2 of 3

USNRDL-TR-749
23 April 1964

A SYSTEM FOR DEVELOPMENT OF MATHEMATICAL MODELS

by
J. W. Hendricks

63-F
HC $3.00
MF $0.70

U.S. NAVAL RADIOLOGICAL DEFENSE LABORATORY

SAN FRANCISCO · CALIFORNIA · 94135

RADIOLOGICAL EFFECTS BRANCH
E. A. Schuert, Head

CHEMICAL TECHNOLOGY DIVISION
L. H. Gevantman, Head

ADMINISTRATIVE INFORMATION

The work reported is part of a project sponsored by the Defense Atomic Support Agency under NWER Program A-7, Subtask 10.001. This project is described in the USNRDL Technical Program Summary for Fiscal Years 1964, 1965, and 1966, dated 1 October 1963, wherein it is identified as Program A-1, Problem 2.

DDC AVAILABILITY NOTICE

Qualified requesters may obtain copies of this report from DDC.

Eugene P. Cooper

Eugene P. Cooper
Scientific Director

D C Campbell, CAPT USN
Commanding Officer and Direc

12ND NRDL P1 (9/63)

ABSTRACT

This paper presents a general quantitative system for association of theory and observation. The system, in formulation and presentation, is directed toward the needs of the user investigator. It is well organized for use with automatic machines in the computing and information processing.

A multivariable distribution approach to model construction is used in accounting for errors and other sources of variation. As usual, hypothesized mathematical descriptions are modified in accordance with the observational data. The requirement of a state of control is regarded as fundamental. The question of unidentifiability is given prominent consideration; and in this connection, the general necessity of calibration is established and emphasized. The principle of maximum likelihood is suggested as the most acceptable ranking criterion for the system; but modifications or decision-theoretic extensions are not precluded. In short, the ordinary restrictive conditions are not imposed in this system.

The fundamental ideas are discussed within the contextual framework of the system. The general principles for applying the system are presented and discussed. The most important classes of models are dealt with mathematically in detail.

DESCRIPTORS

1. Model
2. Mathematical
3. Development
4. System
5. Observation
6. Experiment
7. Error
8. Theory
9. Hypotnesis
10. State-of-Control
11. Identifiability
12. Calibration
13. Description
14. Method
15. Science
16. Structure
17. Relationship
18. Estimation
19. Prediction
20. Probability
21. Distribution
22. Density
23. Multivariable
24. Specification
25. Designation
26. Computers
27. Automation
28. Likelihood
29. Ranking
30. Fitting
31. Regression
32. Least-Squares
33. Transformation

SUMMARY

The Problem

We are concerned here with the problem of mathematical model development in complex observational situations. We wish to adjust hypothesized mathematical formulations in accordance with the observational data. This problem is of course basic to all quantitative science. Investigators, whatever their specialized disciplines, are currently confronted with a bewildering hodgepodge of scientific methodologies. To the great majority, preoccupied with their individual specialties, the extensive literature of potentially applicable mathematics, statistics, philosophy of science, etc., will always be practically inaccessible. Also, those techniques which have ordinarily been accepted and employed, in general tend to embody unrealistic assumptions, inherent inadequacies, or both.

A generally valid system is needed which is reasonably comprehensible, at least in application. In addition, the system should impose no significant inconvenience on the user investigator. More or less generally valid systems have been attempted previously; however, restrictions on their use or problems of application have apparently precluded their acceptance.

Results

In this paper, a completely general system for model development is presented. The entire formulation has been carried out as an attempt to match the viewpoint of the working investigator rather than that of the mathematical statistician. The system, in all of its component methodologies, is principally directed to the understanding and convenience of the investigator. The valid association of theory and observation is accomplished without resort to restrictive conditions. Except for the most general and fundamental criteria, such as the principal of maximum likelihood which is herein suggested as most appropriate, any needed assumptions are regarded as solely the investigators responsibility and as necessary parts of his initial hypotheses.

The system organization is specifically oriented toward use in conjunction with large-scale automatic computing systems. It is presumed that most investigators will have access to such equipment in the near future, and that eventually no routine mathematical analysis of significant complexity will be attempted without such access. Thus, the matter of the investigators convenience will be resolved.

Fundamental ideas are discussed briefly, within the contextual framework of the system. The general principles for applying the system are presented and discussed. The most important classes of models are dealt with mathematically in detail. It is hoped that this paper will also serve as a handbook of general investigative methodology.

CONTENTS

ABSTRACT 1
DESCRIPTORS ii
SUMMARY iii

INTRODUCTION 1
Background Note 3

THE BASIC PRINCIPLE OF THE SYSTEM 6
Principle 8

ASSOCIATION OF THEORY AND OBSERVATION 8

DESIGNATION OF THE PHENOMENON DESCRIBED BY THE MODEL 9

DISTRIBUTIONS MAY RESULT FROM EITHER RANDOM OR SYSTEMATIC SELECTION 11

MATHEMATICAL THEORY AND DISCUSSION 13
Initial Hypotheses 14
The Error Transformation 16
Figure 1 18
Surface Density Functions 17
Figure 2 20
Concordant Hypotheses 19
Available Techniques 19
Purposeful Distributions 23

MODEL TYPES 24
Type I Model: Simple Measurement with Error 24
Type I Schematic 27
Type II Model: All Object Variables Distributed 26
Type II Schematic 29
Type III Model: Structural Relationships 28
Type III Schematic 30
Prediction 31
Type IV Model: Regression on Observable Variables 31
Type IV Schematic 33
Type V Model: Discrete Distribution 34
Type V Schematic 36
Attribute Distributions 37

Type VI Model: Constrained Distributions. 37
Type VI Schematic. 38
Type VII Model: Variform or Discontinuous Models. 37
Type VII Schematic 40
Indirect Observation 39
Other Model Types. 39

PROBLEMS IN COMPUTATION 41

UNIDENTIFIABILITY. 43
Figure 3. 45

INTERPRETATION OF RESULTS 46
Surfaces of Unidentifiability. 46
Calibration. 47
Considerations in Ranking Models 48
State of Control. 50
Figure 4. 51

CONCLUDING REMARKS. 51

REFERENCES. 53

INTRODUCTION

This paper presents a general quantitative system for association of theory and observation. Most phenomena of scientific interest are at least subject to the possibility of quantitative description. Phenomena subject to quantitative description are also subject to compact mathematical representation. It is not necessary that such representation be made; a complete set of quantitative observations or data points contain in themselves all available information, by definition. However, it is our faith in the orderliness of creation, together with past supporting experience, which engenders in us the desire to fill in the space between the data points. We prefer to search for mathematical descriptions which not only account for all past observations, but also all possible observations, in order that prediction will be possible. If a description (or model) is simple and elegant, and at the same time serves to correctly predict a reasonable number of subsequent observations, we feel that understanding of the subject phenomenon has been achieved.

Thus our stated purpose in model development is description and not necessarily decisive action. It is of course implied that description eventually leads to decisions and actions, but actions of character and purpose yet to be established. Specifically, the criteria of future decisions, and the utility values of the possible ultimate consequences, are at present unknown.

The decision-theoretic approach has attained great stature, but not every investigator can adopt a completely behavioristic philosophy. Probably most will remain fundamentally realists. Nevertheless, nothing precludes the inclusion of decision theoretic considerations into the methods of this paper.

We are concerned here with the problem of model development in complex observational situations. In such cases, simple and elegant descriptions are difficult to obtain. Economic pressure or urgent necessity may force us to resort to oversimplifications and unjustified assumptions. Unfortunately, descriptions obtained under such conditions are not likely to be useful in prediction, and their contribution to

understanding is likely to be small. The usual strategy in dealing with complex problems is to attempt a division into more manageable subproblems. Hopefully, the subproblems will respond to the available investigative procedures. Here too, the results are not always successful.

In this paper an operational system is proposed for use in development of complex mathematical models. To this end, consider that complexity in observational situations can be conveniently divided into two classes. First, the complexity may be due to a multiplicity of influential variables. Second, the complexity may be due to the presence of interacting phenomena.

In the first instance, if the number of influential variables is very large, it may be best, in the judgment of the investigator, that the variables be grouped and their combined group influences be observed and described. For example, a group of influences, unobservable individually, may be represented in combination as a random variable. Or, if the number of variables is not so large as to justify grouping, the investigator might not wish to so simplify his theory. Instead, in spite of the complexity, he may elect to describe deterministically the relationships between the numerous variables taken individually.

In the second instance, it may be the case that two or more distinct phenomena are so interrelated that they cannot be observed in isolation. For example, most measurement phenomena are subject to interaction with those phenomena which are the errors of observation. Also, it is a common occurrence, that observations can only be made indirectly. That is to say, the unobservable variables of primary interest can only be determined as functions of variables which are directly observable.

Any system for development of complex mathematical models must be sufficiently general to encompass these situations. In the first instance, means must be provided for either grouping, or individually relating, the influences of variables, or both simultaneously. In the second instance, means must be provided for separately describing the effects of interrelated phenomena. The system herein presented is generally applicable, it is reasonably simple in conception, and it lends itself well to use of automatic machines in computation and in information processing.

The need for an organized or systematic approach to model development has become increasingly evident in recent years. This is due in part to the accelerating trend toward greater complexity of research problems in all of the quantitative sciences. But in addition, a wide variety of problems are either inadequately treated or are completely

intractable with currently available methods. New methods are introduced here which have resulted in the solution of the previously intractable problems, and which at the same time, have provided a basis for the organization of model development which is presented in this paper.

It is perhaps helpful to mention some major fields of research which will benefit from use of this system. For example, in the biological sciences, reaction rate models are assuming increasing importance, and stochastic models are more widely employed than ever. The large number of variables and interacting phenomena involved in biological systems has resulted in a generally unsatisfied demand for accurate mathematical models. Operations research is concerned almost exclusively with mathematical model development. With reference to these models, decisions are frequently made which are of great and immediate importance to society; consequently, the risk in resorting to inadequate or inaccurate models is large. The social sciences generally are in need of accurate and useful mathematical models for complex phenomena. For example, econometric analysis has been hampered by the lack of adequate theory in model construction; although considerable progress has been made in this field with linear and polynomial models. Also, the physical sciences are no less influenced by the trend toward research interests in complex phenomena. The relatively recent introduction of multidimensional pulse-height analyzers (high-speed, digital, multivariable data point recorders) is intensifying the need for precise methods in data interpretation.

At this time, it is not inappropriate to suggest that a reasonable degree of standardization be introduced in scientific method and reporting. Potentially, this can occur as a useful byproduct of the system application. In some circumstances, where problems in research management exist, the basis may be provided thereby for control of research quality. But most useful perhaps, is the improvement in communication which may result. It should be possible to readily compare or combine models developed under widely varying research environments. Finally, use of generally accepted procedures together with automatic data processing can serve to free scarce professional investigators from many routine or peripheral aspects of research.

Background Note

As stated, the general problem with which we are concerned is that of the meaningful association of theory and observation. Here, a practical user-oriented system has been presented for the development of mathematical models. The calculus of probabilities and the Principle of Maximum Likelihood have been employed. First to apply the calculus

of probabilities for these purposes was Thomas Bayes. Carl Friedrich Gauss was first to formulate and apply the Principle of Maximum Likelihood in 1809 (see R. A. Fisher (4), pp. 20-21, and H. F. Trotter (13), pp. 127-). Gauss approached the general problem from the point of view of his least squares theory.

In the interim, the completely general problem has been rarely considered, at least from the practical point of view. On the other hand, restricted problems have been intensively investigated. In particular, methods assuming linearly related or normally distributed variables dominate the literature. These of course include such approaches as analysis of variance, correlation analysis, and regression analysis. With respect to model development under restricted conditions, linear and polynomial regression have found wide acceptance. However, a persistent interest has been maintained by a relatively small number of investigators in the more general problem of more than one variable subject to error. Many techniques, valid for restricted conditions, have been developed. Excellent reviews of this work are offered by M. G. Kendall (6), and Albert Madansky (8).

Several attempts to provide user-oriented systems stand out. W. E. Deming (3), in 1943, presented a general system based on the principle of least squares. Many somewhat less general expressions of the least squares theory have been successfully reduced to practice using digital computers. For example, R. H. Moore and R. K. Zeigler (10), 1960 were among the first to develop a useful and well-documented program for generalized non-linear regression.

The most significant, as well as the most general, approach previously offered is probably thatof Trygve Haavelmo (5), in 1944. The treatment is thorough, it is a system in the sense of this paper, and it is oriented, at least in spirit, to the needs of the user investigator (in this case econometricians). Also, the mathematical formulation constitutes at least one or two additional model types (which are not included in this paper). However, the formulation is quite complex and, as acknowledged by the author, appears not to be readily manageable in practice. (For example, two nonlinear transformations and two integrations are generally required for derivation of the observation model hypothesis.) Also, the initial hypotheses required may be much more appropriate to economic theory than to that of other disciplines. Nevertheless, it is unfortunate that this work has not received more widespread attention.

