Technical Report #111
April, 1967
EXPERIMENTAL STRATEGY

by
G. E. P. Box
University of Wisconsin

(Distribution of this document is unlimited.)

Paper to be read at International Biometrics Conference, Sydney, Australia, September, 1967.
This research was partially supported by the United States Navy through the Office of Naval Research, under Contract NONr-12 02 (17), Project NR042-222.
EXPERIMENTAL STRATEGY

by

G. E. P. Box
University of Wisconsin

At first sight the planning of experiments seems to be a highly arbitrary and uncertain science. For example, suppose that we collected ten groups of experimenters competent in a particular field of science or technology, locked them in ten separate rooms, presented each group with the same problem and asked them to submit their proposal for a suitable experimental plan. While on the one hand it is clear that the nature of the final conclusions drawn after the trials had been run would depend far more heavily on the particular plan submitted than on the data actually collected, on the other hand it is certain that no two groups of experimenters would present the same plan.

1. INDETERMINANCY OF EXPERIMENTATION

Some of the indeterminancies are indicated by the following questions:

1. Which variables or factors should be studied?
2. Which levels of a given factor should be considered?

If the factor concerned is qualitative like "variety of seed" this may involve choice from a very large population of possibilities; if the factor is quantitative this involves a choice from an infinite number of possibilities and in particular involves the choice
of "scales" in which the variables are to be considered.

3. When the variables are quantitative we have the question closely related to (2) of which metrics and which transformations of the variables should be used in the definition of the "factor levels." For example, suppose the amounts $x_A$ and $x_B$ of two nitrogenous fertilizers A and B were under study; should $x_A$ and $x_B$ themselves be regarded as the factors or would it be better to consider, for example, the total amount $z_1 = x_A + x_B$ of nitrogenous fertilizer as one factor and the ratio $z_2 = x_A / x_B$ as the other.

An orthogonal design constructed on the first basis (Figure 1 (a)) corresponds to a highly non-orthogonal arrangement on the second basis (Figure 1 (b)). Everyone knows that you should not use designs which look like that in Figure 1 (b)!

4. Finally, and again related to (2) and (3), we have the question how complex a model (and hence how elaborate an experimental arrangement) is necessary in a particular situation.

2. ITERATIVE NATURE OF THE EXPERIMENTAL LEARNING PROCESS

The situation appears rather more hopeful as soon as one remembers that an isolated group of trials of the kind so far discussed is in fact inevitably part of an iterative sequence. Thus, if we think in terms of the
of "scales" in which the variables are to be considered.

3. When the variables are quantitative we have the question closely related to (2) of which metrics and which transformations of the variables should be used in the definition of the "factor levels." For example, suppose the amounts $x_A$ and $x_B$ of two nitrogenous fertilizers A and B were under study; should $x_A$ and $x_B$ themselves be regarded as the factors or would it be better to consider, for example, the total amount $z_1 = x_A + x_B$ of nitrogenous fertilizer as one factor and the ratio $z_2 = x_A / x_B$ as the other.

An orthogonal design constructed on the first basis (Figure 1 (a)) corresponds to a highly non-orthogonal arrangement on the second basis (Figure 1 (b)). Everyone knows that you should not use designs which look like that in Figure 1 (b):

4. Finally, and again related to (2) and (3), we have the question how complex a model (and hence how elaborate an experimental arrangement) is necessary in a particular situation.

2. **Iterative Nature of the Experimental Learning Process**

The situation appears rather more hopeful as soon as one remembers that an isolated group of trials of the kind so far discussed is in fact inevitably part of an iterative sequence. Thus, if we think in terms of the
EXPERIMENTAL STRATEGY

by

G. E. P. Box
University of Wisconsin

At first sight the planning of experiments seems to be a highly arbitrary and uncertain science. For example, suppose that we collected ten groups of experimenters competent in a particular field of science or technology, locked them in ten separate rooms, presented each group with the same problem and asked them to submit their proposal for a suitable experimental plan. While on the one hand it is clear that the nature of the final conclusions drawn after the trials had been run would depend far more heavily on the particular plan submitted than on the data actually collected, on the other hand it is certain that no two groups of experimenters would present the same plan.

