

**Model Specification: The Views of Fisher
and Neyman, and Later Developments***

By

E.L. Lehmann**

University of California, Berkeley

Technical Report No. 226

October 1989

* A modified version of the 1988 Fisher Lecture delivered Aug. 24, 1988
at the Joint Statistical Meetings in New Orleans.

** Research supported by NSF Grant No. DMS-8908670.

Department of Statistics
University of California
Berkeley, California

Where do probability models come from? To judge by the resounding silence over this question on the part of most statisticians, it seems highly embarrassing. In general, the theoretician is happy to accept that his abstract probability triple $(\Omega, \mathcal{A}, \mathbf{P})$ was found under a gooseberry bush, while the applied statistician's model "just grewed".

A.P. Dawid (1982)

Abstract

Since Fisher's formulation in 1922 of a framework for theoretical statistics, statistical theory has been concerned primarily with the derivation and properties of suitable statistical procedures on the basis of an assumed statistical model (including sensitivity to deviations from this model). Until relatively recently, the theory has paid little attention to the question of how such a model should be chosen. In the present paper, we consider first what Fisher and Neyman had to say about this problem and in Section 2 survey some contributions statistical theory has made to it. In Section 3 we study a distinction between two types of models (empirical and explanatory) which has been discussed by Neyman, Box, and others. A concluding section considers some lines of further work.

**Model Specification: The Views of Fisher
and Neyman, and Later Developments***

By E.L. Lehmann
University of California, Berkeley**

***A modified version of the 1988 Fisher Lecture delivered on Aug. 24, 1988 at the
Joint Statistical Meetings in New Orleans.**

****Research supported by NSF Grant No. DMS-8908670.**

1. The views of Fisher and Neyman

A general framework for theoretical statistics was proposed by Fisher (1922) in his fundamental paper "On the mathematical foundations of theoretical statistics." In it, Fisher defines the principal task of statistics to be "the reduction of data," and states that "this object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a sample. The law of distribution of this hypothetical population is specified by relatively few parameters,..." In other words, Fisher states that the first step is the construction of a low-dimensional parametric model.

On this basis, Fisher divides the problem of statistics into three types:

(1) Problems of Specification.

(That is, the problem of specifying the parametric model).

(2) Problems of Estimation.

(This terminology for the second stage is explained by the fact that Fisher in this paper is concerned only with point estimation. More generally, this type of problem could be described as the derivation of a suitable statistical procedure.).

(3) Problems of Distribution.

(This refers to the distribution of the estimator derived in (2). A more general description of this last stage would be assessing the performance of the procedure obtained in (2).)

The present paper is concerned only with the first of these stages, and it is with considerable interest that one wonders what Fisher has to say about it. Disappointingly, his discussion is confined to a single paragraph which is dominated by the first sentence:

"As regards problems of specification, these are entirely a matter for the practical statistician,..."

This statement constitutes Fisher's answer to the question raised in the title of the present paper and the quotation at its beginning. Fisher's statement implies that in his view there can be no theory of modeling, no general modeling strategies, but that instead each problem must be considered entirely on its own merits. He does not appear to have revised this opinion later; the index to the 5-volume collection of his papers (published by the University of Adelaide) has only one entry under "Specification, problems of" — the 1922 statement cited in the preceding paragraph.

Actually, following this uncompromisingly negative statement, Fisher unbends slightly and offers two general suggestions concerning model building:

(a) “We must confine ourselves to those forms which we know how to handle”,

and

(b) “More or less elaborate forms will be suitable according to the volume of the data”.

We shall return to both of these suggestions later.

To Neyman, not only the practice but also the theory of modeling was a central concern. In 1959 he introduced a course in the subject into the Berkeley curriculum and taught it with few interruptions until the time of his death in 1981. We shall here discuss three comments of his on modeling.

1. Models of complex phenomena are constructed by combining simple building blocks which, “partly through experience and partly through imagination, appear to us familiar and, therefore, simple.”

Although not making exactly the same point, this comment is somewhat reminiscent of Fisher’s suggestion that we should restrict attention to models we know how to handle.

2. An important contribution to the theory of modeling is Neyman’s distinction between two types of model: “interpolatory formulae” on the one hand and “explanatory models” on the other. The latter try to provide an explanation of the mechanism underlying the observed phenomena; Mendelian inheritance was Neyman’s favorite example. On the other hand an interpolatory formula is based on a convenient and flexible family of distributions or models given a priori, for example the Pearson curves, one of which is selected as providing the best fit to the data. (Neyman 1939). The same distinction is emphasized in the writings of George Box, who uses the terms “empirical model” and “theoretical” or “mechanistic” model for the two concepts (mechanistic since it identifies the underlying mechanism).

