DISCRIMINATION AMONG MECHANISTIC MODELS

G. E. P. Box and W. J. Hill

ABSTRACT

This paper is concerned with research, the object of which is to discover the mechanism for a particular phenomenon leading to a specific mathematical model. Such investigations are distinguished from those in which the object is merely to estimate the output \( y \) of a process over a range of values of the input \( \xi_1, \xi_2, \ldots, \xi_k \). Frequently, a number of possible mechanisms are suggested from theoretical considerations leading to a number of different mathematical models. To discriminate among these a sequential procedure is developed in which calculations made after each experiment determine the most discriminatory process conditions for use in the next experiment. The method is illustrated with examples.

1. USING MATHEMATICAL MODELS TO DESCRIBE PHYSICAL PHENOMENA

The experimenter is often concerned with the study of systems in which there is a dependent variable \( y \) related to independent variables \( \xi = (\xi_1, \xi_2, \ldots, \xi_k)' \). For example, he might be studying the yield \( y \) of a chemical reaction under various conditions of temperature \( \xi_1 \), pressure \( \xi_2 \), and catalyst concentration \( \xi_3 \). Any such phenomenon is theoretically capable of representation by a mathematical equation

\[
E(y) = f(\theta, \xi),
\]

(1.1)
which takes account of the mechanism of the process. In equation (1.1) 
\( \theta = (\theta_1, \theta_2, \ldots, \theta_p)' \) is the vector of \( p \) parameters which are usually physical constants of the system under study. We refer to this as the **true** mathematical model. In practice such a model accounting for the phenomenon under study in terms of its exact mechanism is seldom obtainable. Many examples occur, however, where the main characteristics of a mechanism can be elucidated and a close mechanistically based mathematical representation is possible.

It should be noted that the objective which we are presently considering is that of finding out "how a system works." The reason for this may be no more than scientific curiosity. However, if we know how the system works and can describe it by a mathematical model, then we can use this knowledge for practical aims such as predicting the behavior of the process under various experimental conditions and, in particular, in finding optimum operating conditions. This last fact leads to some confusion because if all we need to do is either to estimate the behavior of the process under various experimental conditions or to find optimum operating conditions, we do not necessarily need a mechanistic model. In some circumstances, an attempt to discover the mechanism merely to develop an operable system would be needlessly time consuming. For example, it may be known that the observed strength \( y \) of an extruded plastic sheet depends on the rate of extrusion \( \xi_1 \) and the rate of cooling immediately after extrusion \( \xi_2 \).
Nevertheless, a study to discover the nature of the true mechanism relating $y$ to $\xi_1$ and $\xi_2$ might be extremely difficult and laborious. It would certainly not be needed if all that was required was to determine how the process should be adjusted to give a plastic sheet of greater strength. This more limited objective could probably be obtained from a purely empirical approach, for example, response surface methods (see [1], [2], [3], [4], [5], [6], [7]). However, when there is a good possibility of obtaining a mechanistically based model with a reasonable experimental effort, this approach and not the purely empirical one ought to be followed. This is so even when the objective is only the practical one of estimating the response $y$ over a particular region of the experimental conditions.

The approach via a mechanistic model is particularly useful in the development of new processes. The chances of meaningful extrapolation are very much greater with such a model. Even here, however, because mechanisms are never perfectly known, we cannot hope for much more than having the model indicate a region which has not yet been investigated but is worthy of experimental study. The chance that such predictions will be borne out may be reasonably good with a mechanistic model. Although empirical models such as interpolation polynomials may be perfectly adequate to represent what is happening in the immediate region of the experiments, they will provide little basis for extrapolation.
2. IDENTIFYING THE CORRECT MODEL

The process of building a mechanistic model can be thought of as involving three stages:

(i) identification

(ii) fitting

(iii) diagnostically checking the adequacy of fit.