1. INDETERMINANCY OF EXPERIMENTATION

Some of the indeterminancies are indicated by the following questions:

1. Which variables or factors should be studied?
2. Which levels of a given factor should be considered?

If the factor concerned is qualitative like "variety of seed" this may involve choice from a very large population of possibilities; if the factor is quantitative this involves a choice from an infinite number of possibilities and in particular involves the choice
Figure 1. Two bases for choosing "factors" in a $3^2$ design
overall gathering of knowledge rather than in terms of the individual groups of trials, we see that we cannot hope to obtain a basis for the planning of experiments which ensures that every investigator will, in exactly the same circumstances, do exactly the same thing. What we can do is to attempt to organize research so that whatever the starting point, the whole process will, over a period of time, tend to lead to the right conclusions. It is with convergence rather than with uniqueness that we ought to be concerned.

As I have suggested elsewhere [1] [2] the iteration consists of the successive and repeated use of the sequence

CONJECTURE - DESIGN - EXPERIMENT - ANALYSIS

as illustrated in Figure 2(a) and is a process in which the creative mind of the experimenter is the essential element Figure 2(b).

The truly iterative nature of some investigations may sometimes be obscured by the length of time taken by each iterative cycle. In this case it will be possible to see the wider picture of iteration only by stepping back and seeing what occurs over months or years. Iteration may skip from one investigator to another, from one country to another and its phases may be long. Even in this situation, however, it is important to bear in mind that it is the overall acquisition of knowledge which we want to make an efficient process.
overall gathering of knowledge rather than in terms of the individual groups of trials, we see that we cannot hope to obtain a basis for the planning of experiments which ensures that every investigator will, in exactly the same circumstances, do exactly the same thing. What we can do is to attempt to organize research so that whatever the starting point, the whole process will, over a period of time, tend to lead to the right conclusions. It is with convergence rather than with uniqueness that we ought to be concerned.

As I have suggested elsewhere [1] [2] the iteration consists of the successive and repeated use of the sequence

\[\text{CONJECTURE} \rightarrow \text{DESIGN} \rightarrow \text{EXPERIMENT} \rightarrow \text{ANALYSIS}\]

as illustrated in Figure 2(a) and is a process in which the creative mind of the experimenter is the essential element Figure 2(b).

The truly iterative nature of some investigations may sometimes be obscured by the length of time taken by each iterative cycle. In this case it will be possible to see the wider picture of iteration only by stepping back and seeing what occurs over months or years. Iteration may skip from one investigator to another, from one country to another and its phases may be long. Even in this situation, however, it is important to bear in mind that it is the overall acquisition of knowledge which we want to make an efficient process.
Figure 1. Two bases for choosing "factors" in a $3^2$ design
Figure 2(a). Iterative Experimentation

Figure 2(b). The iterative experimental process in relation to the experimenter
3. SOME CLASSES OF PROBLEMS

Within this iterative context let us consider some particular problems of investigation. Many investigations are undertaken with the purpose of elucidating some aspect of a relation

\[ y_u = f(x_u | \theta) + e_u \]

where \( y_u \) is the observed response in the \( u \)th experimental run when \( k \) variables \( \xi \) are held at levels \( x_u \), \( \theta \) is a set of \( p \) parameters and \( e_u \) is the error. The state of ignorance with which we start and the state of comparative knowledge which we wish to brought to, will determine our approach. It is realized, of course, that no real problem ever quite fits any prearranged category. With this proviso it is perhaps helpful to distinguish the following types of problems and for reference purposes I have given each a tentative name.

<table>
<thead>
<tr>
<th>SUPPOSED U KNOWN</th>
<th>OBJECTIVE</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>Determine the subset ( \xi ) of important variables from a given larger set ( \Xi ) of potentially important variables</td>
<td>Screening variables</td>
</tr>
<tr>
<td>( \xi )</td>
<td>Determine the empirical &quot;effects&quot; of the known input variables ( \xi ).</td>
<td>Empirical model building</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Determine a local interpolation approximation ( f(\xi</td>
<td>\theta) ) to ( f(\xi</td>
</tr>
<tr>
<td>( f )</td>
<td>Determine ( f )</td>
<td>Mechanistic model building</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Determine ( \theta )</td>
<td>Mechanistic model fitting</td>
</tr>
</tbody>
</table>
3. SOME CLASSES OF PROBLEMS

Within this iterative context let us consider some particular problems of investigation. Many investigations are undertaken with the purpose of elucidating some aspect of a relation

\[ y_u = f(x_u | \theta) + e_u \]

where \( y_u \) is the observed response in the \( u \)th experimental run when \( k \) variables \( \xi \) are held at levels \( \xi_u \), \( \theta \) is a set of \( p \) parameters and \( e_u \) is the error. The state of ignorance with which we start and the state of comparative knowledge which we wish to brought to, will determine our approach. It is realized, of course, that no real problem ever quite fits any prearranged category. With this proviso it is perhaps helpful to distinguish the following types of problems and for reference purposes I have given each a tentative name.