In this paper, we have particularly aspired to simple, easily acceptable, and manageable formulations for all model types. For example, the "intersection-projection" transformation of this paper,

effected by substitution, is essential to simple formulation where models involve structural relationships. This device seems obvious enough, as will become apparent: it is frequently applied automatically, without specific reference, in trivial cases. However, it has apparently not been used previously in general formulations of the type here under consideration.

In this paper, it is emphasized that distributions need not be random; and here also, the proper use of systematic distributions is indicated. A similar view on the concept of random selection is expressed by R. B. Braithwaite (1), 1953. However, there seems to be no previous recognition in the literature, and certainly no emphasis, that existent systematic distributions, and in particular purposeful systematic distributions (as in experimental designs), are appropriate and frequently necessary elements in formulation of concordant sets of hypotheses.

The definitive consolidation of work on unidentifiability was given in a paper by T. C. Koopmans and O. Reiersøl (7), in 1950. They begin with recognition of the fact that a completely general formulation of the problem of statistical inference must encompass not only observable populations, but in consideration of errors of observation or other disturbances, the theory or structure of the underlying true phenomena which is thought of as generating the observed distribution must also be considered. The problem of identification is then said to be that of drawing inferences from the probability distribution of observed variables to the underlying structure. This formulation, by Koopmans and Reiersøl, may be thought of as constituting a non-geometric, non-operational statement of the Basic Principle of the present system. Both expressions are closely related to the classic view on the influence of error as expressed by Gauss (13), pp. 1,2: that errors are either constant (systematic) or irregular (random), and that the constant error cannot be estimated from the observations.

In addition to the above introductory contribution, Koopmans and Reiersøl go on to express in greater detail the general concept of identification and to give examples. Their paper makes two other points which are pertinent here. First, we are warned (pp. 169-170) against the temptation to specify models in such a way as to force identifiability, since scientific honesty demands that specification of a model be based on prior knowledge of the phenomenon studied. Second, they point out (p. 179) that even if all parameters are not identifiable, it remains possible to construct identifiable functions of these parameters which constitute useful scientific information. Thus, the appropriate direction of subsequent work would seem to be clearly indicated.

In view of the above background, it then becomes relevant to ask why a workable and acceptable system of adequate generality has not been previously constructed. Several possible explanations come to mind. Principally, the general necessity of resorting to specially constructed calibration models and ultimately to actual calibration in order to resolve problems of unidentifiability in complex observational situations has apparently not been recognized. The emphasis in the literature has been on determination of those conditions necessary for complete identification in restricted situations. In this connection, refer again to Madansky (8). There appears to be no recognition that the information required to verify such conditions must be obtained in most cases by actual calibration. Further, any other means of obtaining the needed information constitutes effective calibration. Thus, it is perhaps a principal contribution of this paper that calibration (or effective calibration), long regarded in classical error theory as an essential feature in scientific method, as applied in simple measurement situations, is here established as also essential in complex observational situations. Complex phenomena are of course widely employed as "controls" in experimentation; but these are trivial instances of calibration.

This does not entirely explain the apparent premature abandonment of generality in most of the literature. It is possible that preoccupation with pencil and paper analytic tractability has played a part. Thus, it is appropriate to point out the advantages of computer-oriented mathematics, particularly in avoiding problems imposed by the limitations of the notation.

As is well known, the concept of state of control is due to W. A. Shewhart (12) 1939. Suitable introductory reading is provided by Munroe (11), Mood (9), and Cramér (2).

THE BASIC PRINCIPLE OF THE SYSTEM

For purposes of this system, the phenomenon under observation is theoretically represented by a distribution of points or events over the Cartesian space of all variables (herein coordinate variables) which are considered to be influential or otherwise of interest in the investigation. It is demonstrable that such a representation may be made. The distribution is generally hypothesized in the compact form of a joint density function, either over the entire space, or over appropriate subspaces. The exact manner in which such representation is accomplished, for the most likely special situations, is resolved in this paper. The general method is also discussed.

Since the model is a density function, it serves to specify the distribution of weight or mass of potentially existent events over the model space. In general, not all arguments of the density function are random variables; some coordinate values may be selected systematically in accordance with a prearranged experimental design. In other words, generally it is not a probability distribution which is specified, if by the term "probability" we imply random selection.

Not all variables of interest, and which are represented in the model, are also observable. Of special interest are the errors of observation and the associated hypothesized true but unobservable variables which they modify. That is to say, we distinguish between observable variables subject to the error of observation, and the associated unobservable variables in fact, which would be free of error could they be observed directly. Denoting actually observed values by subscript o, and errors introduced in the process of observation by e, the relationship is as follows:

$$x_o = x + e_x$$

In this situation, all three variables are of interest and all three are necessarily included among the coordinate variables of the model space over which the phenomenon of interest, including those aspects of the phenomenon which are exclusively related to the process of observation, is hypothetically represented. Thus, there exists a subspace of those variables subject to observation. To these we now direct our attention.

The subspace of the observable variables may be thought of as the means whereby the phenomenon as represented by the model is revealed to the observer. In the same way that we might view the 3-dimensional interior of a house through a window, a 2-dimensional aperture, the subspace of observable variables functions as a window through which we may view the higher-dimensional representation. Indeed, by definition, no other avenue is available. However, it should be noted that what is observable by one method of observation (one set of instruments, one group of investigators, etc.) is not necessarily observable by another. Thus, the same phenomenon might be viewed through a number of different windows.

In this system, the method of viewing consists of projecting the event mass image of the model from the model space of all variables of interest, onto the space of observable variables. This is usually accomplished by integrating out the unobservable variables. That is to say, for each point of the observable space, the distributed event mass, as given by the joint density function model, is summed for all values

of the unobservable variables. The resultant joint marginal density function serves to specify the distribution of potentially observable events.

A precise analogy, which is at the same time an actual example, is given by the process of X-ray photography. Here the model is 3-dimensional and consists of the object being photographed. The observable space is a plane, the essentially 2-dimensional X-ray film. The density function over the observable space is given by the distribution of precipitated silver.

The basic principle of the system should be emphasized.

Principle: The distribution of apparent events over the space of observable variables constitutes our total knowledge of the real unobservable universe. The apparent distribution is literally a projected image of the distribution of real events over higher unobservable spaces.

ASSOCIATION OF THEORY AND OBSERVATION

As additional foundation for construction of the system, it is presumed that quantitative scientific knowledge is best advanced by the following sequence of operations:

1. Consider past theory and observation.
2. Formulate new and generally incomplete hypotheses or models.
3. Employ observational data in completing the one or more formulated models.
4. Employ additional observational data in evaluating and comparing the completed models or theories.

In the event that complete models are hypothesized, the third operation may be dispensed with. However, this is not generally advisable, since in many if not most cases, such a procedure would be tantamount to guesswork.

It is presumed that the joint density function model will not be completely specified but that, in consideration of past theory and observation, only the mathematical form of the model will be hypothesized. That is to say, the hypothesized density function involves unknown constants, parameters to be evaluated in consideration of observational data not influential in formulation of the incomplete model.

In this system, the principle of maximum likelihood is suggested as being most generally acceptable for the purpose of evaluating or estimating the unknown constants. Given a set of observed points or selections, the individual events are presumed to be independent in the stochastic sense. Whether or not observations or selections are random, it will subsequently be seen that the presumption is justifiable. For each point observed, the observable-event density is expressed as a function of the unevaluated parameters. The product of all such functions is the likelihood function, defined over the parameter space. The product of independent event densities ordinarily yields the density of the model space joint event. However, as a function of the parameters to be estimated, it is said to express the likelihood of the parameter space joint event. The maximum likelihood parameter estimators are given by the coordinates of that point in the parameter space for which the likelihood function is maximized. However, as will be seen, these estimates are not necessarily unique. Following evaluation or estimation of the parameters, the completed model may be compared to alternate theories, using likelihood as the ranking criterion, and the same data in each case.

DESIGNATION OF THE PHENOMENON DESCRIBED BY THE MODEL

Given a particular density function model, there is a question as to whether or not the representation has meaning. It has meaning only to the extent that the phenomenon represented can be otherwise uniquely designated.

We speak of the phenomenon which is the object of our investigation, the object phenomenon, as an entity; thus we imply that its essence is unchanging. In a sense, we presume that the variation represented by the mathematical model is superficial variation, having no influence on the essential character of the object phenomenon. It is this essential unchanging character which must be uniquely designated if the quantitative description of variation is to be unambiguously interpretable. It is therefore reasonable to require, in addition to the specifications of the mathematical model, a characterization, consisting of a set of designating statements or classifications which, whenever or wherever applicable, serve to effect the valid association of the object phenomenon and its mathematical model. Thus, by definition, the object phenomenon occupies the class intersection.

In a given instance, the association is invalid to the extent that the model fails to describe the object phenomenon as designated.

Assuming that the model parameters are so selected, according to the initial data, that a good fit of the model to the observations is obtained, it cannot be said in a later failure, whether it is the quantitative hypothesis, or the designating classifications, which are in error. From one point of view, the quantitative model is inadequate in that a source of excessive variation exists which is not represented. Or alternatively, the original investigator was careless or unlucky in that he failed to completely designate the conditions under which he made his observations.

From another point of view, a phenomenon which is adequately designated, or which is reproducible, is necessarily in a state of control. A phenomenon inadequately designated, or unreproducible, is out of control by definition. However, in the same sense that not all variables of interest are observable, not all significant classifications are obvious. Where a state of control is not existent, it is theoretically up to the investigator as to whether or not he will attempt to account for the unrepresented variation by altering the form of the mathematical model, or by placing further restrictions in the form of classifications on the phenomenon to which the model applies, and thereby narrowing the scope of the investigation.

It is generally the unsatisfactory solution to correct a poor fit by narrowing the scope of the investigation. It is better, if possible, to interpret a poor fit as specification error in the form of the mathematical model. In practice however, the investigator may be deprived of the choice which is rightfully his in theory. If the phenomenon is not in a state of control, that is to say not reproducible, then it is possible that a new mathematical model, so specified in form as to adequately describe the observed variation, will not be useful for the original purposes of the investigation.

From the above discussion, two things are apparent. First, it is essential that models be reexamined under the widest possible range of conditions within the limits imposed by the classifications. In practice, experience with phenomena similarly classified may justify some relaxation of this requirement. Second, the care taken in classifying the object phenomenon and in writing the designating statements, must correspond to the effort expended in developing and establishing the quantitative theory. These remarks apply with particular force in the application of this system, since in its use we aspire to a high order of discrimination.

In conclusion, at least one fundamental condition must apply in use of the system. We require in all cases that the object phenomenon as designated be reproducible or in a <u>state of control</u>. More precisely, we require that the finite set of observed events which are actually

obtained constitute a representative sample of a potentially infinite population of observations; specifically that population which is uniquely associated with the object phenomenon as designated.

DISTRIBUTIONS MAY RESULT FROM EITHER RANDOM OR SYSTEMATIC SELECTION

It has been stated that in this system variables need not be randomly distributed in order to be described by a density function. Before this idea is put to use, it should be discussed in some detail. In a given observational situation, the values of a given variable may be either purposefully selected, or they may occur as a natural consequence of the phenomenon under investigation; however, in either case they may be either randomly or systematically distributed.

The idea of random selection is historically associated with the idea of a lottery, or blind selection after mixing of objects distinguishable only by sight. The important characteristics, for the purpose of this discussion, are first that the objects are contained and consequently the distribution is stable, and second that mixing is employed. To say that the distribution is stable is to say that a state of control exists.

The mixing idea is related to a subjective interpretation of probability as degree of reasonable belief. Participants in the lottery agree that it will be fair if their mutual ignorance of future selections is assured; hence mixing prior to selection, or "random" selection, is required.

If we also impose the requirement of replacement following each selection, then the population of selections becomes potentially infinite and, in consideration of the mixing, the individual selections are stochastically independent. Thus, at least in this case, the usual acceptable conditions are established for construction of a simple likelihood function.

However, in valid application of the system we require that the object phenomenon as designated exist in a state of control. This being the case, the ultimate finite set of selections (or sample) obtained is representative by definition. The additional requirements of random or stochastically independent selection are only incidental means to attainment of that which is already provided. Once we are in possession of a representative sample, the means of its attainment are

immaterial. We can arbitrarily treat the individual selections as having been obtained under conditions of stochastic independence, (and form the likelihood function accordingly) whether or not such was the case. Specifically, systematic or ordered selection is not invalidated, since a representative sample can be reordered or mixed without affecting either its representative character or the ultimate maximum likelihood estimators. However, there is more to be said.