3. The last comment of Neyman’s we mention here is that to develop a “genuine explanatory theory” requires substantial knowledge of the scientific background of the problem*. This requirement is agreed on by all serious statisticians but it constitutes of course an obstacle to any general theory of modeling, and is likely a principal reason for Fisher’s negative feeling concerning the possibility of such a theory.

* The same general idea is expressed by John Stuart Mill who writes (p.344 of the tenth edition of his *System of Logic*, Vol. 1 (1879)):

“The guesses which serve to give mental unity and wholeness to a chaos of scattered particulars, are accidents which rarely occur to any minds but those abounding in knowledge and disciplined in intellectual combinations.”

2. Where do models come from?

Several statisticians with extensive applied experience have objected to this question, stating that they know exactly where their models come from. However, it seems difficult to find explicit statements of how models are obtained, which raises the question in a different form: Is applied statistics, and more particularly model building, an art, with each new case having to be treated from scratch*, completely on its own merits, or does theory have a contribution to make to this process?

(i) **A reservoir of models.** One crucial (although somewhat indirect) contribution of theory is indicated by Fisher's and Neyman's references to "those forms which we know how to handle", and to building blocks which "appear to us familiar and, therefore, simple." These references presuppose the existence of a reservoir of models which are well understood and whose properties we know. Probability theory and statistics have provided us with a rich collection of such models. One need only think of the various families of univariate and multivariate distributions, of the different kinds of stochastic processes, of linear and generalized linear models, and so on. The list seems inexhaustible and furnishes the modeler with an indispensable tool.

One aspect of models that is of particular importance to realistic modeling is the way they are described or characterized. (For some references to the literature on characterization of distributions, see for example Galambos (1982).)

Example 2.1. As an illustration consider the model that is traditionally described by

$$(2.1) \quad X_1, \dots, X_n \text{ are i.i.d. with normal distribution } N(0, \sigma^2).$$

If asked to justify the normality assumption, a statistician might refer to the central limit theorem and for corroboration might cite previous experience with similar data. There is however an alternative approach which may be more convincing. The model (2.1) with $n > 1$ can be characterized by the two conditions

$$(2.2a) \quad \text{the } X\text{'s are independent}$$

and

$$(2.2b) \quad \text{the joint density of the } X\text{'s is spherically symmetric, i.e. the density is the same at all points equidistant from the origin.}$$

(For a discussion of the equivalence of (2.2) and (2.1) see for example Feller (1966, p.77/78) and Letac (1981).)

* Although even artistic endeavors require techniques which can be systematized and learned.

The meaning of condition (2.2b) is much easier to grasp than that of normality. As an example suppose that X_1 and X_2 are the coordinates of the impact of a shot fired at a circular target, where we shall assume that the expected point of impact is the bull's eye which is taken to be the origin. If the method of aiming is asymmetric in the horizontal and vertical directions (as is typically the case), the assumption of circular symmetry will usually not be justified; on the other hand, one can imagine automated computer methods of aiming for which this symmetry does hold.

Example 2.2. As a second example, consider the assumption that a variable X is distributed according to an exponential distribution. This family of distributions is characterized by the property of "no aging" i.e. the distribution of the lifetime remaining at time t (given that life has not terminated prior to t) is the same as the distribution of the lifetime X at birth. In many situations this property clearly will not hold and an exponential distribution therefore be unsuitable. Situations in which the assumption may be reasonable are described for example in Mann, Schafer, and Singpurwalla (1974).

Example 2.3. Consider finally the Poisson distribution. In many applications it arises as the distribution of the number of events occurring in a fixed interval, where the events are generated by a Poisson process. The latter is characterized by two assumptions: (a) the numbers of events occurring in non-overlapping intervals are independent, and (b) the probability of more than one event occurring in a very short interval is of smaller order than the probability of a single occurrence.

Simple characterizations such as those of Examples 2.1-2.3 not only provide relatively clear criteria for the applicability of a model in a given situation but they may also suggest, when the assumptions are not satisfied, in what way the assumptions are violated and on this basis lead to ideas as to how the model should be modified. For example, a Poisson model may not be suitable because of the presence of "multiple births", and this possibility can be incorporated into the model by specifying the distribution of the "litter size". (For a discussion of this idea from a modeling point of view see for example Neyman and Scott (1959) and Cox and Isham (1980); some related ideas are considered by Feller (1943)).