The statistical literature has tended to concentrate on stages (ii) and (iii). Less attention has been given to (i). To illustrate what identification involves suppose we consider the catalytic reaction

$$2 \, A \rightarrow B + C.$$ 

If we assume the dual site mechanism, outlined in [10]

(a) $A + s \rightarrow A_s$ adsorption,

(b) $2 \, A_s \rightarrow B_s + C_s$ surface reactions,

(c) $B_s \rightarrow B + s$ desorption,

(d) $C_s \rightarrow C + s$ desorption,

where $s$ is an active catalytic site, we see that there are four different possible rate controlling steps. From each of these possible rate controlling steps can be derived a reaction rate model. These candidate models are

(a) 

$$r = \frac{k_1 \Phi \left( \xi_A - \sqrt{\frac{\xi_B \xi_C}{K}} \right)}{(1 + \sqrt{\frac{\theta_A}{K}} \xi_A \xi_C + \theta_B \xi_B + \theta_C \xi_C)} ,$$
(b) \[ r = \frac{\gamma \theta_A^2 \left( \xi_A^2 - \frac{\xi_B \xi_C}{K} \right)}{(1 + \theta_A \xi_A + \theta_B \xi_B + \theta_C \xi_C)^2} , \]

(c) \[ r = \frac{k_2 \theta_B \xi_B \Phi \left( \frac{\xi_A^2}{\xi_C} - \frac{\xi_B}{K} \right)}{(1 + \theta_A \xi_A + \theta_B \xi_B + \theta_C \xi_C)} , \quad (2.2) \]

(d) \[ r = \frac{k_4 \theta_C \xi_C \Phi \left( \frac{\xi_A^2}{\xi_B} - \frac{\xi_C}{K} \right)}{(1 + \theta_A \xi_A + \theta_B \xi_B + \theta_C \xi_C)} , \]

where \( r \) is the rate of reaction, \( \theta, \Phi, k_1, k_2, k_4, K, \) and \( \gamma \) are constants of the system, and \( \xi \) are partial pressures. If we took into account other mechanisms besides (2.1), we could postulate many more candidates which could conceivably describe the system.

It might be expected that all one needs to distinguish among rival models is to run a number of careful experiments at various experimental conditions. It might then be hoped that one candidate model would fit the data very well whereas the others would show a marked lack of fit. Unfortunately, it is easy to collect data that are well fitted by a large number of different models. Different research groups commonly claim widely varying mechanisms for the same chemical system. Each can produce data which prove that their mechanism is the right one. One such reaction
is the Water Gas Shift reaction [9]

\[ \text{CO} + \text{H}_2\text{O} \xrightarrow{\text{catalyst}} \text{CO}_2 + \text{H}_2, \]

where there are at least ten or more different candidates.

The above situation can occur because the model that one group favors is never really "put in jeopardy" by the experiments they conduct. As an illustrative example of this consider the two candidate mechanisms for a chemical reaction involving components A, B, and C

(a) \[ A \rightarrow B \rightarrow C, \]
(b) \[ A \rightarrow B \leftrightarrow C. \]

In Figure 1 the concentration of B is plotted against time \( \xi \) for both reactions. Obviously, in order to discriminate between these two mechanisms it is not enough to observe B only over small values of \( \xi \) where the curves can almost be identical in shape. We need to observe B at higher values of \( \xi \).

3. SEQUENTIAL DESIGNS FOR DISCRIMINATION

In the problem considered here, it is supposed that an experimenter is investigating a physical system in which the underlying theory suggests at the outset that any one of \( m (>1) \) models might describe the system. It is also supposed that after \( n-1 \) preliminary experimental runs, the results are not conclusive as to which model is best. In order to discriminate among these \( m \) rival models the experimenter may wish to perform further
Figure 1. Two Rival Mechanisms for a Chemical Reaction Involving Components A, B, and C.
experimental runs. If experiments are conducted in sequence, the experimenter can consider each result before he runs the next experiment. In particular, computations prior to each new experiment can indicate where the next experiment ought to be conducted to provide maximum discrimination among m rival models. A sequential approach for \( m = 2 \) was that of Hunter and Reiner [8]. Here experiments were performed at that set of conditions which essentially gave the greatest estimated difference in response. This method, however, did not take into account the magnitude of the error of the estimated difference.