It is interesting to consider some advantages and disadvantages with respect to random versus systematic selection in those situations wherein the investigator has the choice, that is where selection is purposeful. A principal advantage in random selection lies in the simplification of the model that is its frequent accompaniment. There may be variables which are influential in the object phenomenon, but which are not of interest in the investigation. The effects of the unwanted variation may be discounted by uniform random selection of observations over the space of the unwanted variables. For example, suppose an investigator wishes to estimate the mean moisture content of a carload of wheat. Suppose that the car is known to have been loaded in several batches, thereby imposing an internal stratification of unknown structure on the carload. If the structure is of interest, then the spatial coordinate variables should be included in the model, along with hypotheses concerning the form of the interfaces. Nothing is wrong with this approach, except that the model is unnecessarily complicated. In practice, observations would probably be taken at random over the volume of the car, from which the sample mean moisture content would be computed.

The question may now be asked as to whether a systematic uniform selection of observations over the car would not be better. To adequately cover the volume with evenly spaced observations, which are sufficiently close together so as not to miss or otherwise give improper weight to the individual strata, would probably require more observations than would be normally required for a random sample. In another situation, cyclic effects might be adversely influential if the systematic selection is in phase. However, systematic selection can also be designed specifically to either mask or detect cyclic or other systematic effects, particularly where additional information is available. In fact, where economic and physical factors are not influential, it is conceivable in any observational situation requiring purposeful selection, that a design for systematic selection of observations can be found which is superior to random selection. Of course economic and physical factors do intervene. If the investigation involved selection of representative bowls of soup from a large container, it is certainly more reasonable to actually mix the soup prior to ladling, rather than to investigate the spacial distribution of the constituents in order to achieve the desired result through deduction.

There is surely no virtue in ignorance for its own sake; yet random selection is sometimes used when there is no compensating advantage. Consider a distribution of purposefully and systematically selected points designed to closely approximate the normal or Gaussian distribution. If selection is properly done, the points may be plotted as a smooth histogram. However, the distribution of the same number of points, selected at random according to the normal distribution, and plotted as a histogram, generally is a fairly ragged configuration. In a situation where the selected points are used in an estimation process, and wherein confidence in the estimate is increased in proportion to the extent that the selected point set is actually representative of a normal distribution, then the systematically selected point set is to be preferred. The additional noise introduced in random sampling can only serve to increase the variance of the estimate.

In any case, for purposes of the system, no distinction is made between randomly distributed variables and systematically distributed variables. With respect to the concept of probability, and to the extent that it is found to be useful in applying the system, probabilities may be simply interpreted as objectively determined measures of degree of reasonable believability or confirmation. To the extent that application of the system results in the generation of such measures, probability can be said to be operationally defined.

MATHEMATICAL THEORY AND DISCUSSION

In the remainder of this paper, it is the purpose to first discuss the general aspect of model construction in the context of the system, then to investigate several useful model types and to outline the specific steps in their construction, and finally to discuss some aspects of implementation.

In the usual practical application of the system, an investigator is expected to select an appropriate standard model type. For example, if his ultimate goal is the ability to forecast a single random variable event, he will probably choose a regression model. The number of standard model types may be increased without limit, but only the most useful types need be catalogued. Eventually, there should be a large number from which the investigator may choose.

For each standard model type, a computer program should be prepared. Having designated the object phenomenon, formulated the hypotheses, specified the associated mathematical forms, and collected his data, the

investigator employs the appropriate computer program in evaluating the parameters. Resort to automatic machine computation and information processing not only eliminates the drudgery but enforces the desired degree of standardization and objectivity in the analysis. Ideally, the computer should yield a report of the investigation, an integrated presentation of the phenomenon as designated, the data, the completed hypotheses, and other useful information.

Initial Hypotheses

In most model types to be discussed, we distinguish between the phenomenon of primary interest, namely the object phenomenon, and that phenomenon which is the act of observation. The hypothesized true but generally unobservable variables of primary interest, the object variables, are generally denoted

$$x_1, x_2 - - - -, x_n$$

The mathematical representation of the object phenomenon is referred to as "the object model". Included in the object model hypotheses, as initially specified are:

1. Joint, marginal, or conditional density functions which describe the distribution of various subsets of the object variables.
2. Equations relating various subsets of the object variables. (If the variables, as specified free of observational error, are naturally so related, the equations are said to describe structural relationships. In other cases, they may describe prediction or regression relationships.)
3. Inequalities relating various subsets of the object variables. (These relationships serve as constraints on the distribution of object variables. They specify the region boundaries, when such exist, over which the joint density functions are defined. Consequently they are part of the density function specifications.
4. In each of these functions or relationships, unevaluated parameters θ will probably be involved; in which case, known or hypothesized constraints on the θ are also included in the initial specifications.

In most cases, the object variables, being unobservable, are specified free of observational error. Corresponding to the individual object variables,are hypothesized but unobservable errors of observation. The errors are denoted

$$e_1, e_2, - - - -, e_n.$$

The mathematical representation, which by hypothesis describes the distribution of errors of observation, is referred to as "the calibration model". This distribution of errors is generally unknown; but the functional form of the joint error density function, which may or may not involve unevaluated parameters, is hypothesized. It is usually preferable to describe the error distribution as a function of the associated object variables. Consequently, the calibration model is generally specified as a conditional joint error density function, an example of which is here denoted

$$\psi\ (e_1,----,e_n/x_1,----,x_n;\ \Phi)$$

Here, ψ is used to denote density functions generically, while the specific function is unambiguously identified by the arguments. Φ denotes a vector of parameters to be evaluated.

For any fixed point $(x_1,----,x_n)$, the associated joint error density is given by the calibration model. Thus, we say that the object variables enter into the conditional error density function parametrically; that is to say, they are treated as constants.

It is appropriate to mention at this point that, although the density functions are not necessarily probability density functions, they have the same mathematical properties. Given a density function ψ:

1. ψ is defined as event mass per unit of the distributed variable space, for every point in the space of all arguments, including those which enter parametrically.
2. ψ is everywhere positive.
3. Event mass is additive.
4. Total event mass over the distributed variable space is unit mass.

We are therefore free to make use of the calculus of probabilities. Specifically in the situation before us, the conditional joint error density function may be expressed as a ratio of two joint density functions.

$$\psi\ (e_1,----,e_n/x_1,----,x_n;\Phi) = \frac{\psi(e_1,----,e_n,x_1,----,x_n;\theta,\Phi)}{\psi(x_1,----,x_n\ ;\theta)}$$

Here, θ denotes a vector of parameters.

The observable variables, true object variables which are masked or modified by the errors of observation, are defined and denoted as follows:

$$x_{0k} = x_k + e_k;\ k = 1,---,n.$$

We refer to these as "the error equations". The term "error" is understood to signify any influence or set of influences which serves to mask or displace the value of an object variable. For example, an unknown lag associated with a time variable may be thought of as error.

Ultimately, we are interested in an observation model, a joint density function model over the space of observable variables. In application of the system, we propose to obtain the distribution by projecting event mass from some hyperspace which includes the observable space as a subspace. In theory, it makes no difference which hyperspace is used, or how the hyperspace model is initially specified. However in practice, certain procedures are more generally useful. For example, it is found that initially transforming or projecting the calibration model onto a space which includes the observable space is a generally useful procedure.

The Error Transformation

Consider the joint density function $\psi(e_1,---,e_n,x_1,---,x_n)$. Event mass is transformed as probability is transformed; that is to say, in accordance with the calculus of probabilities, the absolute value of the Jacobian of the transformation is used. Using the error equations as the equations of transformation, the transformation is linear, and $/J/$ is always plus one. Thus, the transformation is effected by simple substitution from the error equations into the density function.

$$\psi(e_1,---,e_n,x_1,---,x_n) = \psi[(x_{01}-x_1),---,(x_{0n}-x_n),x_1,---,x_n]$$

$$= \psi(x_{01},---,x_{0n},x_1,---,x_n)$$

Dividing both sides of the resulting equation by the marginal density $\psi(x_1,---,x_n)$, we obtain the desired conditional density function. The transformation is denoted as follows:

$$\psi(e_1,---,e_n/x_1,---,x_n) \rightarrow \psi(x_{01},---,x_{0n}/x_1---,x_n).$$

The geometry of the transformation, for one object variable, is shown in Figure 1.

Surface Density Functions

Another generally useful technique in derivation of the observation model hypothesis consists of eliminating by substitution variables which enter parametrically into conditional density functions. The substitution is valid if the variable to be eliminated is a single-valued function of other variables which enter into the function parametrically. Geometrically, the effect of this operation is in two stages. First, the joint event density function is effectively intersected by the hypersurface of the single valued function; then, the conditional event mass, which is thereby distributed over the hypersurface, is effectively projected in the direction of the variable to be eliminated, onto the subspace of the remaining variables.

For example, suppose that $x_1 = F(x_2,---,x_n; \theta)$. This is a structural relationship between object variables. We also suppose that F is a single-valued function. Here again, θ is a vector of parameters. Given a conditional density function which is defined for all fixed points of the $x_1,----,x_n$ space, the existence of the structural relationship exclusively associates the density function with points on the F-surface, which is imbedded in the $x_1,----,x_n$ space. Since F is single-valued in the x_1 direction, the density function can be mapped one-to-one (projected) onto the $x_2,----,x_n$ subspace. The subspace and the F-surface are of the same dimensionality, of course. It is not necessary that F be single-valued; but structural relationships can usually be so specified, and a mapping which is not one-to-one can result in a significant loss of information.

If the conditional density function is the previously discussed calibration model, the F-intersection is executed as follows:

$$\psi(e_1,---,e_n/x_1,---,x_n;\Phi) = \psi(e_1,---,e_n/F(x_2,---,x_n,\theta),x_2,---,x_n;\Phi)$$

$$= \psi(e_1,---,e_n/x_2,---,x_n;\theta,\Phi;F)$$

Functions of this type may be thought of as generalized or "surface conditional density functions", over the F-surface imbedded in the original $e_1,---,e_n,x_1,x_2,---,x_n$ space.

We denote this transformation in either of two ways: either to the surface conditional density over the F-surface imbedded in the higher original space,

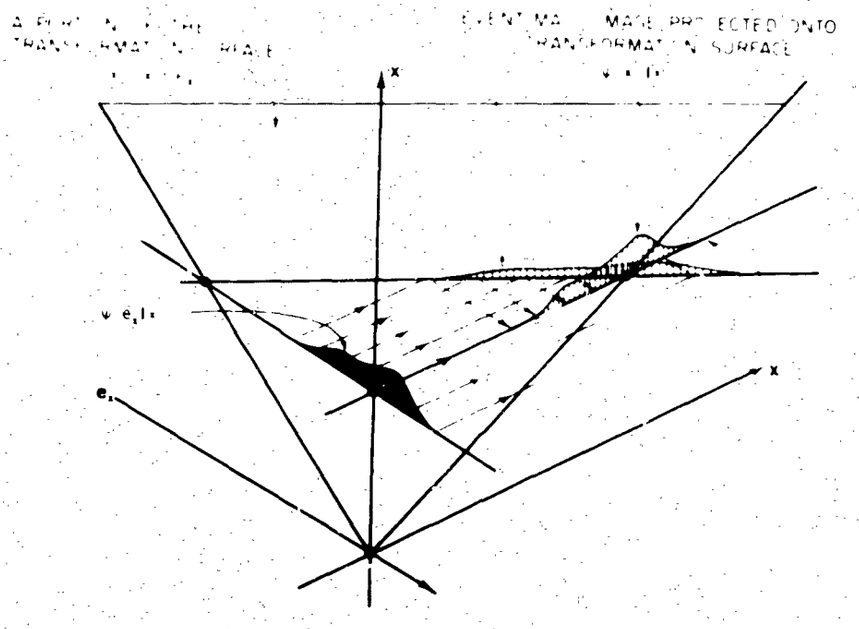


Fig. 1 Illustrating the Geometry of the Error Transformation $\psi(e_x/x) \rightarrow \psi(x_0/x)$

$$\psi(e_1,---,e_n/x_1,---,x_n;\Phi) \rightarrow \psi(e_1,---,e_n/x_2,---,x_n;\theta,\Phi;F)$$

or if no confusion results, by projection to the ordinary conditional density over the subspace.

$$\psi(e_1,---,e_n/x_1,---,x_n;\Phi) \rightarrow \psi(e_1,---,e_n/x_2,---,x_n;\theta,\Phi)$$

The geometry is shown in Figure 2.

Concordant Hypotheses

We now consider an important criterion in the construction of any model. It is necessary that the model components, as initially specified, provide sufficient information for derivation of the observation model hypothesis. In general, the observation model is a joint density function over the space of observable variables. Consider that the number of ways can be very large in which a joint density function can be represented as the product of marginal and conditional density functions. For only three variables, there are thirty ways. Also consider that the observation model joint density function or any of its potentially numerous factors must be obtained from the originally specified model components, some of which are also joint density functions, by an initially unknown series of transformations or projections in accordance with the system concept. Further, the conditions of the investigation usually predetermine some of the initial model specifications; imposed combinations which may be awkward in the analysis. Thus, included in the general problem of hypotheses formation is the not generally insignificant problem of satisfying the sufficient information criterion.

