Technical Report Number 65

February 1966

Design of Experiments for Model Discrimination in Multiresponse Situations

by William J. Hill* and William G. Hunter,
University of Wisconsin

* W. J. Hill was supported by a fellowship grant from Monsanto.

This research was supported in part by the National Science Foundation under grant GP-2755.
1. INTRODUCTION

Multiresponse data, which often arise in experimental situations, can be analysed in terms of a model of the following form

\[ \eta_j^{(v)} = f_j^{(v)} (\theta_1^{(v)}, \theta_2^{(v)}, \ldots, \theta_p^{(v)}; \xi_1, \xi_2, \ldots, \xi_k) \]

\[ = f_j^{(v)} (\theta^{(v)}, \xi) \quad j = 1, 2, \ldots, r \]  \hspace{1cm} (1.1)

where there are \( r \) response functions \( f_j^{(v)} (\theta^{(v)}, \xi) \) each of which depends upon \( p \) parameters \( \theta^{(v)} \) and \( k \) experimental variables \( \xi \). The superscript \( v \) is an index to designate the particular model under study. This notation is necessary when more than one model is being considered. The number of parameters need not be the same for all models, that is, \( p \) is a function of \( v \). An experimental run consists of setting the levels of the variables at some preselected values \( \xi = (\xi_1, \xi_2, \ldots, \xi_k)' \) and making simultaneous observations \( y = (y_1, y_2, \ldots, y_r)' \) on the \( r \) responses. For example, consider the equations appropriate for a second-order chemical reaction of the type \( A + B \). The appropriate model consists of two response functions

\[ \eta_1^{(1)} = \left\{ 1 + \theta_1^{(1)} \xi_1 \exp \left( -\theta_2^{(1)}/\xi_2 \right) \right\}^{-1} \]

\[ \eta_2^{(1)} = 1 - \left\{ 1 + \theta_1^{(1)} \xi_1 \exp \left( -\theta_2^{(1)}/\xi_2 \right) \right\}^{-1} \]  \hspace{1cm} (1.2)

where \( \eta_1^{(1)} \) and \( \eta_2^{(1)} \) are the concentrations of \( A \) and \( B \) respectively, \( \xi_1 \) is the reaction time, \( \xi_2 \) is the temperature, and \( \theta_1^{(1)} \) and \( \theta_2^{(1)} \) are the parameters for this model. (The response functions (1.2) can be derived from the following
set of equations:

\[-\theta_1^{(1)} / \xi_1 = \theta_1^{(1)} \left( \frac{\eta_1^{(1)}}{\xi_1} \right)^2 \text{ and } \eta_2^{(1)} = 1 - \eta_1^{(1)} \text{ where} \]

\[\theta_1^{(1)} = \theta_1^{(1)} \exp \left( -\theta_2^{(1)}/\xi_2 \right) \text{ and } \eta_1^{(1)} = 1 \text{ when } \xi_1 = 0. \]

An experimental run consists of measuring the amounts of A and B (\(y_1\) and \(y_2\), respectively) that are present after the reaction has proceeded for time \(\xi_1\) at temperature \(\xi_2\). A parameter estimation procedure for multiresponse data of this type has been developed by Box and Draper (1965) so that it is now possible to consider a number of associated problems in experimental design. In the present paper we discuss one such problem, that of model discrimination. Suppose that \(n-1\) runs have been made and the results are not conclusive as to which of \(m(>1)\) rival models is the best model. An experimenter in such a situation might wish to perform further experimental runs in order to discriminate among these \(m\) rival models.

The following equations, for example, are the two response functions appropriate for a third-order chemical reaction of the type \(A + B\)

\[
\eta_1^{(2)} = \left( \frac{1 + 2\theta_1^{(2)}}{1 + 2\theta_1^{(2)}} \right) \xi_1 \exp \left( -\theta_2^{(2)}/\xi_2 \right)^{-\frac{1}{2}},
\]

\[
\eta_2^{(2)} = 1 - \left( \frac{1 + 2\theta_1^{(2)}}{1 + 2\theta_1^{(2)}} \right) \xi_1 \exp \left( -\theta_2^{(2)}/\xi_2 \right)^{-\frac{1}{2}},
\]

where \(\eta_1^{(2)}\) and \(\eta_2^{(2)}\) are the concentrations of A and B respectively, \(\xi_1\) is the reaction time, \(\xi_2\) is the temperature, and \(\theta_1^{(2)}\) and \(\theta_2^{(2)}\) are the parameters for this model. (The response functions (1.3) can be derived from the
following set of equations:

\[-3n_1(2)/\partial \xi_1 = \delta(2) \left( \eta_1(2) \right)^3 \text{ and } n_2(2) = 1 - n_1(2) \text{ where} \]

\[\delta(2) = \delta_1(2) \exp (-\delta_2(2)/\xi_2) \text{ and } n_1^2 = 1 \text{ when } \xi_1 = 0.\]

With the available data \((\xi_u = (\xi_{1u}, \xi_{2u})', \text{ and } y_u = (y_{1u}, y_{2u})', u = 1, 2, \ldots, n-1)\) both models (1.2) and (1.3) appear adequate. To resolve this ambiguity, the experimenter may wish to perform further runs and, if so, the question arises as to which runs should be selected for this purpose.