In considering the general problem (\( m \geq 2 \)), we first decide upon some relevant discrimination criterion and then choose experimental conditions which maximize this measure at each stage. In the measure of discrimination one must not only take account of the difference in response given by the models to be discriminated among but also take into consideration the variance of the estimated response. Thus, as shown in Figure 2, if the two lines represent the estimated relationship \( y \) on \( x \) and the crossed areas indicate the limits of error (say 95% confidence region), then it will by no means be true that the best point to discriminate between the two models will be that where there is maximum divergence between the two lines. This is so, since clearly it is the divergence relative to the limits of error to be considered rather than divergence itself at any chosen point.
Figure 2. Two Rival Prediction Equations and Their Corresponding Error Regions.
4. A DISCRIMINATION CRITERION

In studying the amount of information supplied by a communication system, Shannon [12] used the concept of entropy where entropy is defined as

\[ S = - \sum_{i=1}^{m} \Pi_i \ln \Pi_i, \] (4.1)

where \( \Pi_i \) is the probability associated with a symbol \( i \). The least possible information corresponds to maximum entropy, that is, when \( \Pi_1 = \Pi_2 = \ldots = \Pi_m = \frac{1}{m} \). An illustration of this situation for \( m = 4 \) is shown in Figure 3(a). It is more desirable to have a situation where \( \Pi_j > \Pi_i \) where \( i \neq j \). Here, the entropy is small and the amount of information is large. See, for example, Figure 3(b). In a communication system, the first situation (Figure 3(a)) usually represents the input state and the second situation (\( \Pi_j > \Pi_i \)) usually represents the output state. Therefore, in order to obtain the maximum information from the system it is desirable to have a maximum change in entropy between input and output. This occurs, of course, when for any given input the output is such that \( \Pi_j = 1 \) and \( \Pi_i = 0 \) for \( i \neq j \). (In a communication system, this means that the probability of the symbol \( j \) being correctly received is unity.) This concept is now applied to the discrimination among \( m \) models where it is desired to go from a noninformative situation like that in Figure 3(a) to a more informative situation like that in Figure 3(b).
Figure 3. Illustration of Entropy for (a) the Least Informative Case, and (b) a More Informative Case.
At input (after n-1 observations) the prior probabilities associated with the m models are \( \Pi_{in-1}, \Pi_{2n-1}, \ldots, \Pi_{mn-1} \). The posterior probabilities found by taking the n-th observation \( y_n \) are \( \Pi_{in}, \Pi_{2n}, \ldots, \Pi_{mn} \). The expected change in entropy between input and output is expressed as

\[
R = \text{entropy at input} - \text{expected entropy at output} = - \sum_{i=1}^{m} \Pi_{in-1} \ln \Pi_{in-1} - (-1) \int \left( \sum_{i=1}^{m} \Pi_{in} \ln \Pi_{in} \right) q(y_n) \, dy_n,
\]

(4.2)

where

\[q(y_n) = \sum_{i=1}^{m} \Pi_{in-1} p_i,
\]

(4.3)

where \( p_i \) is the probability density function of the n-th observation under model \( i \). Since the posterior probability associated with the \( i \)-th model is

\[\Pi_{in} = \frac{\Pi_{in-1} p_i}{q(y_n)},
\]

(4.4)

substitution of (4.4) into (4.2) gives

\[
R = \sum_{i=1}^{m} \Pi_{in-1} \int p_i \ln \frac{p_i}{q(y_n)} \, dy_n.
\]

(4.5)

By substituting into (4.5) the following inequality (Corollary 3.1 and Example 3.2 of Kullback [11])

\[
\sum_{j=1}^{m} \Pi_{jn-1} p_i \ln \frac{p_i}{p_j} \geq \sum_{j=1}^{m} \Pi_{jn-1} p_i \ln \frac{p_i}{q(y_n)},
\]

(4.6)
we have

\[ R \leq D, \quad (4.7) \]

where

\[ D = \sum_{i=1}^{m} \sum_{j=i+1}^{m} \Pi_{in-1} \Pi_{jn-1} \left( \int p_i \ln \frac{p_i}{p_j} \, dy_n + \int p_j \ln \frac{p_j}{p_i} \, dy_n \right). \quad (4.8) \]

The expression in the parentheses of equation (4.8) is Kullback's total measure of information (or divergence) when discriminating between two hypotheses \( H_i \) and \( H_j \). The function \( D \) represents the maximum change in entropy expected from \( y_n \) and is a measure of discrimination among \( m \) rival hypotheses or models.