Overspecification, or too much information, is no less of a problem, since it can result in an inconsistent, or at best, an inadequate model. Thus, a set of initially specified model components is required, which is sufficient for derivation of the observation model hypothesis, and which is also necessary to the goals of the investigation. Such a set will be called a "concordant set of hypotheses". A principal problem in model construction is, of course, selection of the appropriate concordant set.

Available Techniques

Having initially specified a concordant set of hypotheses, the stated immediate goal is derivation of the observation model hypothesis. In general, we wish to combine the initially specified model components in such a way as to eliminate all unobservable variables and to retain

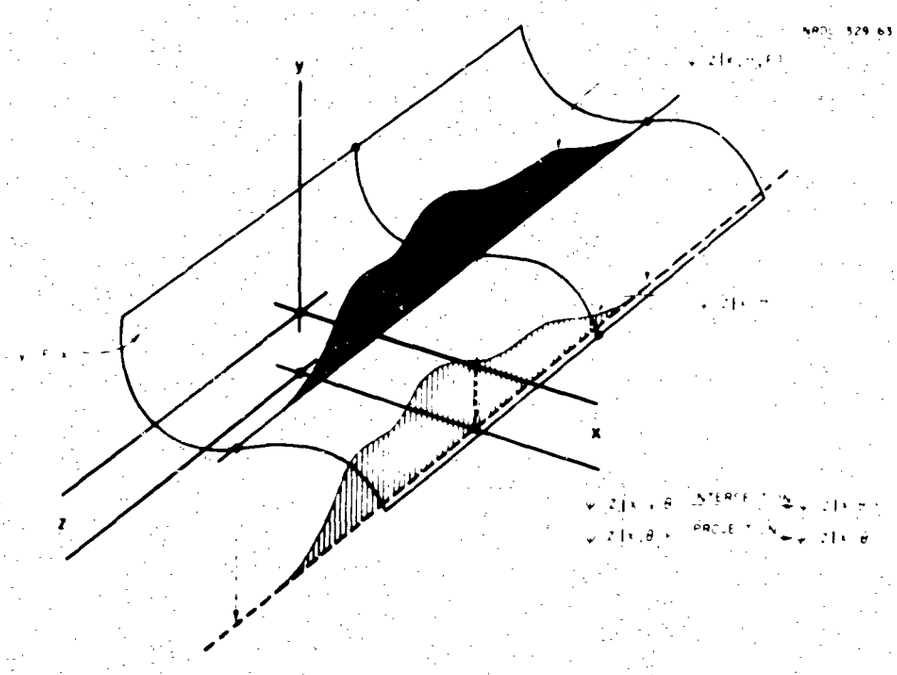


Fig. 2 Illustrating the Intersection-Projection Transformation

$$\psi(Z/x,y;\theta) \xrightarrow{\text{Intersection}} \psi(Z/x;\theta;F)$$

$$\psi(Z/x;\theta;F) \xrightarrow{\text{Projection}} \psi(Z/x;\theta)$$

all observable variables. There are eight manifest techniques which might be used if the occasion demands, and there is a naturally imposed order of precedence in their use:

1. Exclude the unobservable variables from consideration in the first place. This presumes that their influence can be justifiably neglected; otherwise a specification error is introduced thereby.

2. Eliminate unobservable error variables by the one-to-one error equation substitution transformation. Of the eight operations, this induces the greatest gain, for the effort expended.

3. Eliminate fixed or "given" variables of a conditional density function, variables which enter parametrically; using single-valued functions of other given variables as the equations in a one-to-one intersection-projection transformation.

4. Eliminate unwanted variables by using a non-linear one-to-one transformation; mapping into a space of variables not previously involved. This presupposes the prior availability or specification of the appropriate transformation equations, equations representing monotonic surfaces.

5. Use a more complicated variant or generalization of the third technique, whenever the available relationships are not single-valued. It is necessary to partition the surfaces into regions of single-valuedness, and to sum the resultant individual projected event densities. This substitution transformation, being many-to-one, results in some loss of information; however in proper use, this would be the necessary loss incurred in viewing the hypotheses from the limited perspective of the observable space.

6. Use a more complicated variant or generalization of the fourth technique, whenever the available equations represent surfaces which are not monotonic. In such cases, it is necessary to partition the surfaces into regions of monotonicity, such that the resultant transformation is many-to-one, specifically not one-to-many. The resultant individual projected event densities are summed in the new space. This is ordinarily the most general technique applied in transforming density functions. However, it is not always possible to avoid the one-to-many partitioning; for example, the transformation surface may be a hypersphere. There are many interesting ways of dealing with such situations; they all require additional hypotheses. In the case of the hypersphere, projected event mass could be

allocated to the near and far surfaces, either in a fixed ratio or in a ratio which is introduced as an additional parameter to be evaluated. Either specification constitutes an additional hypothesis.

7. Eliminate variables by intersecting a joint density function, and projecting the surface distribution of joint event mass onto an appropriate subspace. This is a further generalization of the third and fifth techniques.

8. Eliminate the influence of unobservable variables from consideration by integration. Sum the event mass over the entire space of unwanted variables, in order to consider only the projected marginal distribution in the subspace of observable variables.

With proper application of these techniques, the observation model hypothesis is derived. However, the observation model is still incomplete, as are the object model and the calibration model, in the sense that there are parameters to be evaluated or estimated.

The parameters are the θ and Φ vectors, of the object model and the calibration model respectively. There are of course other ways in which parameters may be introduced in the hypotheses. Some will be discussed; they are here denoted ξ. Now, given N observations, the data are: $(x_{01i}, x_{02i}, \text{-----}, x_{0ni})$, $i=1, \text{---}, N$. The likelihood function is given by

$$L = \prod_{1}^{N} \psi(x_{01i}, \text{----}, x_{0ni}; \theta, \Phi, \xi).$$

The likelihood is usually maximized by maximizing log L. That is to say, a point is found in the parameter space of θ, Φ, and ξ, subject to possible initially hypothesized constraints, for which log L attains its maximum value. Of course, N must at least equal the number of unevaluated parameters in order that they be determinate. However, we are specifically concerned with those situations wherein observations are not in short supply and the parameter values are consequently overdetermined. Thus, in the presence of sampling variation and specification error, the parameter values must be estimated. Actually, many difficulties can occur in obtaining the maximum likelihood estimates, so the subject is dealt with separately later on, and practical methods are indicated.

The values obtained for the θ, Φ, and ξ may now be used to complete the associated models. The completed object model may now be said to provide the best description of the object phenomenon as designated,

consistent with the hypotheses and the data, according to the principle of maximum likelihood.

Also consider that the error distribution phenomenon can be the phenomenon of primary interest. The maximum likelihood Φ values yield the completed calibration model. However, in cases where the calibration model is constrained by the object model hypotheses, its use should be suspect for points in the space of object variables which are not in the immediate neighborhood of the object phenomenon events.

It is interesting to note, in application of the system generally, that rounding off of observed values to the number of digits known to be or suspected of being significant can result in loss of information which is necessary in precise evaluation of the parameters; in particular where errors are relatively small but significant. Observations should probably be recorded to the maximum precision obtainable from the observational method in use. This is clearly contrary to existing practice in many observational situations.

Purposeful Distributions

As the final topic in this general discussion of model construction, consider that it is frequently necessary to purposely impose a distribution where none occurs as a natural consequence of the phenomenon under investigation. The distinction between purposefully distributed variables and naturally distributed variables is useful in the discussions of various model types to follow. This distinction is solely for the convenience of the investigator, and is of no theoretical or philosophical significance not previously discussed. The act of purposefully distributing or weighting a subset of variables, observable or otherwise, simply imposes on the investigator the requirement that the distribution be separately specified in the object model. An alternative is to consider that the purposeful distribution is a stated condition which serves to designate the object phenomenon. This last procedure necessarily restricts the scope of the investigation; nevertheless, it is sometimes appropriate. For example, consider the situation wherein there are variables in which we have no interest. Then the object phenomenon, as designated, may reasonably involve purposeful uniform point selection, either random or systematic, over the space of the unwanted variables, in order to simplify the model specification.

However, in general, the most useful procedure is to treat the purposeful distribution as a separate but constituent phenomenon, separately specifying the form of the density function and, if necessary, evaluating its parameters either separately or concurrently with the object model parameters. Since, by definition, the functional form is under the direct or indirect control of the investigator, specification error should be minimal if not non-existent.

Purposeful distributions are most likely to be directly observable, although this is not necessarily the case; therefore, it might seem most reasonable to specify purposeful distributions over the space of observable variables. Unfortunately, many useful object model components, taken in conjunction with a directly specified observable distribution, are not likely to form a concordant set of hypotheses. Consequently, it is frequently preferable to specify the purposeful distribution over the space of the associated object variables. An important exception occurs in regression and prediction models, where the specified object variables can also be observable variables.

MODEL TYPES

Type I Model: Simple Measurement with Error

In this first and simplest of all observational situations, the phenomenon under investigation, by hypothesis, is represented by a single point $(\theta_1, \theta_2, \text{----} \theta_n)$ in the space of object variables, the coordinates of which, the θ_i, are the unknown parameters to be evaluated. These hypothesized true but unobservable values are masked by the errors of observation. In other words, we wish to perform a simple act of measurement involving error.

For clarity, but without loss of generality, the system is illustrated first for a Type I model of two object variables. Thus, the six variables of interest which define the model space are: x_1, x_2, the object variables; e_1, e_2, the corresponding errors of observation; and x_{01}, x_{02}, the observable variables.

The object model is given by the equations:

$$x_1 = \theta_1; \quad x_2 = \theta_2$$

These equations are structural relationships. θ_1 and θ_2 are the object model parameters, constants to be evaluated.

The calibration hypothesis is appropriately specified as a conditional joint error density function as follows:

$$\psi(e_1, e_2 / x_1, x_2; \Phi)$$

Again Φ is a vector of unevaluated parameters.

Having formulated the initial hypotheses, the immediate concern is with derivation of the observation model hypothesis. To this end, the initial hypotheses must be combined in such a way as to effect elimination of the unobservable variables, and introduction of the observable variables. Geometrically, we wish to project the combined-model event mass distribution from the model space, or its subspaces, into the subspace of observable variables. First the error equation transformation is applied.

$$\psi(e_1,e_2/x_1,x_2;\Phi) \rightarrow \psi(x_{01},x_{02}/x_1,x_2;\Phi)$$

Thus, the joint distribution of observable values is given for any fixed point (x_1,x_2) in the object variable space. However, by our hypothesis, the object model, the phenomenon of primary interest, is restricted to a single point in the x_1,x_2 space, namely (θ_1,θ_2).

Geometrically, at this stage of the derivation, the model is a joint density function, defined over a plane which is imbedded in the 4-dimensional x_{01},x_{02},x_1,x_2 space. The plane is the 2-dimensional intersection of two 3-dimensional hyperplanes, which are defined for $x_1=\theta_1$ and $x_2=\theta_2$ respectively. The joint density function is given by

$$\psi(x_{01},x_{02};\ \theta_1,\theta_2,\Phi)$$

The expression does not involve x_1 and x_2. Consequently, we choose to ignore the 4-dimensional environment of the distribution; but in doing so we effect the desired one-to-one transformation, or projection into the 2-dimensional subspace of observable variables. Thus, the derivation of the Type I observation model hypothesis is accomplished.

The likelihood function can now be formed, and the parameter estimates obtained. The values obtained are denoted $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\Phi}$. $\hat{\theta}_1$ and $\hat{\theta}_2$ complete the object model. The $\hat{\Phi}$ values obtained, complete the calibration model. However, the calibration model is suspect for points of the object variable space which are not in the immediate neighborhood of $(\hat{\theta}_1,\ \hat{\theta}_2)$.

In conclusion of the Type I model discussion, note that this 2-dimensional derivation is symmetrical with respect to the object variables; consequently, it extends readily to any number of object variables. For simplicity in the discussions to follow, variables which are symmetrically treated in the derivation are represented as vectors. In fact, all model types as represented are extended to any number of

dimensions in consideration of the indicated variables as vectors. Thus, the Type I situation and that of other model types are also presented without explanation in a more compact schematic suitably illustrated for convenient reference.

In the Type I illustration, the conditional error distribution is indicated by contours of equal event density superimposed on the space of the object variables. This superimposition so serves to indicate the appearance of the distribution of observed values over the observable space, which may also be thought of as superimposed; and appropriately in this case, the coordinate axes coincide. This superimposition device will also be used in depicting the situation for other model types.