<table>
<thead>
<tr>
<th>SUPOSED UNKNOWN</th>
<th>OBJECTIVE</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>Determine the subset ( \xi ) of important variables from a given larger set ( \xi ) of potentially important variables</td>
<td>Screening variables</td>
</tr>
<tr>
<td>( \xi )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f )</td>
<td>Determine the empirical &quot;effects&quot; of the known input variables ( \xi ).</td>
<td>Empirical model building</td>
</tr>
<tr>
<td>( \theta )</td>
<td></td>
<td>Response surface methods</td>
</tr>
<tr>
<td>( f )</td>
<td>Determine a local interpolation approximation ( F(\xi</td>
<td>\theta) ) to ( f(\xi</td>
</tr>
<tr>
<td>( \theta )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f )</td>
<td>Determine ( f )</td>
<td>Mechanistic model building</td>
</tr>
<tr>
<td>( \theta )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta )</td>
<td>Determine ( \theta )</td>
<td>Mechanistic model fitting</td>
</tr>
</tbody>
</table>
Figure 2(a). Iterative Experimentation

Figure 2(b). The iterative experimental process in relation to the experimenter
In the time available I will briefly sketch some of the procedures that have been employed to deal with these problems taking rather more time over problem (4).

4. SCREENING VARIABLES WITH SEQUENCES OF FRACTIONAL FACTORIALS

One source of variable screening problems is the "trouble shooting" investigation. An example of such a problem and of iterative experimentation leading to its solution is given in [3]. A filtration process which should have given filtration times \((y)\) in the neighborhood of 40 minutes in fact gave times close to 70 minutes thus creating a production bottleneck. A set \(\mathcal{S}\) of seven variables which hopefully included those responsible for the trouble were (1) type of water, (2) type of raw material, (3) temperature of filtration, (4) hold up time, (5) use of recycle, (6) rate of caustic addition, and (7) type of filter cloth. It seemed likely that at most two of these variables had any real effect over the relevant ranges. The problem was to discover, in the presence of large inherent variability, which ones, if any, of these candidate variables were responsible.

A 1/16 replicate of a \(2^7\) (resolution III) design was first run. The design has the property that it supplies a \(2^2\) factorial replicated twice in every pair of variables. On the assumption that only main effects and two factor interactions need be considered the data from this eight-run design could be explained in terms of effects \((1,3,6)\); \((1,3,1\times 3)\); \((1,6,1\times 6)\) or \((3,6,3\times 6)\). Here, for example, 1 refers to the main effects of factor 1, and \(1\times 3\) refers to its interaction effect with 3. One should not be distressed by ambiguity which still exists at this stage but elated at the thought that the
number of possibilities has been drastically reduced by a very few experimental runs. Proceeding to the next iterative cycle a second block of eight runs was made. This second block was specifically selected to resolve the ambiguities revealed by the first set. It lead to the conclusion that what was being seen were the effects of type of water (1), rate of caustic addition (6) and their interaction (1×6) and this lead to a speedy solution to the problem.

Iterative experimentation with fractional factorials as building blocks has proved a potent tool in the solution of many similar problems.

5. EMPIRICAL MODEL BUILDING

It is perhaps worth saying that the randomized block, latin square, and factorial designs have since their inception been used as building blocks in the iterative learning sequence. The possibility of rapid convergence of the iterative sequence depends on the efficiency with which the dual processes of data generation (design) and data analysis are performed and how they illuminate and stimulate (and do not stifle) the ideas of the investigator. This notion was clear in Fisher's general attitude towards the use of these designs in scientific investigation.

When the variables are quantitative and the experimental error is not too large it may be profitable to attempt to estimate the response function within some area of immediate interest. In many problems the form of the true response function $f(\xi;\theta)$ is unknown and cannot economically be obtained but may perhaps be locally approximated by a polynomial or some other graduating function $F(\xi,\theta)$ and designs have been developed with this end in mind. The essentially
number of possibilities has been drastically reduced by a very few experimental runs. Proceeding to the next iterative cycle a second block of eight runs was made. This second block was specifically selected to resolve the ambiguities revealed by the first set. It lead to the conclusion that what was being seen were the effects of type of water (1), rate of caustic addition (6) and their interaction (1×6) and this lead to a speedy solution to the problem.