In order to use \( D \) in a discrimination criterion, the following distribution theory is developed for the \( n \)-th observation. Under model \( i \) \((i=1, 2, \ldots, m)\) it is assumed that the observation \( y_n \) is normally distributed with expected value \( \eta = \mathbb{E}(y_n) \) and known variance \( \sigma^2 \). The probability density function of \( y_n \) given \( \eta \) and \( \sigma \) is

\[ p_i(y_n | \eta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(y_n - \eta)^2 \right\}. \quad (4.9) \]

If \( \eta \) is linear in the parameters \( \theta_i \) \((i=1, 2, \ldots, m)\) or can be approximately expressed as a linear function of \( \bar{\theta}_i \) in the region of the parameter estimates \( \bar{\theta}_i \), then from a Bayesian development \( \eta \) is normally distributed about \( \bar{y}_n^{(i)} \).
with standard deviation $\sigma_i$. The predicted value of $y_n$ under model $i$ using the first $n-1$ observations is $\tilde{y}^{(i)}_n$ and the variance of $\tilde{y}^{(i)}_n$ is $\sigma_i^2$. (See Appendix for this linear theory development and for the calculation of $\sigma_i^2$.) The probability density function of $\eta$ given $\sigma$ is then written as

$$p_1(\eta \mid \sigma) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left\{ -\frac{1}{2\sigma_i^2} (\eta - \tilde{y}^{(i)}_n)^2 \right\} . \quad (4.10)$$

Therefore, the probability density function of $y_n$ under model $i$ given $\sigma$ and the first $n-1$ observations is

$$p_i = p_1(y_n \mid \sigma) = \int p_1(y_n \mid \eta, \sigma) p_1(\eta \mid \sigma) \, d\eta . \quad (4.11)$$

Substituting equations $(4.9)$ and $(4.10)$ into $(4.11)$ and integrating, we have

$$p_1 = \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_i^2)}} \exp \left\{ -\frac{1}{2(\sigma^2 + \sigma_i^2)} (y_n - \tilde{y}^{(i)}_n)^2 \right\} . \quad (4.12)$$

When the results of $(4.12)$ are substituted in the discrimination function $D$ of equation $(4.8)$ we have

$$D = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=i+1}^{m} \prod_{n-1}^{i-1} \prod_{j-1}^{n-1} \left\{ \frac{(\sigma_i^2 - \sigma_j^2)^2}{(\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2)} + \frac{(\tilde{y}^{(i)}_n - \tilde{y}^{(j)}_n)^2}{\sigma_i^2 + \sigma_j^2} \right\} . \quad (4.13)$$

To select the $n$-th experiment that attains maximum expected discrimination among the $m$ rival models, one chooses those operating conditions that maximize $D$. After $y_n$ is observed the current standing of each model can
be checked by calculating its corresponding posterior probability from equation (4.4). The procedure is repeated until the posterior probabilities indicate that one model is clearly superior to the others.

5. EXAMPLES

It is usually found that it is only possible to conduct experiments in any given situation over a particular region of $\mathcal{F}$. This region will be called the operability region $O$. Thus, for example, with a particular apparatus it might be impossible to conduct experiments outside temperature and pressure ranges. It will be noted that the operability region $O$ is usually very extensive and it would normally be quite impossible to conduct experiments adequately covering all of $O$. Cases may occur where during the course of an investigation we may be led to redesign our apparatus to extend the operability region in a particular direction of interest.