Type II Model: All Object Variables Distributed

The Type II model is characterized by the fact that all object variables are distributed by hypothesis. That is to say, the object model involves specification only of a joint density function, or its marginal and conditional density function factors. The Type I model is actually a trivial special case, wherein the specified distribution assigns all of the event mass to a single point.

The Type II illustration shows the case of two object variables. The initial hypotheses and the derivation of the observation model hypothesis are both presented. The generality of the vector notation should be considered. Any number of Type II variants are admissible thereby.

1. $x \quad (x_1,---,x_n)$

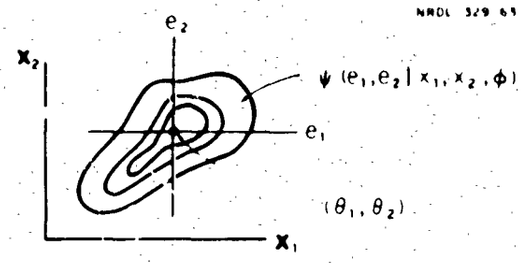
2. $\psi(x) \quad \psi(x_1,---,x_n)$

3. $\psi(x_1,---,x_n)$ need not be specified directly, as a single joint density function; but may be specified in factored form. For example: $\psi(x_1,---,x_k).\psi(x_{k+1},---,x_n/x_1,---,x_k)$.

4. Each of the factors may be either a purposeful distribution or a natural distribution. (In an investigation devoted exclusively to calibration, all factors might be purposefully distributed.)

Of particular interest is that special case wherein a subset x, of object variables, is observable directly without error. That is to say, $e_x = 0$. Let y denote those other object variables which are subject to errors of observation. Then the initial hypotheses are $\psi(x,y;\theta)$ and $\psi(e_y/x,y;\Phi) \quad \psi(e_x,e_y/x,y;\Phi)$. The derivation is as follows:

TYPE I MODEL: SIMPLE MEASUREMENT WITH ERROR



Object Model $x \quad \theta$

Calibration Model $\psi(e_x | x; \phi)$

Derivation of the Observation Model Hypothesis

1. $\psi(e_x | x, \phi) \rightarrow \psi(x_0 | x; \phi)$
2. $\psi(x_0 | x; \phi) \rightarrow \psi(x_0; \theta, \phi)$

Observations $x_{0i} = (x_{01}, x_{02}, \text{------}, x_{0n}), \quad i = 1, \text{-----}, N$

Likelihood

$$L = \prod_{i}^{N} \psi(x_{0i}; \theta, \phi)$$

$$\log L = \sum_{i}^{N} \log \psi(x_{0i}; \theta, \phi)$$

$$(\log L)_{max} = \sum_{i}^{N} \log \psi(x_{0i}; \hat{\theta}, \hat{\phi})$$

$$\psi(e_y/x,y;\Phi) \rightarrow \psi(y_0/x,y;\Phi)$$

$$\psi(x,y;\theta)\cdot\psi(y_0/x,y;\Phi) \rightarrow \psi(y_0,x,y;\theta,\Phi)$$

However, $x \quad x_0$, so that

$$\psi(y_0,x,y;\theta,\Phi) \quad \psi(y_0,x_0,y;\theta,\Phi)$$

and finally

$$\int_{-\infty}^{+\infty} \psi(y_0,x_0,y;\ \theta,\Phi)dy = \psi(x_0,y_0;\theta,\Phi)$$

Type III Model: Structural Relationships

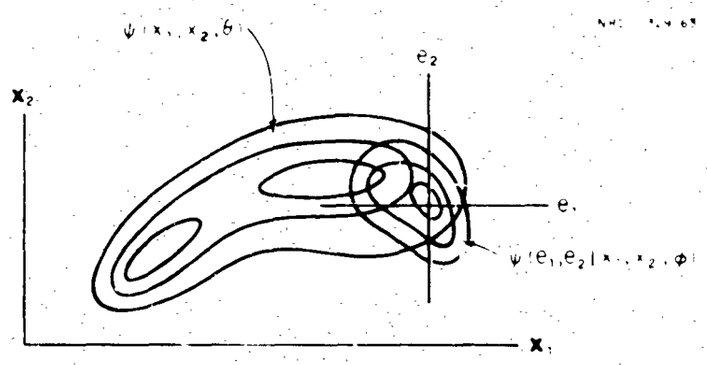
The Type III model object variables are related to each other in the initial hypotheses according to one or more specified equations. It is also stipulated here that the variables, as specified, do not include or involve the errors of observation. As previously stated, such relationships are known as "structural relationships".

The relationships are presumed to be single-valued in their respective y directions, as indicated in the Type III schematic. It is usually the case that such single-valued representation can be specified.

In the intersection operation (Derivation 2), whenever more than one structural relationship is involved, it is of course most expeditious to so order the substitutions that no variable is reintroduced, after being previously eliminated.

In addition to the specified structural relationships and the calibration model, a joint marginal distribution $\psi(x)$, or its constituent factors must be hypothesized, as shown in the Type III illustration. This marginal distribution specification is necessary in order to obtain the concordant set of hypotheses. In the usual situation, we would expect that all of the x object variables would be purposefully distributed; but this need not be the case. Components of this marginal distribution may occur naturally and be of interest in the investigation.

TYPE II MODEL: ALL OBJECT VARIABLES DISTRIBUTED



Object Model $\psi(x, \theta)$

Calibration Model $\psi(e_x \mid x; \Phi)$

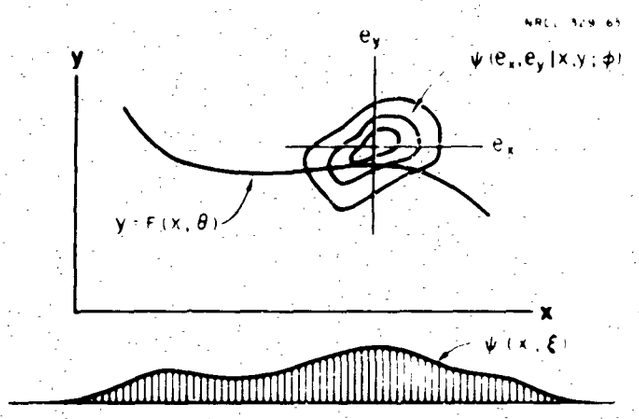
Derivation of the Observation Model Hypothesis

1. $\psi(e_x \mid x; \Phi) \rightarrow \psi(x_0 \mid x; \Phi)$
2. $\psi(x; \theta) \cdot \psi(x_0 \mid x; \Phi) = \psi(x_0, x; \theta, \Phi)$
3. $\int_{-\infty}^{+\infty} \psi(x_0, x; \theta, \Phi) dx = \psi(x_0; \theta, \Phi)$

Likelihood

$$(\log L)_{max} = \sum_{i}^{N} \log \psi(x_{0i}, \hat{\theta}, \hat{\Phi})$$

TYPE III MODEL: STRUCTURAL RELATIONSHIPS



Object Model

1. *Joint Marginal Density Function* $\psi(x; \xi)$
2. *Structural Relationships* $y = F(x, \theta)$

Calibration Model $\psi(e_x, e_y | x, y; \phi)$

Derivation of the Observation Model Hypothesis

1. $\psi(e_x, e_y | x, y; \phi) \rightarrow \psi(x_0, y_0 | x, y; \phi)$
2. $\psi(x_0, y_0 | x, y, \phi) \rightarrow \psi(x_0, y_0 | x; \theta, \phi; F)$
3. $\psi(x; \xi) \quad \psi(x_0, y_0 | x; \theta, \phi) = \psi(x_0, y_0, x; \theta, \phi, \xi)$
4. $\int_{-\infty}^{+\infty} \psi(x_0, y_0, x; \theta, \phi, \xi) dx = \psi(x_0, y_0; \theta, \phi, \xi)$

Prediction

Frequently, there is less interest in the theoretical structural relationship which underlies the errors of observation, than there is interest in predicting observed values y_0 for given observed values x_0. In such cases, the usual practice is to hypothesize no error in x. Thus

$$\psi(e_x, e_y/x,y;\Phi) \quad \psi(e_y/x,y;\Phi)$$

$$\psi(e_y/x,y;\Phi) \quad \rightarrow \psi(y_0/x,y;\Phi).$$

$$\psi(y_0,x,y;\Phi) \quad \rightarrow \psi(y_0/x;\theta,\Phi;F)$$

$$\psi(x;\xi)\cdot\psi(y_0/x;\theta,\Phi) \quad = \quad \psi(x,y_0;\theta,\Phi,\xi)$$

but $x \quad x_0$; so

$$\psi(x,y_0;\theta,\Phi,\xi) \qquad \psi(x_0,y_0;\theta,\Phi,\xi)$$

The prediction function is then given by

$$\bar{Y}_0(x_0;\hat{\theta},\hat{\Phi}) = \int_{-\infty}^{+\infty} y_0 \, \psi(y_0/x_0;\hat{\theta},\hat{\Phi})dy_0$$

This is ordinarily considered to be a kind of regression; but here we call it "prediction". The name "regression" is reserved here for cases involving no underlying structural relationship.

Type IV Model: Regression on Observable Variables

In those analyses wherein all variables are distributed under the initial hypotheses, and wherein it is desired to describe the most representative or mean values of each of a subset of observable variables y_0 as single-valued functions of the remaining observable variables $x \quad x_0$ (i.e., the x are free of observational error), we will say that the development is a case of "regression". We say that the single-valued functions are "regression relationships". They are defined as follows:

$$\bar{y}_0 = F(x_0,\hat{\theta}) = \int_{-\infty}^{+\infty} y_0 \, \psi(y_0/x_0;\, \hat{\theta})dy_0$$

The necessary constituents of the object model are a joint conditional density function, which in general is given by $\psi(y/x;\theta)$, and the structural relationships $x = x_0$ which are also included as a consequence of the model type definition. These necessary constituents together with the calibration model completely determine the regression relationship set $\bar{y}_0 = F(x_0,\theta)$. Consequently, the problem of selecting a concordant set of initial hypotheses is not trivial. The derivation of the observation model hypothesis is shown in the schematic.

The joint marginal density function $\psi(x_0;\xi)$ of the directly observable subset x_0 or any of its component factors may or may not be of theoretical interest. Since this distribution enters into the observation model simply as a factor, and since it is directly observable and therefore may be purposefully and completely specified initially, it (or any of its component factors) can be entered into log L as an additive constant. As such, it has no influence in the maximization, or ultimately on the values $\hat{\theta}$ and $\hat{\Phi}$.

In common application of regression analysis, no calibration model is specified; there is no error in the y variables and therefore

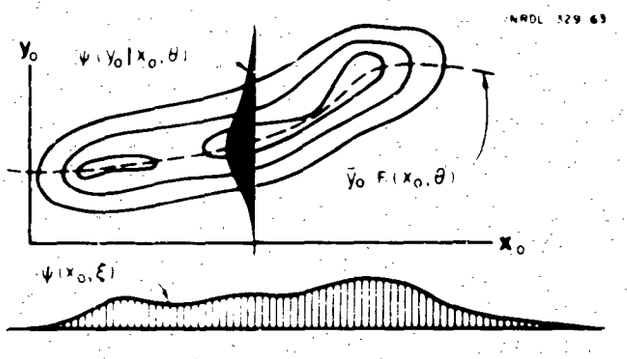
$$\psi(y/x_0;\theta) = \psi(y_0/x_0,\theta)$$

Also commonly, this conditional density function is initially hypothesized for a single y_0 variable and is usually specified as normal:

$$N(y_0,\, \sigma_{y_0}) = N\,[F(x_0,\theta),\, \sigma_{y_0/x_0}]$$

Further, in ordinary application of regression analysis, the observable joint marginal distribution is not of theoretical interest. Also σ_{y_0} may be thought of as constant and ignored. Thus any specified functional form for the regression relationship is concordant, and the process reduces to curve fitting according to the principle of least squares:

TYPE IV MODEL: REGRESSION ON OBSERVABLE VARIABLES



Object Model

1. *Structural Relationships* $x \equiv x_0$
2. *Joint Conditional Density Function* $\psi(y|x,\theta)$
3. *Observed Joint Marginal Density Function* $\psi(x_0,\xi)$
4. *Regression Relationships* $\bar{y}_0 = F(x_0,\theta) = \int_{-\infty}^{+\infty} y_0 \psi(y_0|x_0,\theta)dy_0$

Calibration Model $\psi(e_x, e_y|x,y;\Phi) \equiv \psi(e_y|x,y;\Phi)$

Derivation of the Observation Model Hypothesis

1. $\psi(e_y|x,y;\Phi) \rightarrow \psi(y_0|x,y;\Phi) \rightarrow \psi(y_0|x_0,y;\Phi)$
2. $\psi(y|x;\theta) \rightarrow \psi(y|x_0;\theta)$
3. $\psi(x_0;\xi)\cdot\psi(y|x_0;\theta) = \psi(x_0,y;\theta,\xi)$
4. $\psi(x_0,y;\theta,\xi)\cdot\psi(y_0|x_0,y;\Phi) = \psi(x_0,y_0,y;\theta,\Phi,\xi)$
5. $\int_{-\infty}^{+\infty}\psi(x_0,y_0,y;\theta,\Phi,\xi)dy = \psi(x_0,y_0;\theta,\Phi,\xi)$
6. $\psi(x_0,y_0;\theta,\Phi,\xi) = \psi(x_0;\xi)\cdot\psi(y_0|x_0;\theta,\Phi)$

Likelihood

$$\log L = \left[\sum_i^N \log\psi(x_{0i};\xi)\right] + \left[\sum_i^N \log\psi(y_{0i}|x_{0i};\theta,\Phi)\right]$$

$$(\log L)_{max} = \sum_{i}^{N} \log \left[\frac{1}{\hat{\sigma}_{y_0} \sqrt{2\pi}} \; e^{-\frac{[y_{0i} - F(x_{0i}, \hat{\theta})]^2}{2 \hat{\sigma}_{y_0}^2}} \right]$$

or

$$Q(\log L)_{min} = \sum_{i}^{N} [y_{0i} - F(x_{0i}, \hat{\theta})]^2 \text{ , for constant } \sigma_{y_0}.$$

Type V Model: Discrete Distributions

In general, for the various model types discussed, the indicated distributions need not be continuous. They may be specified only for a finite number of values of certain of the variables. Suppose that certain variables $x \equiv (x_1, ---, x_k, ---, x_m)$ by hypothesis assume a number $q(k)$ of values ξ_{kj}. That is to say, for each x_k,

$$x_k = \xi_{kj} \; ; \; j = 1, ----, q(k)$$

In the manner of the Type I model, these are structural relationships. However, since each variable assumes a multiplicity of values, the individual surface distributions of event mass must be summed in projection, according to the fifth of the previously listed available techniques.