An assumption that makes an appearance in both Examples 2.1 and 2.3 and which generally enjoys great popularity is the assumption of independence. As has recently been emphasized by Kruskal (1988), this assumption is often made rather casually. It is for example frequently taken for granted that successive observations by the same observer are independent. That this in fact tends not to be the case was noted by Karl Pearson (1902) who carried out some experiments for this purpose. The issue is discussed by Student (1927) and Cochran (1968). Further references can be found in

Kruskal's paper. Unfortunately, the independence assumption is not only very seductive but the resulting inferences are liable to serious error when the assumption is not justified.

Classes of models for dependent observations taken in series are treated in the theory of time series. Observations which, while not independent, retain the complete symmetry of the iid assumption are called **exchangeable**. The theory of exchangeable events was initiated by de Finetti; for references and recent developments see, for example, Kingman (1978), Koch and Spizzichino (1981), Galambos (1982), and Diaconis and Freedman (1984).

(ii) **Model selection.** Procedures for choosing a model not from the vast storehouse mentioned in (i) but from a much more narrowly defined class of models are discussed in the theory of **model selection**. A typical example is the choice of a regression model, for example of the best dimension in a nested class of such models. The best fitting model is of course always the one that includes the largest number of variables. However, this difficulty can be overcome in a number of ways, for example by selecting the dimension k which minimizes the expected squared prediction error, i.e. the expected squared difference between the next observation and its best prediction from the model.

This measure has two components

$$(2.3) \quad E(\text{squared prediction error}) = (\text{squared bias}) + (\text{variance}).$$

As the dimension k of the model increases, the bias will decrease. At the same time the variance will tend to increase since the need to estimate a larger number of parameters will increase the variability of the estimator. Typically there will exist a minimizing value k_0 which then provides the desired solution. The value of k_0 will tend to increase as the number of observations increases and its determination thus constitutes an implementation of Fisher's suggestion that "more or less elaborate forms will be suitable according to the volume of the data." [In practice one minimizes not the expected squared prediction error which depends on unknown parameters but a suitable estimator of this expected value].

Areas in which model selection has been successfully employed include, besides regression, the choice of an appropriate ARMA model, choosing the order of a Markov chain, or the maximal order of the interactions to be included in factorial and loglinear models. See for example Breiman and Freedman (1983: regression), Poskitt (1987: ARMA models), Katz (1981: Markov chains), Linhart and Zucchini (1986), and Edwards and Havranek (1987).

As described above, this approach appears to make the choice of model automatically, solely on the basis of the data. However, this view of model selection ignores a

preliminary step: the specification of the class of models from which the selection is to be made. This specification often will also be quite routine along the line: we are modeling a smooth surface — so let's approximate it by a polynomial regression equation including all terms up to a given degree. In choosing the variables to include in the list of possibilities we may be quite liberal and let the procedure make the choice from this list (as well as the degree of the polynomial) from the data. However, in other cases, the choice of class of models may be strongly informed by subject matter considerations.

(iii) **Modeling a sequence of dichotomous trials.** The great difficulty of model specification is well illustrated by the history of models for sequences of dichotomous trials such as tosses with a coin, births classified by sex, defective and nondefective items on an assembly line, clinical trials, or free throws in basket ball.

In his 1710 paper, "An Argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes," Arbuthnot assumed a binomial distribution, and on this basis tested (and rejected) the hypothesis that $p = 1/2$. The binomial model of course is the result of two assumptions: (i) the constancy of the success probability p ; (ii) the independence of the trials. Heyde and Seneta (1977) and Stigler (1986) discuss the history of the struggle with these assumptions which went on throughout the 19th century, and which led to the development of more general alternative models and of tests of the binomial hypothesis.

A curious debate sprang up in the early 20th century in publications by Marbe (1899, 1916, 1919), Sterzinger (1911), and Kammerer (1919), in which these authors claimed that the results of probability theory contradict the workings of the real world and are therefore not applicable to reality. These beliefs were based on their theories of the "bunching" of events (Knäuelungstheorie) of Sterzinger, the "uniformity" (Gleichförmigkeit) of the world (Marbe), or Kammerer's principle of seriality, bolstered by inconsistencies which the authors thought they had discovered between the observed numbers of runs in various dichotomous sequences such as coin tosses or births and the probabilities of such runs calculated on a binomial model. The arguments show that what is really being put in question is the assumption of independence. The flaws in some of this work are analyzed by Bortkiewicz (1917) and Czuber (1923). (For a recent study of related issues, see Gilovich, Vallone, and Tversky (1985), and Tversky and Gilovich (1989)).