Iterative experimentation with fractional factorials as building blocks has proved a potent tool in the solution of many similar problems.

5. EMPIRICAL MODEL BUILDING

It is perhaps worth saying that the randomized block, latin square, and factorial designs have since their inception been used as building blocks in the iterative learning sequence. The possibility of rapid convergence of the iterative sequence depends on the efficiency with which the dual processes of data generation (design) and data analysis are performed and how they illuminate and stimulate (and do not stifle) the ideas of the investigator. This notion was clear in Fisher's general attitude towards the use of these designs in scientific investigation.

When the variables are quantitative and the experimental error is not too large it may be profitable to attempt to estimate the response function within some area of immediate interest. In many problems the form of the true response function \( f(\xi|\theta) \) is unknown and cannot economically be obtained but may perhaps be locally approximated by a polynomial or some other graduating function \( F(\xi,\beta) \) and designs have been developed with this end in mind. The essentially
In the time available I will briefly sketch some of the procedures that have been employed to deal with these problems taking rather more time over problem (4).

4. SCREENING VARIABLES WITH SUBQUENCES OF FRACTIONAL FACTORIALS

One source of variable screening problems is the "trouble shooting" investigation. An example of such a problem and of iterative experimentation leading to its solution is given in [3]. A filtration process which should have given filtration times \( y \) in the neighborhood of 40 minutes in fact gave times close to 70 minutes thus creating a production bottleneck. A set \( \mathcal{E} \) of seven variables which hopefully included those responsible for the trouble were (1) type of water, (2) type of raw material, (3) temperature of filtration, (4) hold up time, (5) use of recycle, (6) rate of caustic addition, and (7) type of filter cloth. It seemed likely that at most two of these variables had any real effect over the relevant ranges. The problem was to discover, in the presence of large inherent variability, which ones, if any, of these candidate variables were responsible.

A 1/16 replicate of a \( 2^7 \) (resolution III) design was first run. The design has the property that it supplies a \( 2^2 \) factorial replicated twice in every pair of variables. On the assumption that only main effects and two factor interactions need be considered the data from this eight-run design could be explained in terms of effects \( (1, 3, 6); (1, 3, 1 \times 3); (1, 6, 1 \times 6) \) or \( (3, 6, 3 \times 6) \). Here, for example, \( 1 \) refers to the main effects of factor 1, and \( 1 \times 3 \) refers to its interaction effect with 3. One should not be distressed by ambiguity which still exists at this stage but elated at the thought that the
iterative nature of so called "response surface methodology" is well known and illustrates the multicomponent character of the more general learning process.

Thus, in a typical investigation to improve a chemical process [4] the investigator would be concerned to learn as he went along (i) the amount of replication needed to achieve sufficient accuracy, (ii) the location of the experimental region of interest, (iii) an appropriate scaling and transformation for the variables, (iv) the degree of complexity of approximating function and hence of design needed.

6. MECHANISTIC MODEL BUILDING

In recent years considerable attention has been given to the possibility of using the true function form \( f(x, \theta) \) to represent the response rather than approximating it by \( R(x, \theta) \). To do this one seeks to discover the mechanism for the phenomenon and to express this mechanism in appropriate mathematical form. In practice we can hope to achieve useful working mechanistic models which take account of the principal elements of the mechanism. The models often are most naturally expressed in terms of differential equations or other non-explicit forms but developments in computing and in the better understanding of non-linear design and estimation have made it possible to cope with the resulting problems. A mechanistic model has the advantage that

1) it contributes to our scientific understanding of the phenomenon under study

2) it should provide a better basis for extrapolation at least to conditions worthy of further experimental investigation
iii) it tends to be parsimonious in the use of parameters and hence to provide better estimates of the response. Results from fitting mechanistic models have sometimes been disappointing because not enough attention has been given to discovering what is the appropriate model. It is easy to collect data which never "place the assumed model in jeopardy" and so it is common (for example, in chemical engineering) to find different research groups each advocating a different model for the same phenomenon and proffering data which "proves" their claim.