For example, consider a Type II model. Suppose that

$$\psi(x,y;\theta,\xi) = \psi(x;\xi) \cdot \psi(y/x;\theta)$$

where

$$\psi(x;\xi) \equiv \psi(x_j;\xi_j) = \eta_j; \; j = 1, ----, q$$

Here, the ξ_j are the discrete set of x values, which may or may not be known initially; and the η_j are the corresponding set of marginal densities, which may or may not be known initially. Thus,

$$\psi(x,y;\theta,\eta_j,\xi) = \eta_j \cdot \psi(y/x_j;\theta);$$

ξ indicating implicit parameters. Referring to the Type II schematic, we also have

$$\psi(x_0,y_0/x,y;\Phi) \qquad \psi(x_0,y_0/x_j,y;\Phi)$$

Consequently,

$$\psi(x_0,y_0,x_j,y;\theta,\Phi,\eta_j,\xi) = \eta_j \cdot \psi(y/x_j;\theta) \cdot \psi(x_0,y_0/x_j,y;\Phi)$$

However, we do not proceed exactly according to the Type II derivation; given the q structural relationships, which geometrically are represented by constant hyperplanes, the situation calls for a series of q intersections of the joint event mass distribution defined by the above joint density function. This is in application of the seventh of the previously listed available techniques. The resulting set of q surface joint density functions is given by

$$\psi(x_0,y_0,y;\theta,\Phi,\eta_j,\xi_j;x_j) \ , \ j = 1,----,q.$$

As stated, the associated projection transformation is many-to-one and the densities must be summed in the subspace. Thus we proceed.

$$\psi(x_0,y_0,x_j,y;\theta,\Phi,\eta_j,\xi) \rightarrow \psi(x_0,y_0,y;\theta,\Phi,\xi,\eta)$$

where

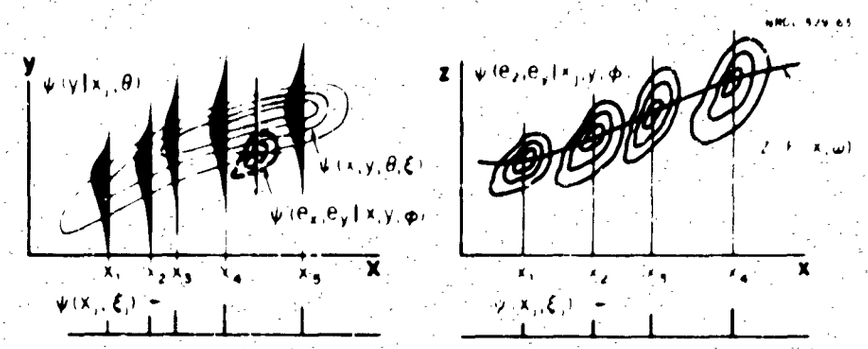
$$\psi(x_0,y_0,y;\theta,\Phi,\xi,\eta) = \sum_{j=1}^{q} \eta_j \cdot \psi(y;\theta,\xi_j) \cdot \psi(x_0,y_0/y;\Phi,\xi_j)$$

The last step is simply to project into the observable space:

$$\psi(x_0,y_0;\theta,\Phi,\xi,\eta) = \int_{-\infty}^{+\infty} \cdot \left[\sum_{j=1}^{q} \eta_j \cdot \psi(y;\theta,\xi_j) \cdot \psi(x_0,y_0/y;\Phi,\xi_j) \right] dy$$

In some situations, it is expected that other nonconstant structural relationships will also be specified in the manner of the Type II model. The effect, in such cases, is simply to require the additional associated intersection-projection substitution transformations in the derivation process, provision for which is indicated in the Type V schematic. The additional structured variables are denoted Z. Two specialized Type V situations are shown in the illustrations.

TYPE V MODEL: DISCRETE DISTRIBUTIONS



Object Model

1. *Joint Marginal Discrete Density Functions*

   $\psi(x;\xi) = \psi(x_j;\xi_j) = \eta_j \; ; \; j = 1,----,q$

2. *Associated x Value Discrete Set, Constant Structural Relationships*

   $x_k = \xi_{kj} \; ; \; j = 1,----,q(k)$

3. *Additional Nonconstant Structural Relationships* $\quad z = F(x_j,y,\omega)$

4. *Joint Conditional Continuous Density Function* $\quad \psi(y|x_j,\theta)$

Calibration Model $\quad \psi(e_x,e_y,e_z|x_j,y,z;\Phi)$

Derivation of the Observation Model Hypothesis

1. $\eta_j \cdot \psi(y|x_j;\theta) = \psi(x_j,y;\theta,\eta_j)$

2. $\psi(e_x,e_y,e_z|x_j,y,z;\Phi) \rightarrow \psi(x_0,y_0,z_0|x_j,y,z;\Phi)$

3. $\psi(x_0,y_0,z_0|x_j,y,z;\Phi) \rightarrow \psi(x_0,y_0,z_0|x_j,y;\Phi,\omega;F)$

4. $\psi(x_j,y;\theta,\eta_j) \cdot \psi(x_0,y_0,z_0|x_j,y;\Phi,\omega) = \psi(x_0,y_0,z_0,x_j,y;\theta,\Phi,\omega,\eta_j)$

5. $\psi(x_0,y_0,z_0,x_j,y;\theta,\Phi,\omega,\eta_j) \rightarrow \psi(x_0,y_0,z_0,y;\theta,\Phi,\omega,\eta_j,\xi_j)$

6. $\sum_{j=1}^{q} \psi(x_0,y_0,z_0,y;\theta,\Phi,\omega,\eta_j,\xi_j) = \psi(x_0,y_0,z_0,y;\theta,\Phi,\omega,\eta,\xi)$

7. $\int_{-\infty}^{+\infty} \psi(x_0,y_0,z_0,y;\theta,\Phi,\omega,\eta,\xi)dy = \psi(x_0,y_0,z_0;\theta,\Phi,\omega,\eta,\xi$

8. or $\psi(x_0,y_0,z_0;\theta,\Phi,\omega,\eta,\xi) = \sum_{j=1}^{q} \eta_j \psi(x_0,y_0,z_0|x_j=\xi_j;\theta,\Phi,\omega)$

Attribute Distributions

The generality of the Type V model is considerably extended if we consider that the x need not be quantitative variables but that the discrete "values" they assume may be qualitative. That is to say, the x_j may represent attributes or classifications, of which the relative likelihoods or weights, the η_j, are under investigation. In such a case, the ξ_j are merely implicit qualitative parameters. Thus, the observation model hypothesis is given by

$$\psi(x,y_0,z_0;\theta,\Phi,\omega,\eta) = \int_{-\infty}^{+\infty} \left[\sum_{j=1}^{q} \eta_j \cdot \psi_j(y;\theta_j) \cdot \psi_j(y_0,z_0/y;\Phi_j,\omega_j) \right] dy$$

or

$$\psi(x,y_0,z_0;\theta,\Phi,\omega,\eta) = \sum_{j=1}^{q} \eta_j \cdot \psi_j(y_0,z_0;\theta_j,\Phi_j,\omega_j)$$

Type VI Model: Constrained Distributions

As previously stated, inequalities which are specified in the object model describe region boundaries of the object model distributions. For this reason, we refer to these distributions as "constrained distributions"; and the ordering relationships are called "constraints".

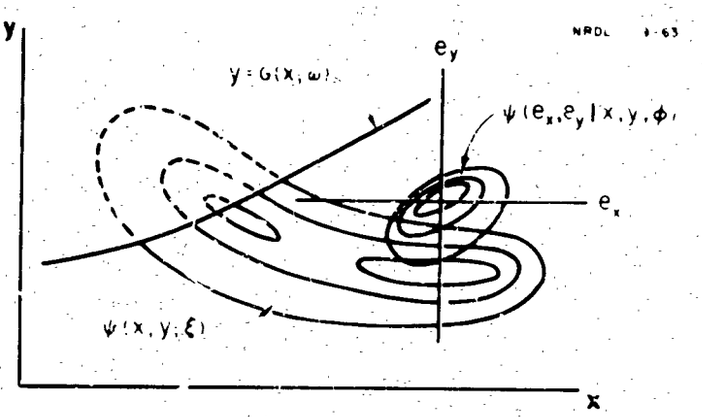
The constrained distribution models which are considered here may also involve structural relationships. Such situations are dealt with in the derivation schematic, but are not included in the Type VI illustration.

The constraining surfaces are denoted $Y = G(x,\omega)$. In general the requirement that parameters ω be evaluated, introduces great difficulty in computation with numerical methods currently available. However, for constant surfaces $Y=\omega$. there is no particular trouble. Also, the problem is resolved completely if the integration can be performed analytically.

Type VII Model: Variform or Discontinuous Models

For models of this type, various regions of the model space are represented by constituent elements of the model which are distinct in mathematical form. The regions are separated by specified constraining surfaces. The constraining surfaces in this case are called "shock fronts", or in the one-dimensional case, "shock points". For example, a melting point or a boiling point is a shock point; in meteorology, storm fronts or air mass boundaries are shock fronts.

TYPE VI MODEL: CONSTRAINED DISTRIBUTIONS



Object Model

1. *Joint Density Function* $\psi(x,y;\xi)$
2. *Structural Relationships* $z = F(x,y,\theta)$
3. *Constraints* $y = G(x,\omega)$

Calibration Model $\psi(e_x,e_y,e_z|x,y,z;\Phi)$

Derivation of the Observation Model Hypothesis

1. $\psi(e_x,e_y,e_z|x,y,z;\Phi) \rightarrow \psi(x_0,y_0,z_0|x,y,z;\Phi)$
2. $\psi(x_0,y_0,z_0|x,y,z,\Phi) \rightarrow \psi(x_0,y_0,z_0|x,y;\theta,\Phi;F)$
3. $\psi(x,y;\xi) \cdot \psi(x_0,y_0,z_0|x,y;\theta,\Phi) = \psi(x_0,y_0,z_0,x,y;\theta,\Phi,\xi)$
4. $\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{G(x,\omega)}\psi(x_0,y_0,z_0,x,y;\theta,\Phi,\xi)dy\right]dx = \psi(x_0,y_0,z_0;\theta,\Phi,\xi,\omega)$

Two Type VII illustrations are used, to show examples of both change in the form of joint density functions, and change in the form of structural relationships. A more compact notation than that of the previous model type has been adopted in the derivation schematic for convenience.

The same computational difficulty, as in the previous model type, may be expected here if the ω require evaluation.

Indirect Observation

First consider a simple measurement situation, such as that of the Type I model; except that one variable, which is not subject to observation, with or without the influence of errors of observation, is determined only as a specified single-valued function of the remaining variables. Under this hypothesis, all of the variables assume constant values, the determination of which is the object of our effort; but specifically we wish to obtain that numerical value which is "only indirectly observable".