In addition to the belief in positive dependence of events that are close together in time and space, which would result in an excess of long runs, there is also a common belief in negative dependence according to which each success decreases the probability of success for the next trial. It is interesting that this misconception concerning dependence was already noted by Laplace (1814, 1917). One particularly lively

paragraph in Chapter 16 of his Philosophical Essay on Probability begins:

“I have seen men, ardently desirous of having a son, who could learn only with anxiety of the birth of boys in the month when they expected to become fathers. Imagining that the ratio of these births to those of girls ought to be the same at the end of each month, they judged that the boys already born would render more probable the births next of girls.”

3. Two types of models.

The distinction between the two types of models mentioned in Section 2: **Empirical** (or interpolatory) on the one hand, and **Explanatory** (or mechanistic) on the other, has been noted in one form or another by many writers and tends to make an appearance in most general discussions of modeling. Models of these two types differ in many respects and the distinction throws light on some aspects of the modeling process.

- (i) **Purpose.** As pointed out by Box (for example in Box and Hunter (1965) and Box and Draper (1987)) and others, the two kinds of models differ in their basic purpose. Empirical models are used as a guide to **action**, often based on forecasts of what to expect from future observations (eg. the number of college applications, demand for goods, or stock market performance). In contrast, explanatory models embody the search for the basic mechanism underlying the process being studied; they constitute an effort to achieve **understanding**.

The following description of the distinction is from a paper dealing with the role of models in ecology (Goldstein, 1977).

“Within a given ecological research program, modeling can be a valuable procedure in helping to address a number of frequently occurring research objectives. A basic research objective is increased fundamental understanding of the system being studied. This need not be an objective of all research programs. Oftentimes, there is a desire to produce a specific output from a given system. In many circumstances, this goal can be achieved through a well-designed manipulation of the system’s inputs, without any attempt to derive a basic understanding as to how the system functions. This type of approach is frequently referred to as an “input-output” analysis and the system is described as a “black box”. Mathematical modeling techniques can be very helpful in this type of analysis as well.”

- (ii) **Environment.** The two kinds of aims and attitudes described in (i) tend to arise in somewhat different environments. This distinction is discussed for example in a paper by Healy (1978) who contrasts technology with science, or applied with theoretical science, as follows:

“I merely want to propose that much of what is commonly described as science comes more appropriately under the heading of technology. I could soften the blow... if I substituted for ‘technology’ the term ‘applied science.’ ” Asking what distinguishes the two, Healy goes on to state that: “I hold that, in contrast to the scientist, the technologist is not concerned with truth at all... The mark of the technologist is that he must act; everything that he does has some sort of deadline. He has to manage therefore, with as much truth as is available to him with the scientific theories current in his time.”

It is interesting to contrast this with the following passage from Box and Hill (1967).

“It should be noted that the objective we are presently considering is that of finding out ‘how a system works’. The reason for this may be no more than scientific curiosity. However, if we know how the system works and can describe it by a mathematical model, then we can use this knowledge for practical aims such as predicting the behavior of the process under various experimental conditions and, in particular, in finding optimum operating conditions. This fact leads to some confusion because if all one need to do is either to estimate the behavior of the process under various experimental conditions or to find optimum operating conditions, we do not necessarily need a mechanistic model. In some circumstances, an attempt to discover a mechanism merely to develop an operable system would be needlessly time consuming.”

Both passages are somewhat defensive, but they defend against attacks from opposite directions. In the expectation that his audience will feel insulted by having their work described as technological rather than scientific, Healy says that he “could soften the blow” of calling what much of statistics is concerned with technology rather than science by using a less offensive terminology. Box and Hill, on the other hand, apologize for being needlessly scientific in a situation which calls for a solution to a practical problem. They point out that developing an explanatory model rather than being satisfied with an empirical one might be justified by more than mere “scientific curiosity”: the knowledge gained in this way might actually be utilized “for practical aims.”

- (iii) **Ad hoc nature vs broad applicability.** The difference in attitudes just described leads quite naturally to different positions concerning the appropriate level of generality of the models and the conclusions derived from them. A scientific model describing the structure of the underlying mechanism and the

* A heated debate concerning the status of the logistic curve as a general law of population growth is discussed in Kingsland (1985).

laws governing it will strive for the most general formulation possible.* This typically requires abstraction and idealization in order to eliminate the specific circumstances of particular situations. General physical laws explaining the motion of bodies or biological laws describing the genetic mechanisms of inheritance apply without regard to the nature of particular bodies and to very general classes of biological organisms.

