "wording may have been chosen inadequately." And Neyman strongly denies that they have any cognitive significance. According to him (Neyman, 1957), the outcome of a test is "an act of will or decision to take a particular action,"... and again in the next paragraph: "These processes are certainly not any sort of 'reasoning'...; they are acts of will."

He states his position even more strongly in his text (Neyman, 1950). A test is a choice between two actions A and B which is "interpreted as the *adoption* or the *acceptance* of one of the hypotheses H or $\bar{H}$ and the *rejection* of the other." He goes on to warn the reader: "The terms 'accepting' and 'rejecting' a hypothesis are very convenient and are well established. It is important, however, to keep their exact meaning in mind and to discard various additional implications which may be suggested by intuition. Thus, to accept a hypothesis H means only to decide to take action A rather than action B. This does not mean that we necessarily believe that the hypothesis is true. Also, if the application of a rule of inductive behavior "rejects H, this means only that the rule prescribes action B and does not imply that we believe that H is false."

([1]) It appeared in the fall of 1934 but bore the imprint of 1935.

However, statisticians tend not to analyze terms as carefully as philosophers do. The line between actions, decisions and acts of will on the one hand, and assertions, statements and judgements on the other, is often rather thin. So it should not come as a surprise that Neyman's own applied writing sometimes crosses this line which he has drawn so firmly. In a paper on the expansion of galaxies, for example (Lovasich, Mayall, Neyman, and Scott, 1961), a hypothesis of stability is being tested. Among the conclusions we read: "If we admit [certain assumptions], then, with outliers included, there is strong evidence that at least some of the clusters (and this would include Coma) are unstable. In this case, the chance of error in asserting instability of at least some of the clusters is less than one in a hundred. If the outliers are excluded, no such evidence exists; the chance of error in asserting instability may be as large as three in four." Similar language is found, for example, in Neyman and Scott (1961).

In denying that the outcome of a test has any cognitive impact, Neyman thus seems to have carried his argument too far: it contradicts the reality even of his own statistical practice. On the other hand, scientific experiments typically are followed by decisions – if only whether or not to make a public statement. A test then offers a substitute for the unattainable, but so desperately desired, ideal of induction. On the basis of the incomplete information provided by the observations, it furnishes a guide to which action to take and thus – Neyman's evocative phrase – to appropriate 'inductive behavior.'

Despite the close connection of their ideas, Neyman and Popper – in stark contrast to the lifelong feuding of Neyman and Fisher – assiduously ignored each other. I can find no reference to Popper in Neyman's writings, and the few times Popper mentions Neyman, it is as translator of von Mises or relates to Neyman's frequentists interpretation of probability. This can, of course, partly be explained by the fact that the two men worked in different disciplines. It is nevertheless surprising.

3. The impact of Neyman's work. Neyman's ideas on testing and estimation channelled statistical theory into new directions which culminated in Wald's general decision theory. How strongly Wald was influenced by Neyman's behavioristic point of view can be seen by the explanation Wald (1950) offers when formulating his general decision problem: "The adoption of a particular decision function by the experimenter may be termed 'inductive behavior,' since it determines uniquely the procedure for carrying out the experiment and for making a terminal decision. Thus, the above decision problem may be called the problem of inductive behavior." (In a footnote he attributes the term to Neyman (1938).)

But Neyman thought of his behavioristic formulation (as Fisher did of his likelihood approach) as an important contribution not only to statistics but also to the philosophical problem of induction. It is of interest to consider the reaction of philosophers of science to this claim. In assessing the reception of Neyman's ideas by the philosophical community, it is necessary to consider more generally the influence of the whole of the modern statistical enterprise based on the work of Fisher, Neyman, Pearson and Wald, and also of the Bayesians, particularly Jeffreys, De Finetti and Savage. This body of work has affected the relevant parts of the philosophy of science in two important ways: By changing the emphasis from a deterministic to a stochastic setting([2]), and by developing methods that permit the implementation of the abstract formulations of the philosophy of science.