For some time statisticians and chemical engineers have been working in cooperation on this problem at Wisconsin. So far two main techniques have resulted

i) a diagnostic input perturbation technique [5] [6] [7]

ii) a sequential discrimination technique [8] [9] [10]

6.1 Diagnostic input perturbation.

Often the model is most naturally expressed in terms of time or space dependent differential equations subject to certain boundary conditions. On integration these yield

\[ E(y) = f(\xi,\theta,t) \]  (1)

where \( t \) is the time or space variable. In a chemical study \( y \) would typically refer to the yield of certain products at time \( t \), and the \( \theta \)'s would be parameters such as rate constants. In some cases the yields of several (say \( t \)) products of reaction would be studied simultaneously. So that there
iii) it tends to be parsimonious in the use of parameters and hence to provide better estimates of the response.

Results from fitting mechanistic models have sometimes been disappointing because not enough attention has been given to discovering what is the appropriate model. It is easy to collect data which never "place the assumed model in jeopardy" and so it is common (for example, in chemical engineering) to find different research groups each advocating a different model for the same phenomenon and proffering data which "proves" their claim.

For some time statisticians and chemical engineers have been working in cooperation on this problem at Wisconsin. So far two main techniques have resulted

i) a diagnostic input perturbation technique [5] [6] [7]

ii) a sequential discrimination technique [8] [9] [10]

6.1 Diagnostic input perturbation.

Often the model is most naturally expressed in terms of time or space dependent differential equations subject to certain boundary conditions. On integration these yield

$$E(y) = f(e, \theta, t)$$  \hspace{1cm} (1)

where \( t \) is the time or space variable. In a chemical study \( y \) would typically refer to the yield of certain products at time \( t \), and the \( \theta \)'s would be parameters such as rate constants. In some cases the yields of several (say \( l \)) products of reaction would be studied simultaneously. So that there
iterative nature of so called "response surface methodology" is well known and illustrates the multicomponent character of the more general learning process.

Thus, in a typical investigation to improve a chemical process [4] the investigator would be concerned to learn as he went along (i) the amount of replication needed to achieve sufficient accuracy, (ii) the location of the experimental region of interest, (iii) an appropriate scaling and transformation for the variables, (iv) the degree of complexity of approximating function and hence of design needed.

6. MECHANISTIC MODEL BUILDING

In recent years considerable attention has been given to the possibility of using the true function form $f(x, \theta)$ to represent the response rather than approximating it by $R(x, \theta)$. To do this one seeks to discover the mechanism for the phenomenon and to express this mechanism in appropriate mathematical form. In practice we can hope to achieve useful working mechanistic models which take account of the principal elements of the mechanism. The models often are most naturally expressed in terms of differential equations or other non-explicit forms but developments in computing and in the better understanding of non-linear design and estimation have made it possible to cope with the resulting problems. A mechanistic model has the advantage that

i) it contributes to our scientific understanding of the phenomenon under study

ii) it should provide a better basis for extrapolation at least to conditions worthy of further experimental investigation
would then be \( \xi \) functions of the form of equation (1) supplying information about the same \( \theta \)'s. Multivariate estimation of the \( \theta \)'s in this situation has been discussed in [10].

When investigating such a system it is usually most convenient to set the variables \( \xi \) at some fixed levels and observe \( y \) at \( m \) specified values of \( t \). The values of \( \xi \) are then reset at a second set of levels and so on through \( n \) experimental runs. It would often be convenient to arrange these \( n \) runs so that the variables \( \xi \) followed some factorial or fractional factorial arrangement.

Now suppose (using non-linear estimation methods [11] [12] [13] [14]) we obtained estimates \( \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n \) of the \( p \) parameters for each of the \( n \) experimental runs. In the simplest application of the technique our assumed model would imply that these quantities should estimate the same set of parameters \( \theta \), and in particular that they should be independent of the levels of the experimental variables \( \xi \). The usual factorial analysis may now be applied not to the \( y \)'s but to the \( \hat{\theta} \)'s. The existence of significant effects points to inadequacy of the model. Much more importantly the analysis supplies a diagnostic tool indicating at precisely what point the inadequacy occurs.

Thus, for example, if \( \theta_2 \) was a particular rate constant and \( \xi_3 \) an initial concentration of a certain reactant then correlation of \( \hat{\theta}_2 \) and \( \xi_3 \) would imply that the inadequacy of the model was to be looked for at the point where \( \theta_2 \) and \( \xi_3 \) were associated. This technique and particularly its contribution to iterative learning is well illustrated in a study of the kinetic mechanism of the catalytic oxidation of methane [6]. With this technique
we do not necessarily have any particular alternative model in mind. We are attempting rather to strain the model in such a way that the nature of any inadequacy which is shown up may jog the imagination of the experimenter to the right kind of new conjecture.