On the other hand, a technological model intended to provide guidance in a particular situation at hand will want to make full use of all the special circumstances of that situation. In particular, it needs to provide a good approximation only over the range of values of interest. Thus a linear regression may be perfectly adequate for the problem under consideration even when it is clear that a linear approximation can provide a reasonable fit only over a very limited range.

- (iv) **The role of subject matter.** An explanatory model, as is clear from the very nature of such models, requires detailed knowledge and understanding of the substantive situation that the model is to represent. On the other hand, an empirical model may be obtained from a family of models selected largely for convenience, on the basis solely of the data without much input from the underlying situation. (Examples of both situations can be found in the writings of Box and his coworkers, for example in Box and Hunter (1965), Box and Hill (1967), Box, Hunter and Hunter (1978), and Box and Draper (1987)).

It is interesting to examine Kepler's elliptical orbits and Mendel's laws of inheritance from this point of view. A first look at Kepler's work may lead to the conclusion that he just tried to fit a simple curve to the available data and found that an ellipse was both mathematically simple and provided an excellent fit. However, closer study reveals that his discovery came about as the result of a great deal of previous thinking and theorizing about the subject matter. For a detailed discussion of this point, see for example Aiton (1975) and Wilson (1975).

Similarly it is tempting to believe that Mendel, often depicted as working in monastic isolation, came to his startling innovative laws solely by contemplating the results of his experiments with peas. Again it emerges from a more detailed study that his conclusions and explanations were by no means based only on his data. (See for example, Orel (1984) and Olby (1985).)

- (v) **Validating the model.** The difference in the aims and nature of the two types of models implies very different attitudes toward checking their validity. Techniques such as goodness of fit tests or cross validation serve the needs of checking an empirical model by determining whether the model provides an adequate fit for the data. Many different models could pass such a test, which reflects the fact that there is not a unique correct empirical model. On the

other hand, ideally there is only one model which at the given level of abstraction and generality describes the mechanism or process in question. To check its accuracy requires identification of the details of the model and their functions and interrelations with the corresponding details of the real situation. And the explanation must provide a detailed fit not only for the particular phenomenon at hand but must also fit appropriately into the existing scientific fabric of related phenomena in which it is to become embedded.

It is interesting to note that the distinction discussed in this section was realized, and considered important, long before the modern scientific age. Following are three quotes, one each from the 16th, 17th, and 18th centuries, as given in Popper (1965).

The first is from Osiander; 'Preface to Copernicus' "De Revolutionibus" (1543), putting a "spin" on this work of which Copernicus would hardly have approved had he lived to see it:*

"There is no need for these hypotheses to be true, or even to be at all like the truth**; rather one thing is sufficient for them — that they should yield calculations which agree with the observations".

Osiander was a Protestant Theologian. The issue became more heated in the following century in the dispute between Galileo and the Catholic Church. The position of the latter as stated by Cardinal Bellarmine in 1615 was that the church would raise no objections if Galileo stated his theory as a mathematical hypothesis, "invented and assumed in order to abbreviate and ease the calculations", provided he did not claim it to be a true description of the world.

As a last expression of this thought, here is a quote from Bishop Berkeley (1734).

"A mathematical hypothesis may be described as a procedure for calculating certain results. It is a mere formalism... judged by its efficiency... The question of truth of a mathematical hypothesis does not arise — only that of its use as a calculating tool."

Although these passages of course refer to deterministic rather than stochastic models, the idea of empirical modeling could hardly be expressed more clearly today.

* For additional discussion of Osiander's Preface see Toulmin (1961) and particularly Blumenberg (1987).

** Throughout this paper it is tacitly assumed that there exists an underlying "true" situation which one is attempting to model. This is an attitude much discussed by philosophers of science. Some recent references are Cartwright (1983), Hacking (1983), Jardine (1986), and Giere (1988).

4. A spectrum of possibilities.

The discussion in the preceding section of empirical and explanatory models ignores the fact that many, perhaps most, actual modeling situations have an intermediate character, exhibiting some features of each of the two types. This was clearly recognized by Kruskal and Neyman (1956) who wrote:

“Although the distribution between true theory and interpolation is sometimes quite sharp in specific cases, the two models of analysis really represent somewhat extreme points of a continuous spectrum... The continuum is by no means precise, nor is it meant to be so...”