The approaches of Fisher, Neyman–Pearson and the Bayesians each have their proponents among philosophers of science. But even the expositions of inductivists and other opponents of Neyman typically find room for a critical discussion of the Neyman–Pearson theory. Thus, Gillies (1973) explains that although the Neyman–Pearson theory "in certain circumstances gives results which directly contradict our own approach," nevertheless "an examination of the Neyman–Pearson theory is not for us an option but a most pressing necessity" because "the Neyman–Pearson theory is still the generally accepted theory of testing statistical hypotheses."

Similarly, Kyburg (1974) justifies devoting a chapter to the "classical" theory of Neyman and Pearson by reminding the reader: "By the 1940's, the view, primarily associated with the name of Neyman that the fundamental form of statistical inference was the choice between statistical hypotheses, had come to dominate other views in the English speaking world. It merely dominated other views among statisticians; but it utterly overwhelmed other views among those whose interest in statistics was primarily practical."

These and the other authors listed in Section 2 discuss Neyman's approach not because of their liking for it but because of its prominence. However, there are also philosophers who strongly support Neyman's point of view. Prominent among these is Giere. In his essay (Giere, 1975), for example, he states that the approaches of Fisher, Neyman–Pearson, Neyman–Wald, and of "subjective Bayesians like Savage" have developed "prototype inductive logics which can deal with interesting scientific cases." — "Of these possibilities," he says, "the Neyman–Pearson–Wald tradition seems to me most promising." In a later section he writes: "The logic adopted here is a version of the theory of statistical hypothesis testing developed by Neyman and Pearson... This choice is based partly on my personal conviction that the Neyman–Pearson viewpoint is basically correct. But it is also justified by the tremendous influence this work has had for nearly half a century. The Neyman–Pearson approach must count as one of the three or four major candidates for a comprehensive inductive logic."

([2]) It has been pointed out to me by Brian Skyrms that historically much of the impetus for this change came from physics.

Another philosopher who strongly supports Neyman's ideas is Hacking who writes in 1980: "In my opinion there have been two important contributions to the study of rationality in this century, both negative. One is by Jerzy Neyman, the other by Imre Lakatos." And later on: "if you have firm beliefs about the future, then you can have detailed beliefs about the future, and you can engage in inductive behaviour such as avoiding betting. That is what Neyman teaches, and of course, I endorse it. There is no such thing as inductive inference, nor a logic of statistical inference or whatever..."

The supporters of Neyman, however, are in the minority. As Giere (1975) points out "in recent years, the most active area of interest for philosophers of science has been the subjective Bayesian approach to inference and decision making. Whatever the causes and reasons for this interest, it is a fact that, of all philosophers seriously concerned with the foundations of probability and statistics, more prefer some form of subjective account than any other approach."

Had he known about it, Neyman would have been interested in the attention philosophers have paid to his ideas, and in the importance statistics has acquired in the discussion of induction. He would no doubt have been disappointed to learn that his approach has not found more acceptance among philosophers. I believe he would have reacted the way he usually did when hearing unwelcome news. He would have shrugged his shoulders and said, "too bad."

REFERENCES

Asquith, P., and Kyburg, H., Eds. (1979), *Current Research in Philosophy of Science*, Philosophy of Science Association, East Lansing.

Braithwaite, R. B. (1953), *Scientific Explanation*, Cambridge University Press.

Fetzer, J. (1981), *Scientific Knowledge*, Reidel, Dordrecht.

Fisher, R. A. (1935), *The logic of inference* (with discussion), J. Roy. Statist. Soc.

Fisher, R. A. (1935), *The Design of Experiments*, Oliver and Boyd, Edinburgh.

Fisher, R. A. (1955), *Statistical methods and scientific induction*, J. Roy. Statist. Soc. (B) 17, pp. 69–78.

Giere, R. (1975), *The epistemological roots of scientific knowledge*, in: *Induction, Probability and Confirmation*, W. L. Harper and C. A. Hooker (Eds.), University of Minnesota Press, Minneapolis.