The following three examples illustrate the application of the discrimination criterion. Although, polynomial models are used in the first two examples, this is not meant to suggest that this discrimination criterion is proposed for the empirical model situation. These examples are essentially intended to illustrate the characteristics of the criterion. More in keeping with the purpose of this study, that is, the elucidation of the underlying mechanism of a system, the third example includes mechanistically based models that might be considered as rivals when an experimenter is studying the reaction of the form $A \rightarrow B$. 
Example 5.1

As a simple illustration of the discrimination criterion, consider the case of the \( m = 2 \) models

\[
\text{Model 1. } \quad E(y) = \theta_{11} \xi, \quad (5.1)
\]

\[
\text{Model 2. } \quad E(y) = \theta_{21} + \theta_{22} \xi,
\]

where the variance is assumed known. See Figure 2 for a graphical representation of the prediction equations for these models and their corresponding error regions. Model 1 implies that the expected value of the response is proportional to \( \xi \) and hence is represented by a straight line passing through the origin. Model 2 implies that the expected value of the response follows a straight line not necessarily passing through the origin. For this particular example it is easy to derive theoretically the course which the discrimination procedure will follow. That is, when \( -\infty < \xi < \infty \), D is maximized if the first discriminatory observation is taken at

\[
\xi_n = - \sum_{u=1}^{n-1} \xi_u. \quad (5.2)
\]

It follows from equation (5.2) that all further maximum discrimination points (taken one at a time) occur at \( \xi = 0 \). Therefore, as one might expect, in order to discriminate between the two first order models (5.1), future settings of \( \xi \) should be chosen at the origin after an initial discrimination point has been chosen according to expression (5.2).
Example 5.2

To illustrate the case $m > 2$, consider the discrimination among the $m = 4$ polynomial models

Model 1. \[ E_1 (y) = \theta_{11} \xi, \]

Model 2. \[ E_2 (y) = \theta_{21} + \theta_{22} \xi, \]

Model 3. \[ E_3 (y) = \theta_{31} + \theta_{32} \xi + \theta_{33} \xi^2, \] (5.3)

Model 4. \[ E_4 (y) = \theta_{41} \xi + \theta_{42} \xi^2, \]

representing respectively a straight line passing through the origin, a straight line not necessarily passing through the origin, a quadratic not constrained to pass through the origin, and a quadratic passing through the origin. The standard deviation $\sigma = 1$ is assumed known. Data are generated from model 3 where $\theta_{31} = \theta_{32} = \theta_{33} = 1$. (In an actual experimental situation, of course, the true model and the corresponding parameter values would not be known.) The operability region for $\xi$ is $0 \leq \xi \leq 4$ and preliminary data are generated at $\xi = 0, 1, 2, 3, 4$. When the prior probabilities are taken to be $\Pi_{10} = \Pi_{20} = \Pi_{30} = \Pi_{40} = .25$, then the posterior probabilities after five data points are $\Pi_{45} = .0024$, $\Pi_{25} = .0058$, $\Pi_{35} = .6583$, and $\Pi_{45} = .3335$. Models 3 and 4 appear to be the important rivals from these initial data. For $n = 6$, the grid ($\xi = 0(.2)4$) is searched to find where $D$ is maximized. Maximum discrimination occurs at $\xi = 0$ and when an observation is generated at this point the posterior probabilities
are $\Pi_{16} = 0.0009$, $\Pi_{26} = 0.0012$, $\Pi_{36} = 0.8777$, and $\Pi_{46} = 0.1202$. The procedure is repeated for $n = 7$, 8, and 9. The discrimination results are summarized in Figure 4. Because the important rivals (models 3 and 4) differ only by a constant, this may partially explain why maximum discrimination occurred at $\xi = 0$ for each run. Since after nine observations the posterior probabilities indicate that model 3 is clearly superior to the other three models, we discriminate in favor of model 3 (true model).