(The authors then proceed, as an amusing pastime, to place various kinds of models on a “fanciful scale” meant to represent this imprecise continuum. To factor analysis, for example on a scale from 0 (pure interpolation) to 10 (fully explanatory) they assign a score of 4).

The corresponding remark applies to the discussion of science vs technology. In this connection it is helpful to recall the distinction made by Kuhn (1970) between normal and revolutionary science. As suggested by Nelder (1986), normal science occupies a position somewhere between technology and revolutionary science, and again there exists a whole spectrum of intermediate shadings. Concerning the role of statistics in these different environments Nelder writes:

“I doubt if statistics has much to offer to revolutionary science... The position is very different, however, with both normal science and technology, though some justification may be needed for grouping them together here. I would argue that there are major differences between normal science and technology, for example in the relevance of cost-benefit assessments to their progress, their long-term objectives, and their attitudes to theory (Healy, personal communication), it is nonetheless true that, on the scale of day to day activity, the procedures of (normal) scientist and technologist will be found to be very much alike. They will both be working within a given theoretical framework and be concerned, for instance, with estimating quantities defined within that framework, with confirming the estimates of others, and with relating their estimates to those predicted by theory.”

The construction of a satisfactory “revolutionary” model is essentially the problem of scientific discovery. Where the ideas for such discoveries come from is one of the central problems in the philosophy of science which has been discussed by many scientists and philosophers. Most scientists agree with Nelder’s implication that there can be no systematic aid to discovery, that it is a matter of imagination and inspiration. One of the most influential philosophers of science, Karl Popper, sums this up by saying (Popper 1965, p. 192) that scientific hypotheses are “the free creations of our own

minds, the result of an almost poetic intuition.”

Yet even in this extreme case the situation is not as clearcut as it seems at first. Examination of some of the classical examples of revolutionary science shows that the eventual explanatory model is often reached in stages, and that in the earlier efforts one may find models that are descriptive* rather than fully explanatory. This is for example true of Kepler whose descriptive model (laws) of planetary motion precede Newton’s explanatory model. To some extent, this remark also applies to Mendel’s revolutionary model for inheritance. His theory does offer an explanation for his laws of segregation and independent assortment in terms of factors responsible for the transmission of genetic material. However, he was not in a position to identify these factors and so obtain the required identification between his model and biological reality.

5. Conclusions.

This paper has been concerned with the question of what contribution statistical theory has to make to model specification or construction. Three areas of such contributions turned out to be:

- (a) **A reservoir of models**, with particular emphasis on transparent characterizations or descriptions of the models that would facilitate the understanding of when a given model is appropriate. (A special problem in this area which to requires additional work is a study of the circumstances under which independence is or is not a suitable assumption.)
- (b) **Model selection**. This is a body of techniques for selecting a particular model (or parametric subfamily of models) from a rather narrowly specified family of models. It seems likely that this approach will be developed much further particularly with the help of techniques from artificial intelligence. The use of expert systems will make it easier to inject subject matter information into this process. On the other hand, it is difficult to see how this approach can break out of a current paradigm and thus lead to revolutionary scientific discoveries. (An interesting and thoughtful “progress report” on the possible role of artificial intelligence in scientific discovery is provided by Langley et al (1987).

* Such descriptive models should perhaps be considered as a distinct third type intermediate between empirical and explanatory models. A division of models into these three types by Berkeley is discussed by Popper (1965, p. 169).

- (c) **Classification of models.** The distinction between explanatory and empirical models provides only one of the many ways in which models can be classified. A better understanding of the modeling process can perhaps be obtained by distinguishing various other types of situations which require different kinds of models. Fundamental differences between the needs and possibilities in the social and physical sciences, for example, are discussed by Lieberman (1985) and in the debates initiated by Freedman (1985, 1987). The corresponding issues regarding biology and physics (with extensive references to the literature) are considered for example in Rosenberg (1985).

Another important distinction is that between deterministic and stochastic models. Of course any statistical model has a stochastic component. This may however enter either only through the errors in the measurements of an essentially deterministic situation (as in Kepler's astronomical data) or also through the basic stochastic nature of the underlying phenomenon (for example, in Mendel's theory).

As a last example we mention the distinction between models which do or do not employ what psychologists have called "hypothetical constructs" (see for example MacCorquodale and Meehl, 1948). These are entities whose existence is required by the theory but has not actually been established although it is hoped that observation will eventually change their status. Mendel's genes or some of the elementary particles are cases in which eventual verification did occur; structures required for certain mental abilities, for example color vision, provide another illustration. On the other hand, the postulated existence of ether and phlogiston in the end had to be abandoned.