6.7 Sequential discrimination.

A different case arises if we have $m$ rival and plausible models which limited experimentation has so far been ineffective in distinguishing between. We ask the question "Where should the next run be performed so as to allow us to best discriminate between the alternatives?" In a recent paper [10] the following approach has been followed. Let $p_i$ ($i = 1, 2, \ldots, m$) be the probability density of the $N'$th observation given the first $N-1$ observations on the assumption that the $i'$th model is true.

Let $\pi_{iN-1}$ denote the probability that the $i$'th model is true when $N-1$ observations have been performed. (We should normally assume that $\pi_{i0} = \frac{1}{m}$ ($i = 1, 2, \ldots, m$)). Then using Bayes' theorem

$$\pi_{iN} = \frac{\pi_{iN-1} p_i}{\sum_{i=1}^{m} \pi_{iN-1} p_j}.$$ 

Let $E_i$ denote the expected value on the assumption that model $i$ is correct. Then given that $N-1$ experimental runs have been performed we choose the $N'$th run so as to maximize the measure of divergence

$$D = \sum_{i=1}^{m} \sum_{j \neq i}^{m} \pi_{iN-1} \pi_{jN-1} \left( E_i \ln \frac{p_i}{p_j} + E_j \ln \frac{p_j}{p_i} \right).$$
we do not necessarily have any particular alternative model in mind. We are attempting rather to strain the model in such a way that the nature of any inadequacy which is shown up may jog the imagination of the experimenter to the right kind of new conjecture.

6.2 Sequential discrimination.

A different case arises if we have m rival and plausible models which limited experimentation has so far been ineffective in distinguishing between. We ask the question "Where should the next run be performed so as to allow us to best discriminate between the alternatives?" In a recent paper \[9\] \([16]\) the following approach has been followed. Let \(p_i\) (\(i = 1, 2, ..., m\)) be the probability density of the \(N\)'th observation given the first \(N-1\) observations on the assumption that the \(i\)'th model is true.

Let \(\pi_{i|N-1}\) denote the probability that the \(i\)'th model is true when \(N-1\) observations have been performed. (We should normally assume that \(\pi_i = \frac{1}{m} \) (\(i = 1, 2, ..., m\)). Then using Bayes' theorem

\[
\pi_{i|N} = \frac{\pi_{i|N-1} p_i}{\sum_{j=1}^{m} \pi_{j|N-1} p_j}
\]

Let \(E_i\) denote the expected value on the assumption that model \(i\) is correct. Then given that \(N-1\) experimental runs have been performed we choose the \(N\)'th run so as to maximize the measure of divergence

\[
D = \sum_{i=1}^{m} \sum_{j|i} \pi_{i|N-1} \pi_{j|N-1} \left( E_i \ln \frac{p_i}{p_j} + E_j \ln \frac{p_j}{p_i} \right)
\]
would then be \( \hat{\theta} \) functions of the form of equation (1) supplying information about the same \( \theta \)'s. Multivariate estimation of the \( \theta \)'s in this situation has been discussed in [10].

When investigating such a system it is usually most convenient to set the variables \( \xi \) at some fixed levels and observe \( y \) at \( n \) specified values of \( t \). The values of \( \xi \) are then reset at a second set of levels and so on through \( n \) experimental runs. It would often be convenient to arrange these \( n \) runs so that the variables \( \xi \) followed some factorial or fractional factorial arrangement.

Now suppose (using non-linear estimation methods [11] [12] [13] [14]) we obtained estimates \( \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n \) of the \( p \) parameters for each of the \( n \) experimental runs. In the simplest application of the technique our assumed model would imply that these quantities should estimate the same set of parameters \( \theta \), and in particular that they should be independent of the levels of the experimental variables \( \xi \). The usual factorial analysis may now be applied not to the \( y \)'s but to the \( \hat{\theta} \)'s. The existence of significant effects points to inadequacy of the model. Much more importantly the analysis supplies a diagnostic tool indicating at precisely what point the inadequacy occurs.

Thus, for example, if \( \theta_2 \) was a particular rate constant and \( \xi_3 \) an initial concentration of a certain reactant then correlation of \( \hat{\theta}_2 \) and \( \xi_3 \) would imply that the inadequacy of the model was to be looked for at the point where \( \theta_2 \) and \( \xi_3 \) were associated. This technique and particularly its contribution to iterative learning is well illustrated in a study of the kinetic mechanism of the catalytic oxidation of methane [6]. With this technique
and the $\bar{N}_{i+1}$ are recalculated after each run until adequate discrimination is obtained. The maximization of $D$ corresponds to choosing the $N'$th run so that the expected change in entropy between the $N-1$'th and the $N$'th runs is maximized.