This example is one in which certain models are specific cases of others. For instance, by setting $\theta_{21} = 0$ in model 2, $\theta_{31} = \theta_{43} = 0$ in model 3, and $\theta_{42} = 0$ in model 4, one obtains model 1 in each case. In these circumstances it might be thought that if model 1 were the true model and since it follows that models 2, 3, and 4 would also be correct, it might be impossible to discriminate among them. In fact, the criterion does discriminate effectively in this case, essentially because the average variances of the estimates of the response, obtained from models with fewer parameters are smaller and thus the posterior probabilities become larger.

This reducible model situation is illustrated when model 1 is taken to be true with $\theta_1 = 4$. The posterior probabilities are shown in Figure 5 at each stage up to $n = 15$ observations. Although the convergence is slow, as would be expected, and although more observations are needed to complete the analysis, there is a marked preference for the simplest model, model 1, even though all the models fit the data. Notice that model 3 which has the most parameters has the lowest probability
Initially
\[ \pi_{10} = .25 \quad \pi_{20} = .25 \quad \pi_{30} = .25 \quad \pi_{40} = .25 \]

Model
\[ 1 \quad 2 \quad 3 \quad 4 \]

After n=5 Preliminary Observations
\[ \pi_{15} = .00 \quad \pi_{25} = .01 \quad \pi_{35} = .66 \]
\[ \pi_{45} = .33 \]

\[ 1 \quad 2 \quad 3 \quad 4 \]

n = 6
\[ \xi = 0 \]
\[ \pi_{16} = .00 \quad \pi_{26} = .00 \quad \pi_{36} = .88 \]
\[ \pi_{46} = .12 \]

\[ 1 \quad 2 \quad 3 \quad 4 \]

n = 7
\[ \xi = 0 \]
\[ \pi_{17} = .00 \quad \pi_{27} = .00 \quad \pi_{37} = .75 \]
\[ \pi_{47} = .25 \]

\[ 1 \quad 2 \quad 3 \quad 4 \]

n = 8
\[ \xi = 0 \]
\[ \pi_{18} = .00 \quad \pi_{28} = .00 \quad \pi_{38} = .90 \]
\[ \pi_{48} = .10 \]

\[ 1 \quad 2 \quad 3 \quad 4 \]

n = 9
\[ \xi = 0 \]
\[ \pi_{19} = .00 \quad \pi_{29} = .00 \quad \pi_{39} = .97 \]
\[ \pi_{49} = .03 \]

\[ 1 \quad 2 \quad 3 \quad 4 \]

Figure 4. Posterior Probabilities for Four Polynomial Models (Model 3 True).
Initially

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
0.25 & 0.25 & 0.25 & 0.25 \\
\end{array}
\]

Model

After \( n = 5 \)

Preliminary Observations

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
0.32 & 0.24 & 0.19 & 0.25 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 10 & \xi = 0.0 & 0.45 & 0.18 & 0.10 & 0.27 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 11 & \xi = 1.8 & 0.47 & 0.19 & 0.09 & 0.25 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 6 & \xi = 0.0 & 0.32 & 0.25 & 0.18 & 0.25 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 12 & \xi = 0.0 & 0.48 & 0.18 & 0.08 & 0.26 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 7 & \xi = 0.0 & 0.39 & 0.19 & 0.13 & 0.29 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 13 & \xi = 1.6 & 0.50 & 0.18 & 0.08 & 0.24 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 8 & \xi = 1.6 & 0.38 & 0.19 & 0.13 & 0.30 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 14 & \xi = 0.0 & 0.51 & 0.17 & 0.07 & 0.25 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 9 & \xi = 4.0 & 0.43 & 0.20 & 0.11 & 0.26 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

\[
\begin{array}{cccc}
n = 15 & \xi = 0.0 & 0.52 & 0.16 & 0.07 & 0.25 \\
1 & 2 & 3 & 4 \\
\end{array}
\]

FIGURE 5. Posterior Probabilities for Four Polynomial Models (Model 1 True).
associated with it after 15 observations.