In addition to these different types of models, it is useful to distinguish between two aspects, both of which are typically present in the same model: the subject matter part of the model and the part played by "error". Here the latter term is meant to include not only measurement error but impurities in the material, changes in temperature or time of day, in fact all the contributions to the observations of the various environmental and observer effects that are extraneous to the subject matter.

Acknowledgments.

This paper greatly benefited from discussions with Persi Diaconis, Lucien Le Cam, and Juliet Shaffer.

6. References.

E.J. Aiton (1975). The elliptical orbit and the area law. In *Four Hundred Years* (Proceedings of conferences held in honour of Johannes Kepler (Beer and Beer, Eds)

Pergamon Press. Oxford.

Arbuthnot, J. (1710). An argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes. *Phil. Trans. Roy. Soc. London* **27**, 186-190.

Blumenberg, H. (1987). The genesis of the Copernican world. MIT Press Cambridge.

Bortkiewicz, L. von (1917). Die Iterationen. Springer, Berlin.

Box, G.E.P. and Draper, N. (1987). Empirical Model-Building and Response Surfaces. John Wiley. New York.

Box, G.E.P. and Hill, W.J. (1967). Discrimination among mechanistic models. *Technometrics* **9**, 57-71.

Box, G.E.P. and Hunter, W.G. (1965). The experimental study of physical mechanisms. *Technometrics* **7**, 23-42.

Box, G.E.P.; Hunter, W.G.; and Hunter, J.S. (1965). Statistics for Experiments. John Wiley. New York.

Breiman, L. and Freedman, D. (1983). How many variables should be entered into a regression equation? *J. Amer. Statist. Assoc.* **78**, 131-136.

Cartwright, N. (1983). How the Laws of Physics lie. Oxford Univ. Press.

Cochran, W.G. (1968). Errors of measurement in statistics. *Technometrics* **10**, 637-666.

Cox, D.R. and Isham, V. (1980). Point Processes. Chapman and Hall. London.

Czuber, E. (1923). Die Philosophischen Grundlagen der Wahrscheinlichkeitsrechnung. Teubner. Leipzig.

Dawid, A.P. (1982). Intersubjective statistical models. In Exchangeability in Probability and Statistics (Koch and Spizzichino, Eds). North Holland.

Diaconis, P. and Freedman, D. (1984). Partial exchangeability and sufficiency. In Statistics: Applications and New Directions (Ghosh and Roy, Eds). Proc. Indian Statist. Instit. Golden Jubilee Internat. Conference.

Edwards, D. and Havranek, T. (1987). A fast model selection procedure for large families of models. *J. Amer. Statist. Assoc.* **81**, 205-213.

Feller, W. (1943). On a general class of 'contagious' distributions. *Am. Math. Statist.* **14**, 389-400.

- Feller, W. (1966). An Introduction to Probability Theory and its Applications. Vol. II, p. 77/8.
- Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy Soc. London (A)* **222**, 309-368.
- Freedman, D. (1985). Statistics and the scientific method. In Cohort Analysis in Social Research: Beyond the Identification Problem (Mason and Fienberg, Eds). Springer. New York.
- Freedman, D. (1987). As others see us: a case study in path analysis (with discussion). *J. Educ. Statist.* **12**, 101-223.
- Galambos, J. (1982). Characterizations of distributions. In *Encycl. of Statist. Sciences*, Vol. 1, 422-428.
- Galambos, J. (1982). Exchangeability. *Enc. Statist. Scie.* **2**, 573-577.
- Galambos, J. (1982). Exponential distribution. *Enc. Statist. Sci.* **2**, 583-587.
- Giere, R.N. (1988). Explaining Science. Univ. of Chicago Press.
- Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* **17**, 295-314.
- Goldstein, R.A. Reality and models. In *Mathematical Models in Biological Discovery* (Solomon and Walter, Eds). Springer. New York.
- Ian, Hacking (1983). Representing and Intervening. Cambridge Univ. Press.
- Healy, M.J.R. (1978). Is statistics a science? *JRSS (A)* **141**, 385-343.
- Heyde, C.C. and Seneta, E. (1977). I.J. Bienaymé. Springer. New York.
- Jardine, N. (1986). The Fortunes of Inquiry. Clarendon Press. Oxford.
- Kammerer, P. (1919). Das Gesetz der Serie. Dentsche Verlags-Anstalt. Stuttgart.
- Katz, R.W. (1981). On some criteria for estimating the order of a Markov Chain. *Technometrics* **23**, 243-249.
- Kingman, J.F.C. (1978). Uses of exchangeability. *Ann. Probab.* **6**, 183-197.
- Kingsland, S.E. (1985). Modeling Nature. University of Chicago Press.
- Koch, G. and Spizzichinco, F. (Eds) (1981). Exchangeability in Probability and Statistics. North Holland. Amsterdam.