**Discrimination among rate models:**

Suppose that we look at the simple reaction

$$A + B$$

Then depending on whether the reaction is of first, second, third, or fourth order the relationships between $y$, the amount of unreacted $A$, and $\xi_1$, time in minutes, and $\xi_2$, the temperature, are

**Model 1.**

$$E(y) = e^{-\theta_{11}\xi_1 e^{-\theta_{12}/\xi_2}}$$

**Model 2.**

$$E(y) = \frac{1}{1 + \theta_{21}\xi_1 e^{-\theta_{22}/\xi_2}}$$

**Model 3.**

$$E(y) = \frac{1}{1 + 2\theta_{31}\xi_1 e^{-\theta_{32}/\xi_2}}^{1/2}$$

**Model 4.**

$$E(y) = \frac{1}{1 + 3\theta_{41}\xi_1 e^{-\theta_{42}/\xi_2}}^{1/3}$$

where the $\theta_{ij}$ are the constants of the system. In each case the initial concentration of $A$ is taken to be unity. In a simulated study, model 2 (second order reaction) was chosen to be the correct model with

$$E(y) = \frac{1}{1 + 400\xi_1 e^{-5000/\xi_2}}$$

with $y$'s normally distributed and standard deviation .05. It was supposed that
the operability region was defined by \( 0 \leq \xi_1 \leq 150 \) minutes and
\[ 450 \leq \xi_2 \leq 600 \degree K. \]
A preliminary \( 2^2 \)-factorial design was chosen within this region at levels \( \xi_1 = 25 \) and \( 125 \) minutes and \( \xi_2 = 475 \) and \( 575 \degree K. \)

The points where maximum discrimination was obtained at each stage are shown in Figure 3. The posterior probabilities are shown in Figure 4 where the initial priors were taken to be locally uniform
\[(\pi_{10} = \pi_{20} = \pi_{30} = \pi_{40} = .25).\]
After only \( N = 8 \) observations, model 2 was correctly favored over the rest even though after the preliminary \( N = 4 \) observations there was a slight preference for model 3.

7. MECHANISTIC MODEL FITTING

Non-linear model fitting it seems to me has too often received emphasis at the expense of model building because the first is a well-defined problem and the second is not. However, in those cases where we can suppose the model to be "known" most emphasis is needed on design and particularly sequential design. The purely numerical problems arising in the choice of an optimal design [15] are much more manageable in the fairly common circumstance where runs can be made one at a time [16] [17]. One can then select runs experimental / which at each stage maximize (or at least make large) some relevant measure of expected information such as the expected change in volume of the confidence region.
the operability region was defined by $0 \leq \xi_1 \leq 150$ minutes and $450 \leq \xi_2 \leq 600 \, ^{\circ} \text{K}$. A preliminary $2^2$-factorial design was chosen within this region at levels $\xi_1 = 75$ and 125 minutes and $\xi_2 = 475$ and $575 \, ^{\circ} \text{K}$.

The points where maximum discrimination was obtained at each stage are shown in Figure 3. The posterior probabilities are shown in Figure 4 where the initial priors were taken to be locally uniform $(\pi_{10} = \pi_{20} = \pi_{30} = \pi_{40} = .25)$. After only $N = 8$ observations, model 2 was correctly favored over the rest even though after the preliminary $N = 4$ observations there was a slight preference for model 3.

7. MECHANISTIC MODEL FITTING

Non-linear model fitting it seems to me has too often received emphasis at the expense of model building because the first is a well-defined problem and the second is not. However, in those cases where we can suppose the model to be "known" most emphasis is needed on design and particularly sequential design. The purely numerical problems arising in the choice of an optimal design [15] are much more manageable in the fairly common circumstance where runs can be made one at a time [16] [17]. One can then select experimental runs which at each stage maximize (or at least make large) some relevant measure of expected information such as the expected change in volume of the confidence region.
and the $E_{iN-1}$ ($i = 1, 2, \ldots, m$) are recalculated after each run until adequate discrimination is obtained. The maximization of $D$ corresponds to choosing the $N$'th run so that the expected change in entropy between the $N-1$'th and the $N$'th runs is maximized.