Example 5, 3

Suppose an experimenter is studying a chemical reaction of the type $A \rightarrow B$. Depending on whether the reaction is of first, second, third, or fourth order, the following $m = 4$ rival models can be considered

Model 1. $E(y) = \exp(-\theta_{11} \xi_1 \exp(-\theta_{12}/\xi_2))$,

Model 2. $E(y) = 1/(1 + \theta_{21} \xi_1 \exp(-\theta_{22}/\xi_2))$,

Model 3. $E(y) = 1/(1 + 2 \theta_{31} \xi_1 \exp(-\theta_{32}/\xi_2))^{1/2}$,

Model 4. $E(y) = 1/(1 + 3 \theta_{41} \xi_1 \exp(-\theta_{42}/\xi_2))^{1/3}$,

where $y$ is the concentration of $A$, $\xi_1$ is the reaction time, $\xi_2$ is the temperature, and $\theta_{ij}$ and $\theta_{12}$ are the parameters of model $i$ ($i=1, 2, 3, 4$). In a constructed study, model 2 is chosen as the true model with $\theta_{21} = 400$ and $\theta_{22} = 5000$. The standard deviation $\sigma = .05$ is assumed known. The operability region in the $(\xi_1, \xi_2)$ space is $0 \leq \xi_1 \leq 150$ and $450 \leq \xi_2 \leq 600$. It is desired to select design points in units of 25 for both $\xi_1$ and $\xi_2$. A preliminary $2^2$ – factorial design is chosen at levels $\xi_1 = 25$ and $125$, and $\xi_2 = 475$ and $575$. Data are generated from model 2 at these points and the initial priors are chosen to be $\Pi_{10} = \Pi_{20} = \Pi_{30} = \Pi_{40} = .25$. After these initial four runs, model 3 is slightly favored over model 2 (true model), whereas models 1 and 4 appear to have little importance. This is indicated by the posteriors $\Pi_{14} = .0069$, $\Pi_{24} = .4290$, $\Pi_{34} = .5008$, and $\Pi_{44} = .0633$ which are shown in Table 1. For $n = 5$ the discrimination function $D$ is
maximized on the grid (ξ₁ = 0 (25) 150, ξ₂ = 450 (25) 600) when (ξ₁, ξ₂) = (125, 600). When an observation is generated at these conditions and when the posterior probabilities are calculated, model 2 is slightly favored. This is seen in Table 1 for n = 5. If the procedure is repeated until n = 8, model 2 is chosen as the best model on the basis of the posterior probabilities. The discrimination results are shown in Figure 6 and Table 1.

Suppose that due to either wild observations or poor prior knowledge the wrong model is initially favored. Then, because of the weighting in D due to the presence of the prior probabilities, is there a possibility that the procedure may never recover from this incorrect weighting and continue to favor the wrong models? Consideration was given to the above question when this example was done again using an initial prior probability for model 3, seven times as great as those for models 1, 2, and 4. In reality, of course, model 2 was the true model and model 3 was the most important rival. Even with this handicap the procedure still rapidly discriminated and unmistakably revealed the true model after five additional observations. The results are shown in Table 2.

6. CONCLUSIONS

The availability of fast electronic computers ensures that calculations of the type envisaged here can be made in a reasonably short space of time. It is therefore possible to consider an important new development in the scientific method. The complexity of models which result
<table>
<thead>
<tr>
<th>n</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>$\gamma$</th>
<th>$\Pi_1$</th>
<th>$\Pi_2$</th>
<th>$\Pi_3$</th>
<th>$\Pi_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>575</td>
<td>0.3961</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>475</td>
<td>0.7232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td>475</td>
<td>0.4215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>125</td>
<td>575</td>
<td>0.1297</td>
<td>0.0069</td>
<td>0.4290</td>
<td>0.5008</td>
<td>0.0633</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>600</td>
<td>0.0984</td>
<td>0.0019</td>
<td>0.5602</td>
<td>0.4291</td>
<td>0.0088</td>
</tr>
<tr>
<td>6</td>
<td>125</td>
<td>600</td>
<td>0.0556</td>
<td>0.0018</td>
<td>0.8639</td>
<td>0.1339</td>
<td>0.0004</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>450</td>
<td>0.7969</td>
<td>0.0021</td>
<td>0.9736</td>
<td>0.0243</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>600</td>
<td>0.0325</td>
<td>0.0032</td>
<td>0.9956</td>
<td>0.0012</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**TABLE 1.** Results when Discriminating Among $m = 4$ Rate Models where $\Pi_{10} = \Pi_{20} = \Pi_{30} = \Pi_{40} = .25$ (Model 2 True).
FIGURE 6. Contours of Unreacted A:
1 - 4 Initial Design Points
5 - 8 Sequentially Chosen Points Using D.
TABLE 2. Results when Discriminating Among \( m = 4 \)
Rate Models where \( \Pi_{10} = \Pi_{20} = \Pi_{40} = .1 \) and
\( \Pi_{30} = .7 \) (Model 2 True).