Sequential design procedures, where at stage \(n-1\) the best settings for the \(n\)-th run are determined, have been proposed for the single-response case \((r=1)\) by Hunter and Reiner (1965) and by Box and Hill (1967); however, no work appears to have been done on the multiresponse case \((r >1)\). In the following sections a multiresponse discrimination function, which is a generalization of the one developed by Box and Hill, is proposed as a design criterion. Accordingly, those settings \(\xi_n = (\xi_{1n}, \xi_{2n}, \ldots, \xi_{kn})'\) of the experimental variables which maximise this function are selected for the \(n\)-th run.

2. FORMULATION

At the outset of an investigation on a particular system, suppose that the underlying theory suggests not one but several (say, \(m\)) rival models of the form (1.1). Suppose at some later stage of experimentation, after multiresponse data of the form

\[\xi_u = (\xi_{1u}, \xi_{2u}, \ldots, \xi_{ku})' \text{ and } y_u = (y_{1u}, y_{2u}, \ldots, y_{ru})', u = 1, 2, \ldots, n-1,\]
are available, the results are still inconclusive as to which is the best model. The experimenter may then wish to perform further experiments to clarify the situation. A design criterion D is proposed which can be used in choosing additional runs. Using it, one selects for the next run those operating conditions $x_n$ in the operability region that will provide maximum discrimination.

Under model $v$ it is assumed that the observations are $y_n = (y_{1n}, y_{2n}, \ldots, y_{xn})'$ from the yet-to-be performed $n$-th run are normally distributed with expected values $\eta_n(v) = (\eta_1(v), \eta_2(v), \ldots, \eta_X(v))$, and known variance-covariance matrix $V$. The probability density of $y_n$ given $\eta_n(v)$ and $V$ is

$$p(y_n \mid \eta_n(v), V) = \frac{|V|^{-1/2}}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (y_n - \eta_n(v))' V^{-1} (y_n - \eta_n(v)) \right\}. \quad (2.1)$$
Suppose that for a region in the \( \hat{\theta}^{(v)} \) space sufficiently close to the parameter estimates \( \hat{\theta}^{(v)} \) (based on n-1 observations), the function for the j-th response can be approximately expressed in the linear form

\[
\eta_j^{(v)} = f_j^{(v)}(\hat{\theta}^{(v)}, \xi_n) + \sum_{k=1}^{p} (\theta_k - \hat{\theta}_k) x_{jn}^{(v)},
\]

where

\[
x_{jn}^{(v)} = \left[ \frac{\partial f_j^{(v)}(\hat{\theta}^{(v)}, \xi_n)}{\partial \theta_k^{(v)}} \right] \theta_k^{(v)} = \hat{\theta}_k^{(v)}.
\]

Locally, if \( \eta^{(v)} \) is assumed to have a uniform prior distribution, then from a Bayesian viewpoint \( \eta^{(v)} \) can be considered as being normally distributed about \( \bar{\eta}^{(v)} \) with variance-covariance matrix \( R_{\eta \eta} \) (equal to the variance-covariance matrix of \( \bar{\eta}^{(v)} \)) where \( \bar{\eta}^{(v)} \) are the r predicted values of the n-th observations using the data from the first n-1 runs. The probability density of \( \eta^{(v)} \) can then be written as

\[
p(\eta^{(v)} | \xi_n) = \left| \frac{R_{\eta \eta}}{(2\pi)^{n/2}} \right|^{-1/2} \exp \left( -\frac{1}{2} (\eta^{(v)} - \bar{\eta}^{(v)}) R_{\eta \eta}^{-1} (\eta^{(v)} - \bar{\eta}^{(v)}) \right).
\]

It follows that the probability density of \( \eta^{(v)} \) under model v given the first n-1 runs and the variance-covariance matrix \( V \) is

\[
p(v | \eta^{(v)}) = \int_{\eta^{(v)}} p(\eta^{(v)} | \xi_n) \frac{1}{p(v | \eta^{(v)})} | \eta^{(v)} \rangle d\eta^{(v)}. \quad (2.3)
\]
After integration, this expression simplifies to

\[ p(v) = \frac{1}{(2\pi)^{n/2}} \sum_{vn} \frac{-1/2 \exp \left( -\frac{1}{2} \sum_{vn} (y_n - \bar{y}_n)^2 \right) \sum_{vn} (y_n - \bar{y}_n)^2}{(2\pi)^{n/2}} , \tag{2.4} \]

where

\[ \sum_{vn} = V + R_{vn} \tag{2.5} \]

If the prior probability associated with model \( v \) for the \( n \)-th run is \( \pi_{v, n-1} \), then the posterior probability associated with model \( v \) is

\[ \pi_{vn} = \frac{\pi_{v, n-1} p(v)}{\sum_{i=1}^{m} \pi_{i, n-1} p(i)} \tag{2.6} \]

For the single-response case, Box and Hill (1967) used information theory as described in Kullback (1959) and Shannon (1948) to develop a function for discriminating among \( m \) rival models.