- Kruskal, W. (1988). Miracles & Statistics. the casual assumption of independence. *J. Amer. Statist. Assoc.* **83**, 929-940.
- Kruskal, W. and Neyman, J. (1956). Stochastic models and their applications to social phenomena. (Unpublished lecture presented at joint session of IMS, ASA, and the Amer. Social Soc., in Detroit, Sept. 1956).
- Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. 2nd Ed. Univ. of Chicago Press.
- Langley, P.; Simon, H.; Bradshaw, G. and Zytkow, J. (1987). *Scientific Discovery: Computational Explorations of the Creative Processes*. MIT Press. Cambridge.
- Laplace, P.S. (1917). *A Philosophical Essay on Probabilities* (Truscott and Emory, Transl.) John Wiley. New York.
- Letac, G. (1981). Isotropy and sphericity: Some characterizations of the normal distribution. *Ann. Statist.* **9**, 408-417.
- Liebertson, S. (1985). *Making it Count*. U.C. Press, Berkeley.
- Linhart, H. and Zucchini, W. (1986). *Model Selection*. John Wiley. New York.
- Luce, A.D. (1986). *Response Times. Their role in inferring elementary mental organization*. Oxford Univ. Press.
- MacCorquadale, K. and Meehl, P. (1948). On a distinction between hypothetical constructs and intervening variables. *Psych. Rev.* **55**, 95-107.
- Makridakis, S., Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski R.; Newton, J.; Parzen, E.; and Winkler, R. (1984). *Forecasting accuracy of major time series methods*. John Wiley. New York.
- Mann, N.R.; Schafer, R.E., and Singpurwalla, N.D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. John Wiley. New York.
- Marbe, K. (1899). *Naturphilosophische Untersuchungen zur Wahrscheinlichkeitstheorie* Leipzig.
- Marbe, K. (vol. 1, 1916; vol. 2, 1919). *Die Gleichförmigkeit in der Welt*. C.H. Beck, München.
- Nelder, J.A. (1986). Statistics, Science and Technology. *JRSS (A)* **149**, 109-121.
- Neyman, J. and Scott, E.L. (1959). Stochastic models of population dynamics. *Science* **130**, 303-308.

- Olby, R. (1985). *Origins of Mendelism* (2nd Ed). Univ. of Chicago Press. Chicago.
- Orel, V. (1984). *Mendel*. Oxford Univ. Press.
- Pearson, Karl (1902). On the mathematical theory of errors of judgment. *Phil Trans. Roy. Soc. London (A)* **198**, 235-299.
- Popper, K. (1965). *Conjectures and Refutations: The Growth of Scientific Knowledge* (2nd Ed).
- Poskitt, D.S. (1987). A modified Hannan-Rissanen strategy for mixed autoregressive-moving average order determination. *Biometrika* **74**, 781-790.
- Rosenberg, A. (1985). *The Structure of Biological Science*. Cambridge Univ. Press.
- Shibata, R. (1981). An optimum selection of regression variables. *Biometrika* **68**, 45-54.
- Sterzinger, D. (1911). *Zur Logik and Naturphilosophie der Wahrscheinlichkeitslehre* Leipzig.
- Stigler, S.M. (1986). *The History of Statistics*. Belknap. Cambridge.
- Student (1917). Errors of routine analysis. *Biometrika* **XIX**, 151-164.
- Thompson, J.R. and Brown, B.W. (Eds) (1987). *Cancer Modeling*. Marcel Dekker. N.Y.
- Tong, H. (1975). Determination of the order of a Markov chain by Akaike's information criteria. *J. Appl. Prob.* **12**, 488-497.
- Toulmin, S. (1961). *Foresight and Understanding*. Indiana Univ. Press.
- Tversky, A. and Gilovich, T. (1989). The cold facts about the "Hot hand" in basketball. *Chance* **2**, 16-21.
- van Fraassen, B. (1989). *Laws and Symmetries*. Oxford Univ. Press.
- Wilson, C.A. (1975). Kepler's ellipse and area rule - their derivation from fact and conjecture. In *Four Hundred Years* (Proceedings of Conferences held in honour of Johannes Kepler. (Beer and Beer, Eds). Pergamon Press. Oxford.