**Discrimination among rate models:**

Suppose that we look at the simple reaction

$$A \rightarrow B$$

Then depending on whether the reaction is of first, second, third, or fourth order the relationships between $y$, the amount of unreacted $A$, and $\xi_1$, time in minutes, and $\xi_2$, the temperature, are

**Model 1.**  
$$E(y) = e^{-\theta_{11}\xi_1} e^{-\theta_{12}/\xi_2}$$

**Model 2.**  
$$E(y) = 1 / \left(1 + \theta_{21}\xi_1 e^{-\theta_{22}/\xi_2}\right)$$

**Model 3.**  
$$E(y) = 1 / \left(1 + 2\theta_{31}\xi_1 e^{-\theta_{32}/\xi_2}\right)^{1/2}$$

**Model 4.**  
$$E(y) = 1 / \left(1 + 3\theta_{41}\xi_1 e^{-\theta_{42}/\xi_2}\right)^{1/3}$$

where the $\theta_{ij}$ are the constants of the system. In each case the initial concentration of $A$ is taken to be unity. In a simulated study, model 2 (second order reaction) was chosen to be the correct model with

$$E(y) = 1 / \left(1 + 400\xi_1 e^{-5000/\xi_2}\right)$$

with $y$'s normally distributed and standard deviation 0.05. It was supposed that
1 - 4 Initial points
5 - 9 Sequentially chosen points

Also shown are theoretical contours of unreacted A

Figure 3. Location of runs in a sequential discrimination study
Initially

<table>
<thead>
<tr>
<th></th>
<th>10 * .25</th>
<th>20 * .25</th>
<th>30 * .25</th>
<th>40 * .25</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After \( n = 4 \)

Observation

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( N = 5 \)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.56</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( N = 6 \)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( N = 7 \)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.09</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( N = 8 \)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Posterior probabilities for four reaction rate models
Initially

<table>
<thead>
<tr>
<th>10^2.25</th>
<th>20^2.25</th>
<th>30^2.25</th>
<th>40^2.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

After \( N = 3 \)

\[
\begin{array}{c|c|c|c|c}
\text{Preliminary Observations} & 1 & 2 & 3 & 4 \\
\hline
1 & .10 & 1.50 & .05 & 1.05 \\
\end{array}
\]

\( N = 5 \)

\[
\begin{array}{c|c|c|c|c}
\xi_1 & 125 & .56 & .43 & 0.01 \\
\xi_2 & 600 & 1 & 2 & 3 & 4 \\
\end{array}
\]

\( N = 6 \)

\[
\begin{array}{c|c|c|c|c}
\xi_1 & 125 & 0.01 & 1.13 & 0.00 \\
\xi_2 & 600 & 1 & 2 & 3 & 4 \\
\end{array}
\]

\( N = 7 \)

\[
\begin{array}{c|c|c|c|c}
\xi_1 & 50 & 0.00 & 0.03 & 0.00 \\
\xi_2 & 450 & 1 & 2 & 3 & 4 \\
\end{array}
\]

\( N = 8 \)

\[
\begin{array}{c|c|c|c|c}
\xi_1 & 100 & 0.00 & 0.00 & 0.00 \\
\xi_2 & 600 & 1 & 2 & 3 & 4 \\
\end{array}
\]

**Figure 4.** Posterior probabilities for four reaction rate models
Figure 5. Location of runs in a sequential discrimination study

(1) - (4) Initial points
5 - 9 Sequentially chosen points

Also shown are theoretical contours of unreacted A
REFERENCES


## EXPERIMENTAL STRATEGY

### 6. April, 1967

<table>
<thead>
<tr>
<th>7a. 17 pp.</th>
<th>7b. 16 references</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>8a. Nonr 1202(17)</th>
<th>9a. Report #111</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b. NR 042-222</td>
<td>9b. --</td>
</tr>
</tbody>
</table>

10. Distribution of this document is unlimited

12. Office of Naval Research
Washington, D. C.

13. Abstract: In this paper the basic iterative nature of experimental learning process is emphasized. The role of iteration is then discussed for a number of specific classes of problems. These include the screening of variables, empirical model building, mechanistic model fitting. Recent contributions to the important problem of mechanistic model building are emphasized.
13. Abstract: In this paper the basic iterative nature of experimental learning process is emphasized. The role of iteration is then discussed for a number of specific classes of problems. These include the screening of variables, empirical model building, mechanistic model fitting. Recent contributions to the important problem of mechanistic model building are emphasized.
REFERENCES