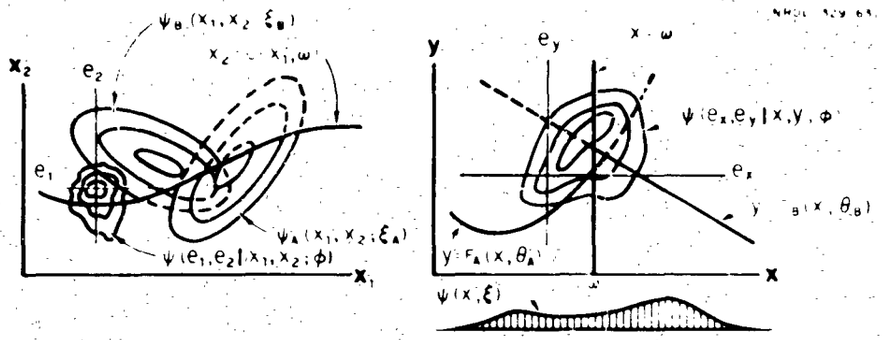
It is helpful to refer to the Type I illustration and schematic, in observing that no amount of observation can provide information about the single-valued functional relationship. The functional form must be specified completely in the initial hypothesis. However, estimators of the hypothesized constant values of those variables subject to observation (and also error of observation) are obtainable through application of the Type I model. But most important, it is known that single-valued functions of maximum likelihood estimators are also maximum likelihood estimators. Consequently, the desired maximum likelihood estimator, for that constant not subject to observation with or without error, is obtained indirectly by simple substitution of the available estimators into the completely specified single-valued function.

Similar considerations apply in mapping distributions, one-to-one or many-to-one, from spaces of which all variables are subject to observation with or without error, onto spaces the variables of which include one or more which are only indirectly observable. The ordinary transformation of coordinates is invoked, and parameters of the new distribution, which are determined thereby, retain the maximum likelihood properties of the primary estimators. However, the equations of transformation must be completely specified.

Other Model Types

The previously discussed model types no doubt encompass a significant fraction of observational situations likely to arise; however,

TYPE VII MODEL: VARIFORM OR DISCONTINUOUS MODELS



Object Model *(For simplicity, only two forms are indicated.)*

Form A	1.	*Joint Marginal Density Function*	$\psi_A(x; \xi_A)$
	2.	*Structural Relationships*	$y = F_A(x, \theta_A)$
Form B	1.	*Joint Marginal Density Function*	$\psi_B(x, \xi_B)$
	2.	*Structural Relationships*	$y \quad F_B(x, \theta_B)$
Shock Front		*(In general more than one required.)*	$x_2 = G(x_1, \omega)$

Calibration Model $\psi(e_x, e_y | x, y; \Phi)$

Derivation of the Observation Model Hypothesis

1. $\psi(e_x, e_y | x, y; \Phi) \rightarrow \psi(x_0, y_0 | x, y; \Phi)$

2.A $\psi(x_0, y_0 | x, y; \Phi) \rightarrow \psi(x_0, y_0 | x; \theta_A, \Phi; F_A)$

2.B. $\psi(x_0, y_0 | x, y; \Phi) \rightarrow \psi(x_0, y_0 | x; \theta_B, \Phi; F_B)$

3.A. $\psi_A(x; \xi_A) \cdot \psi_A(x_0, y_0 | x; \theta_A, \Phi) = \psi_A(x_0, y_0, x; \theta_A, \Phi, \xi_A)$

3.B. $\psi_B(x; \xi_B) \cdot \psi_B(x_0, y_0 | x; \theta_B, \Phi) = \psi_B(x_0, y_0, x; \theta_B, \Phi, \xi_B)$

4. $$\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{G(x_1, \omega)} \psi_A(x_0, y_0, x_1, x_2; \theta_A, \Phi, \xi_A) dx_2 \right] dx_1$$

$$+ \int_{-\infty}^{+\infty} \left[\int_{G(x_1, \omega)}^{+\infty} \psi_B(x_0, y_0, x_1, x_2; \theta_B, \Phi, \xi_B) dx_2 \right] dx_1 = \psi(x_0, y_0, \theta_A, \theta_B, \Phi, \xi_A, \xi_B, \omega)$$

many situations can arise to which the available model types do not readily conform. For example, the various models may be combined in configurations of great complexity; also, it will probably be expedient to select certain special cases for particular emphasis. To a lesser extent, it may be useful to strive for greater generality. Hopefully, the number of available useful model types will grow continually as experience is gained in application of the system.

In a large class of observational situations, not all events which are potentially observable are actually observed. Thus in general, there exists an "attenuation function" or "efficiency of observation function", α, defined over the space of object variables. In such cases, the object variable joint density function becomes

$$\psi'(x) \quad \alpha(x) \cdot \psi(x)$$

As a final remark, in combining the various model types into more complex configurations, whenever possible it is better to evaluate parameters of all constituent hypotheses at the same time, in a single observation model.

PROBLEMS IN COMPUTATION

So far we have carefully sidestepped the possible problems of computation. In construction of the observational model, difficulties can arise in nonlinear transformations and projections requiring integration. Following construction of the observation hypothesis and the likelihood function, the problem of maximizing log L is encountered.

With respect to the problem of integration, presently available methods appear to be inadequate for the task at hand. The most useful system applications will probably involve integration over regions of many dimensions. Classical methods are accurate but require excessive computation in the higher dimensional cases. Monte Carlo methods are probably too imprecise for most system applications.

For maximizing the likelihood, there are several useful techniques available; principally the Newton-Raphson method for solution of simultaneous nonlinear equations, and direct search methods such as the method of steepest ascent. However, it is more probable that new, more powerful methods will be required.

It is our stated purpose to employ automatic computing machines for most expeditious application of the system techniques. Thus, although analytic solutions may be feasible in some cases, numerical methods are to be preferred for the generality of their application in automatic computation. However, application of numerical methods for integration in the model derivation and for maximization of the likelihood imposes the requirement that the two (or more) techniques must be effected simultaneously. This compounds the difficulties considerably.

These considerations have led this author into an investigation of numerical methods suitable for use in the system applications. The basic idea is to partition the multi-dimensional region of interest by a series of cuts. As the cutting proceeds the region is represented by a number (usually large) of subregions. Prior to each cut, the subregion to be cut and the direction of cutting are selected according to appropriate indices. Alternatively, an index for termination of cutting is used in place of the subregion selection index. The resulting subregions are then represented by a selected point set. Operations on this representative point set can then be performed to effect the solution of a wide variety of multivariable numerical analysis problems, including integration and direct search optimization.

The principal work to date has been in the area of multiple integration. The problem of non-variables-separable functions has been resolved by selecting as the representative point for each subregion, the intersection of (n-1)-dimensional mean value surfaces, where n is the dimensionality of the region of integration. These methods have met with considerable success. The method of integration is presently programmed for an IBM-704 computer with a 4000-word magnetic core storage. Multiple integrations can be performed for functions of up to ten variables. For example, the integral over the unit interval of $\mathrm{Exp}(x_1 \cdot x_2 \cdot x_3 \cdot x_4)$ has been obtained to better than six significant digits of precision in about seven minutes. For this result, 32,455 points and 1,930 subregions were required. Considerable improvement is expected if more core storage is available; in particular with respect to increasing the speed and the upper limit on the number of variables.

With respect to the problem of simultaneous integration and maximization, developmental work is in progress. The approach used is to generate a representative set of points as above but over a higher dimensional space, which includes as subspaces both the space of variables over which integration is to take place and the parameter space over which maximization of the likelihood is to take place. In the parameter space, the selected index for cutting is the likelihood value or its logarithm. Subregion midpoints may be used as representative points in this space. Integration is best held to relatively low precision, except for points in the immediate neighborhood of likelihood maximum points.

For the purpose of exploring multidimensional surfaces, such as the likelihood surface, a representative point set is particularly valuable. Localized maxima or saddle points are not likely to be mistaken for the absolute maxima. Also, singularities and discontinuities tend to be readily detected. Such a characterization of the likelihood surface should probably be reported out as auxiliary information. It could find use in future decision theoretic applications of the model as developed.

In any case, it is hoped that the capability for solution of the system computation problems will soon be available. This author's work* is in preparation. Perhaps the main point to be made at this time is not that the methods proposed here for computation will necessarily be those which are ultimately best for the general system application, but that whatever methods are employed, they must be based on computer-oriented mathematics.

Specifically, the methods to be used in computation must not be based on pencil and paper mathematics. It is inconceivable that investigators can ever hope to deal with problems at a practical level of complexity, if they are to be restricted to the present-day pencil and paper analytic techniques. One is led to believe that a new kind of mathematical analysis must evolve, which is exclusively devoted to computer-oriented mathematics.

Also, if our goal is to see our new mathematical methods actually put to use on a broad scale, it is necessary that most of the labor in application be removed. Investigators generally do not aspire to be also mathematicians. They will not be disposed to accept and apply elaborate new procedures, unless they are also attractively packaged and automated.

UNIDENTIFIABILITY

Throughout this paper we have emphasized the Basic Principle of projection into the observable variable space. At this point, we consider the consequences with regard to that information which is necessarily lost in projection.

In the usual practical situation, it is not possible to uniquely determine (or estimate) some subset of the real event model parameter

*See "A Methodology for Numerical Analysis of Functions of Many Variables, with Emphasis on Multiple Integration", by J. W. Hendricks. (Technical Report to be published).

values. A parameter is said to be "identifiable" if and only if a unique value or its consistent estimate is obtainable, given full knowledge of the distribution of observations. In general, unidentifiability is due to the loss of information in projecting the distribution of real events into the apparent event space of observable variables. However, in trivial instances, a case of unidentifiability in effect can be caused by specification of superfluous parameters.

The concept of identification may be more clearly explained by means of an example. In considering the error equations, it is evident that for each observed event x_0, there is an infinite class of possibly true event pairs (x,e_x) which could serve to explain the observed event. It follows that for any given distribution of observed events $\psi(x_0)$, there is an infinite class of joint density functions $\psi(x,e_x)$ which could have served to generate $\psi(x_0)$. In general, in attempting to identify the true joint density function $\psi^*(x,e_x)$, we may or may not be concerned with parameter evaluation.

However, in this system we are committed to parametric representation of the initial hypotheses. The functional forms hypothesized for $\psi(x,\theta)$ and $\psi(e_x/x;\Phi)$ serve to restrict the infinite class of density functions $\psi(x,e_x;\theta,\Phi)$ from which the given $\psi(x_0;\theta.\Phi)$ could have been generated. Individual members of the class are identified by unique values of the θ and Φ. To state that the true joint density function $\psi^*(x,e_x;\theta^*.\Phi^*)$ is identifiable, is to state that the true parameter values θ^* and Φ^* are all individually identifiable.

Let the θ^* and Φ^* be exclusively location or translation parameters (i.e., not shape parameters) of a true structural relationship F^*, and of an associated conditional error distribution respectively. These θ^* and Φ^* are particularly susceptible to unidentifiability for the reasons given. Consider the situation of Figure 3, where a large error bias, uniformly applied over the x,y space, is indicated. Observe that erroneous values of θ,namely θ' in the illustration are also compatible with the observations under the assumption of little or no translation (or bias) of the error distribution.

In such a situation, the false hypothesis F' would, for the indicated observations, actually represent an equal likelihood alternative to the true structural relationship F^*. F^* and F' are of course only two examples of the infinite family of structures which would represent equal likelihood alternatives in this case.

Bias of course is not necessarily constant over the space; there may be more complex interaction. Also, it is not only the location parameters which interact. However, unidentifiability associated with interaction between the object phenomenon and the error phenomenon is but one

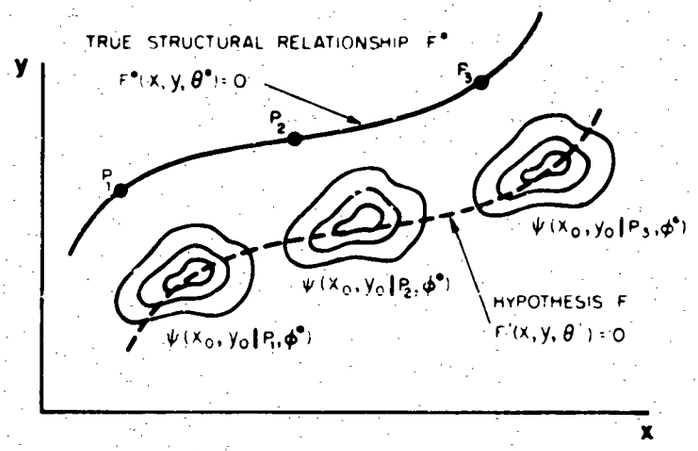


Fig. 3 Illustrating Unidentifiability of Location Parameters

case in point, although perhaps the most important case. Either the object phenomenon or the error phenomenon can be of themselves inherently unidentifiable.

With respect to identification, we are not at all concerned with the problem of estimation. Whether or not the number of observations (or sample size) is large enough to indicate that the parameters are jointly determinate or over-determinate, it still may not be possible to identify some subset (or even any) of the individual specified parameters. However, it should be stated that this is not necessarily an unhappy state of affairs.