<table>
<thead>
<tr>
<th>n</th>
<th>( \xi_1 )</th>
<th>( \xi_2 )</th>
<th>Y</th>
<th>( \Pi_1 )</th>
<th>( \Pi_2 )</th>
<th>( \Pi_3 )</th>
<th>( \Pi_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>575</td>
<td>0.3961</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>475</td>
<td>0.7232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td>475</td>
<td>0.4215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>125</td>
<td>575</td>
<td>0.1297</td>
<td>0.0018</td>
<td>0.1071</td>
<td>0.8752</td>
<td>0.0159</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>600</td>
<td>0.0984</td>
<td>0.006</td>
<td>0.1567</td>
<td>0.8402</td>
<td>0.0025</td>
</tr>
<tr>
<td>6</td>
<td>125</td>
<td>600</td>
<td>0.0556</td>
<td>0.011</td>
<td>0.4790</td>
<td>0.5197</td>
<td>0.0002</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>450</td>
<td>0.7969</td>
<td>0.0020</td>
<td>0.8496</td>
<td>1.484</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>600</td>
<td>0.0600</td>
<td>0.0035</td>
<td>0.9884</td>
<td>0.0081</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
<td>600</td>
<td>0.2140</td>
<td>0.0002</td>
<td>0.9900</td>
<td>0.0098</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
from other than rather simple mechanisms may make it almost impossible for the experimenter to judge what constitutes a worthwhile experiment at a given stage of an investigation. Placing the computer in the iterative circuit shown in Figure 7 enormously helps the experimenter to increase the discriminatory power of his experiments. The use of techniques like the above will increase the efficiency of experimentation and should reduce the frequency with which contradictory conclusions are reached in any given situation.
FIGURE 7. Iterative Experimental Procedure.
\[
L(\theta_1 | \sigma, y) = \frac{1}{(\sqrt{2\pi} \sigma)^{n-1}} \exp\left\{ -\frac{1}{2\sigma^2} (T_1 \theta_1 + (\tilde{\theta}_1 - \tilde{\theta}_1)' X_1' X_1 (\theta_1 - \tilde{\theta}_1) \right\}.
\] (7.7)

If a locally uniform prior distribution is assumed for \( \theta_1 \) then from Bayes theorem the posterior probability density function of \( \theta_1 \) after \( n-1 \) observations is

\[
p(\theta_1 | \sigma, y) = \frac{L(\theta_1 | \sigma, y)}{\int L(\theta_1 | \sigma, y) \, d\theta_1}.
\] (7.8)

Upon simplification (7.8) reduces to

\[
p(\theta_1 | \sigma, y) = \frac{1}{(\sqrt{2\pi} \sigma)^{p}} \exp\left\{ -\frac{1}{2\sigma^2} (\theta_1 - \tilde{\theta}_1)' X_1' X_1 (\theta_1 - \tilde{\theta}_1) \right\}.
\] (7.9)

Therefore, \( (\theta_1 - \tilde{\theta}_1) \) is normally distributed about zero with variance-covariance matrix \( (X_1' X_1)^{-1} \sigma^2 \).

Using the estimates of \( \theta_1 \) based on the first \( n-1 \) observations, we have the following linearized expression for the expected value of \( y_n \) under model 1.
REFERENCES