Gillies, D. A. (1973), *An Objective Theory of Probability*, Methuen, London.

Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge University Press.

Hacking, I. (1980), in: *Scientific Discovery, Logic and Rationality*, T. Nickles (Ed.), Reidel, Dordrecht.

Harper, W. L., and Hooker, C. A. (Eds.) (1974), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. 2: *Foundations and Philosophy of Statistical Inference*, Reidel, Dordrecht.

Howson, C., and Urbach, P. (1989), *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle.

Kyburg, H. (1974), *The Logical Foundations of Statistical Inference*, Reidel, Dordrecht.
Lovasich, J. L., Mayall, N. U., Neyman, J., and Scott, E. L. (1961), *The Expansion of Clusters of Galaxies*, in: Proc. 4th Berkeley Symp. Math. Statist. Probability, Vol. 3, University of California Press, pp. 187–227.
Maxwell, G., and Anderson, R. M., Jr. (1975), *Induction, Probability, and Confirmation*, University of Minnesota Press, Minneapolis.
Neyman, J. (1938), *L'estimation statistique traitée comme un problème classique de probabilité*, Actualités Sci. Indust., No. 739, pp. 25–57.
Neyman, J. (1947), *Raisonnement inductif ou comportement inductif*, Proc. Internat. Statist. Conference 3, pp. 423–433.
Neyman, J. (1950), *First Course in Probability and Statistics*, Holt, New York.
Neyman, J. (1951), *Foundations of the general theory of statistical estimation*, Actualités Sci. Indust., No. 1146, pp. 83–95.
Neyman, J. (1952), *Lectures and Conferences on Mathematical Statistics*, 2nd ed., Graduate School, U.S. Department of Agriculture, Washington, D. C.
Neyman, J. (1955), *The problem of inductive inference*, Commun. Pure Appl. Math. 8, pp. 13–45.
Neyman, J. (1957), *'Inductive behavior' as a basic concept of philosophy of science*, Rev. Inst. Internat. Statist. 25, pp. 7–22.
Neyman, J. (1966), *Behavioristic Points of View on Mathematical Statistics in Political Economy and Econometrics — Essays in Honour of Oskar Lange*, PWN — Polish Scientific Publishers, Warszawa.
Neyman, J. (1971), *Foundations of behavioral statistics*, in: *Foundations of Statistical Inference*, V. P. Godambe and D. A. Sprott (Eds.), Holt, Rinehart and Winston, Toronto.
Neyman, J. (1977), *Frequentist probability and frequentist statistics*, Synthèse 36, pp. 97–131.
Neyman, J., and Pearson, E. S. (1933), *On the problem of the most efficient tests of statistical hypotheses*, Philos. Trans. Roy. Statist. Soc. Sec. A. 231, pp. 289–337.
Neyman, J., and Scott, E. L. (1961), *Magnitude — redshift relation of galaxies in clusters in the presence of instability and absorption*, Astron. J. 66, pp. 581–589.
Pearson, E. S. (1955), *Statistical concepts in their relation to reality*, J. Roy. Statist. Soc. (B) 17, pp. 204–207.
Popper, K. (1935), *Die Logik der Forschung*, Springer, Wien. English translation: *The Logic of Scientific Discovery* (1959), Basic Books, New York.
Popper, K. (1976), *Unended Quest*, Open Court, La Salle.
Rosenkrantz, R. D. (1977), *Inference, Method and Decision*, Reidel, Dordrecht.
Seidenfeld, T. (1979), *Philosophical Problems of Statistical Inference*, Reidel, Dordrecht.
Stegmüller, W. (1973), *Personelle und statistische Wahrscheinlichkeit*, Vol. 2: *Statistisches Schliessen, statistische Begründung, statistische Analyse*, Springer, Berlin.
Wald, A. (1950), *Statistical Decision Functions*, Wiley, New York.

Department of Statistics
University of California
Berkeley, CA 94720, U.S.A.

Received on 8.3.1994