INTERPRETATION OF RESULTS

Surfaces of Unidentifiability

Thus the question of identifiability arises for models generally, and in particular for individual parameters. Unidentifiability is evidenced by the fact that the likelihood function in such cases possesses no unique maximum. However in general, all pertinent parameter information is contained in the relationship between parameters which is implied by the locus of equal likelihood maximum points in the parameter space, namely

$$L(\theta,\phi) = L_{max}.$$

Thus at least, the functional relationship between individually unidentifiable parameters is always identifiable in this sense.

Loci of likelihood maximum points, here called "surfaces of unidentifiability", are likely to be multi-dimensional surfaces. For simplicity in discussion, consider only monotonic surfaces. Regardless of the dimensionality of the parameter space, the dimensionality of such a surface is equal to the number of parameters which must be evaluated, arbitrarily or otherwise, in order to establish a unique maximum likelihood point in the parameter space. However, it is neither necessary nor desirable that parameter values be assigned arbitrarily, thereby imposing information which is unjustified by either observation or theory.

It should be remembered that our ultimate goal is model development. To the extent that non-unique parameter estimates constitute an unsatisfactory conclusion to the investigation, the investigator may wish to modify his specified hypotheses or take some other appropriate action. This he is uniquely qualified to do. He alone may possess valid auxiliary information, not previously included in the specifications; and he alone establishes the goals and criteria of the investigation. For example, he may know of no mechanism involved in observation which could justify the presumption of bias ($E(e_x) \neq 0$). He might therefore choose to assume no error bias (using a criterion of maximum simplicity).

A system requirement of no error bias would tend to dispose of unidentifiability due to interaction between the calibration and object models. However, it is specifically not the object of this discussion to require, or even investigate, conditions for complete identifiability of either models or individual parameters. In no case is it reasonable

that, for computational or analytic convenience, we impose assumptions and conditions which are not realistic or generally acceptable in scientific investigation.

Thus, characterization of the set of all equal likelihood maximum points in the parameter space is the appropriate ultimate conclusion in any single application of this system. Further, it is essential that the computer output be of such a form as to facilitate the investigator's interpretation and use of the results. The characterization can take the form of:

1. A unique evaluation or estimate for each parameter.
2. A set of functional relationships between parameters which are individually unidentifiable.
3. A representative subset of the likelihood maximum points.
4. A combination of these.

With respect to computer output, in some simple cases, tabulation may suffice. In other cases, graphical or analytic representation of the surface of unidentifiability would be required. Analytic representation might possibly require "fitting" a specified functional form to a representative set of points either in reapplication of the system or by appropriately modifying the initial hypotheses. However, graphical representation of selected cross-sections would seem also to be generally convenient.

In some trivial cases, superfluous unidentifiable parameters can be eliminated by simple substitution, using the indicated relationship. However in general, more elaborate procedures are required; and in some cases, the unidentifiability will probably never be satisfactorily resolved.

Calibration

Presumably, the most common form of unidentifiability is that associated with errors in observation. In measuring a table with a previously unused ruler, we can never be absolutely sure that the scale is not significantly either too short or too long, and if so by how much. Classically the scale is calibrated against a known and accepted standard. It clearly serves no purpose to check the measurement against other previously unused scales, assuming no other information. Or in general, when calibration models include unevaluated location parameters, no purpose is served in substituting other methods of observation, the associated calibration models for which also involve unevaluated location parameters.

Actual calibration against appropriate "standard object phenomena" is much to be desired, perhaps more so for the more complex phenomena and models. In application of the system for the purpose of calibration, we seek more complete specification of the calibration model prior to its use in conjunction with other nonstandard object models. The ideal standard object phenomenon might consist of a finite, more or less uniform lattice of event points distributed over the object variable space, at least encompassing the range of variables within which the object phenomena to be investigated are likely to be centered. Parameter values of the standard object phenomenon model are most appropriately established by definition or with reference to other known standards.

Thus, in the usual calibration situation, the calibration model parameters alone remain to be evaluated. Unidentifiability due to error is thereby eliminated, unless there is unidentifiability inherent in the error phenomenon model as specified. However, it is conceivable that not all standard object phenomenon parameter values will be established previously. In such a case, they may be evaluated along with the calibration model parameters, providing no unidentifiability is introduced by the interaction.

A requirement that actual calibration be employed in all cases may be unnecessarily restrictive. Given a surface of unidentifiability or the set of all maximum points, in general we desire that the appropriate number of parameter values necessary for complete identification be included in the model specifications. This does not mean that all parameters evaluated by specification must be calibration model parameters. They need only constitute that subset of the necessary size about which the most information, not otherwise employed, is available. Actual calibration is only one possible source of such information.

In the usual practical application of the system, where unidentifiability due to error is present, actual calibration is probably to be preferred over other means of effecting complete identification. In each specific instance, only the investigator is qualified to decide.

Considerations in Ranking Models

An investigator may choose initially to include an additional parameter in the model (a "fudge factor"), hopefully to counteract a potential weakness in his theory, or as a device in theoretical exploration. The introduction of additional parameters, above and beyond those which are truly justified by current theory, will not necessarily result in unidentifiability. Parameters which are superfluous in this sense may take on estimated values which are unique but effect no significant influence in the model. Another possibility is that the additional flexibility thereby introduced into the model will result in a higher

but misleading value for the maximum likelihood; a better fit is obtained to that variation in the observations which is attributable solely to the random component in sampling. Since this is always a dangerous possibility, the need for large samples and verification is again evident.

A deficiency in the number of specified unknown parameters may or may not result in unidentifiability. However, such a deficiency always constitutes a specification error and a poor fit of the model to the observational data must result, with a consequent low value for the maximum likelihood.

In general, we can expect a higher maximum value for the likelihood with a more complicated model; but as always, we do not further our goal of model development unless a state of control is also established. Perhaps the maximum likelihood values should also be considered in resolving questions of state of control.

Recall also that maximum likelihood is an acceptable criterion for selection of the best model from a number of models, providing that the same data is used in each case. Thus a direct maximum likelihood comparison of object phenomenon models, which have been developed using different methods of observation, is not necessarily valid. Additional information concerning a state of "mutual control" on the designated phenomena may be required. Also, an adjustment must be made for the difference in sample size.

Availability of the likelihood value makes its use convenient in ranking models; also the associated difficulties do not seem to be too serious. However, this method of ranking is probably not that which is ultimately best.* Consider that the ability to predict has been associated, in an earlier section, with the idea of understanding. Prediction is only temporal extrapolation. Thus we can regard the ability to extrapolate in the direction of other variables as a useful criterion of the achievement of understanding, and consequently a useful criterion in ranking models.

Every model, and in fact every functional model component, can be regarded as an implicit function of any variable in nature which is not already involved in the functional expression. Model components, by their specification, are presumed to be unchanging in the coordinate directions of those variables which are omitted. Thus, no variable is excluded from consideration in this sense.

*Note that "Goodness of Fit", in its specialized sense, is not considered in this discussion.

Clearly, the ability to predict is also evidence of a state of control; and further, the ability to extrapolate in the direction of other variables is evidence of a generalized state of control. Thus, given some quantitative measure of generalized state of control, such measure may also be appropriate if not superior to likelihood, in ranking models.

State of Control

Recall that the object phenomenon, as designated, actually constitutes a class of phenomena, namely that which is the logical intersection of classes established by the individual designating statements. Within the designated class intersection, there is generally room for some detectable variation (i.e., acceptable variation in "representative" sets of observations), random or otherwise. Any such variation which is excessive, according to some criterion, is by definition evidence of lack of control.

In any serious attempt at model development, an indicated significantly large lack of control can hardly be tolerated. The investigator will usually choose not to narrow the scope of his investigation. From the information available, he will probably attempt to adjust his initial specifications. Parameters will tend to vary in the degree to which they are reproducible. He may therefore specifically devote his attention to those model parameters the values of which are relatively out of control.

Criteria for control are best established with reference to the distribution of observations. However as stated, the acceptable residual variation is also influential in the parameter space. Consider a situation involving unidentifiability, wherein a number of independent investigations are conducted; and in each case, with the same object phenomenon, the same method of observation, and the same initially specified hypotheses. Then from each investigation, a surface of unidentifiability results.

Assuming only one member in the designated object class, or a "perfect state of control", for large sample size we would expect that all of the surfaces would effectively coincide. However, in practical model development, even assuming no sampling error, perfection is not likely to be achieved. A more likely result is depicted in Figure 4. Here, four surfaces E_1,------,E_4 are shown to be intersecting; although in general, there is no reason that they should.

In the parameter space, measures of state of control could refer to variation about some "central point". For example, the least squares point for distances normal to each surface, where such exists, might

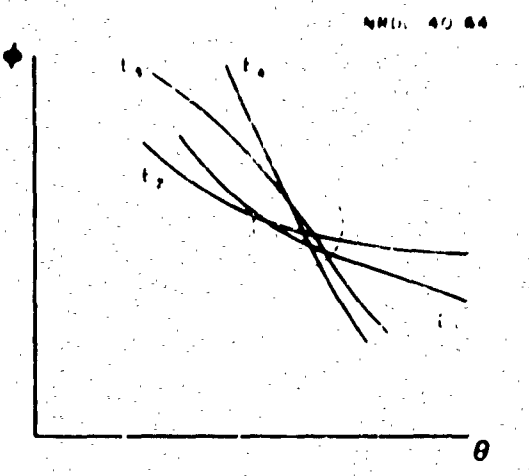


Fig. 4 Evidence of Specification Error or Lack of Control

be optimum. Any decreasing function of the minimum sum of squares could then be used to indicate degree of control.

The dotted circle of the illustration can be said to define a region of acceptable risk; and in the same sense, the least squares point may be considered to indicate the safest parameter estimates. This is a barefaced behavioristic point of view; but it seems to be reasonable under the circumstances.

Thus, in these ways, the development of mathematical models may be pursued.

CONCLUDING REMARKS

We have been primarily concerned with the elimination of undue restrictions which are imposed on the investigator in his acceptance of currently available procedures for data analysis; a corollary concern

has been to point out some important deficiencies of ordinary investigative procedures. To this end, an extension and synthesis of scientific methods has been attempted.

A subsidiary goal has been to contribute to the elimination of inefficiency in scientific investigation through contribution to methods of organization and presentation. Ultimately, automation must effect the greatest influence. In view of these goals, it is hoped that this paper may also serve as a procedural guide or handbook.

By the term "initial specifications", it is not intended to imply that no labor is involved in their formulation. In fact, it is likely that this will constitute the investigators most difficult problem, since it is here that his unique professional skills find their greatest application. However, it is here also that the statistician may prove most helpful. In particular, he may assist in or even direct construction of the calibration model. Also, he may assist in the actual calibration.

The development of mathematical models must of course proceed concurrently with the development of the associated scientific discipline. In the usual situation, initial hypotheses are formulated with reference to previously developed object models, and in the light of new information. Consequently, no overwhelming difficulty need attend construction of the initial hypotheses at any one stage of the development. The same conditions of course apply in formulation of the calibration model hypotheses.

REFERENCES

1. Braithwaite, R. B. (1953). Scientific Explanation, A Study of the Function of Theory, Probability and Law in Science. Cambridge, at the University Press.

2. Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

3. Deming, W. E. (1943). Statistical Adjustment of Data. John Wiley and Sons.

4. Fisher, R. A. (1950). Statistical Methods for Research Workers. Hafner Publishing Company.

5. Haavelmo, T. (1944). The Probability Approach in Econometrics. Econometrica, Vol. 12, July Supplement.

6. Kendall, M. G. (1951). Regression, Structure, and Functional Relationships. Part I. Biometrika 38:11-25.
   _____ (1952) Part II. Biometrika 39:96-108.

7. Koopmans, T. C. and Reiersøl, O. (1950). The Identification of Structural Characteristics. Annals of Mathematical Statistics 21:165-181, June.

8. Madansky, A. (1959). The Fitting of Straight Lines When Both Variables are Subject to Error. American Statistical Association Journal, March, pp. 173-206.

9. Mood, A. M. (1950). Introduction to the Theory of Statistics. McGraw-Hill.

10. Moore, R. H. and Zeigler, R. K. (1960). The Solution of the General Least Squares Problem with Special Reference to High Speed Computers. Los Alamos Scientific Laboratory Report LA-2367.

11. Munroe, M. E. (1951). Theory of Probability. McGraw-Hill.

12. Shewhart, W. A. (1939). Statistical Method From the Viewpoint of Quality Control. Department of Agriculture, The Graduate School.

13. Trotter, H. F. (1957). Gauss's Work (1802-1826) on the Theory of Least Squares, An English Translation. Atomic Energy Commission Technical Report 3049; Statistical Techniques Research Group, Princeton University.