A generalization of this discrimination function in the multiresponse situation can be obtained by replacing \( y_n \) everywhere by \( \bar{y}_n \) which gives

\[ D = \sum_{v=1}^{m} \sum_{i=v+1}^{m} \pi_{v, n-1} \pi_{i, n-1} \left\{ \int_{\bar{y}_n} p(v) \ln \frac{p(v)}{p(i)} \, d\bar{y}_n \right\} + \int_{\bar{y}_n} p(i) \ln \frac{p(i)}{p(v)} \, d\bar{y}_n \tag{2.7} \]
Substituting (2.4) into this expression and integrating we have

\[
D = \frac{1}{2} \sum_{v=1}^{m} \sum_{i=v+1}^{m} \Pi_{i,n}^{v,n} \Pi_{i,n}^{v,n-1} \text{trace} \left[ \sum_{vn}^{v} \sum_{in}^{i} + \sum_{in}^{i} \sum_{vn}^{v} - 2I_r \right] 
\]

\[
+ \left( \bar{y}_{n}^{(v)} - \bar{y}_{n}^{(i)} \right)^{T} \left( \sum_{vn}^{v} \sum_{in}^{i} \right) \left( \bar{y}_{n}^{(v)} - \bar{y}_{n}^{(i)} \right), \quad (2.8)
\]

where \(I_r\) is an \(r \times r\) identity matrix. Expression (2.8) can be used as a design criterion for selecting experimental conditions for discriminating among \(m\) rival models when \(r\) responses can be measured and the variance-covariance matrix \(V\) for the \(r\) responses is known. (When \(r = 1\), Equation (2.8) is, of course, identical to the discrimination function developed by Box and Hill (1967).) For maximum discrimination the experimenter chooses those operating conditions \(z_n\) that maximize \(D\) and then performs the \(n\)-th run at these optimum operating conditions. Once \(y_n\) are observed the relative merit of each model is checked by calculating its corresponding posterior probability from (2.6). If it is still not possible to distinguish among the rival models, this procedure could be repeated until the posterior probabilities indicate that one model is clearly superior to the others.
3. **EXAMPLE**

Suppose an experimenter is studying a chemical reaction of the type \( A \rightarrow B \) for which two rival models (1.2) and (1.3) are being considered. We will use a constructed example in which all the data are generated from model (1.2) (denoted as model 1 as opposed to its rival, model 2) where the true values of the constants are \( \theta_1(1) = 400 / \theta_2(1) = 5000 \). (In an actual experimental situation, of course, the true model and the corresponding parameter values would be unknown.) The variance-covariance matrix

\[
V = 1.25 (10)^{-3} \begin{pmatrix}
2 & 1 \\
1 & 2
\end{pmatrix}
\]
is assumed known.

The region in which experiments can be conducted in the \((\xi_1, \xi_2)\) space is limited by the following constraints: \(0 \leq \xi_1 \leq 150\) and \(450 \leq \xi_2 \leq 600\).

It is desired to select design points in units of 25 for both \(\xi_1\) and \(\xi_2\). Suppose the initial set of four runs shown in Table 1 have been performed.

Before considering the double-response case \((r = 2)\), it is interesting to consider the single-response case \((r = 1)\) where only the data on response A are used. Before any experimentation the prior probabilities \(\Pi_{10}\) and \(\Pi_{20}\) associated with models 1 and 2, respectively, are taken to be \(.5\).

Even after the initial four runs both models appear to fit the data equally well as indicated by their respective posterior probabilities which are given in Table 2. In fact, model 2 (the wrong model) is slightly favoured at this stage \((\Pi_{24} = .5386\) vs. \(\Pi_{14} = .4614\)). The discrimination function \(D\), Equation (2.8) with \(m = 2\), \(r = 1\), and \(n = 5\), is maximised on the grid \(\xi_1 = 0(25)\) 150,

\[
\xi_2 = 450(25)600
\]

when \((\xi_{15}, \xi_{25}) = (125, 600)\). An observation on response A is generated at this point, the resulting posteriors being \(\Pi_{15} = .5663\) and \(\Pi_{25} = .4337\). This discrimination point has succeeded in reducing the plausibility of model 2, and now model 1 is slightly favoured. When this design procedure is repeated for \(n = 6\), the posterior probability associated with model 1 increases substantially, as it continues to do for \(n = 7\) and \(n = 8\). These probability results are summarised in Table 2. If Wald's sequential/ratio test (1947) is performed on the posterior probabilities with Type I and Type II errors equal to .01, model 1 would be selected as the better model on the basis of these eight runs.

When data on both responses are used, however, we can discriminate in favour of model 1 after only five runs. That is, using all the data in Table 1 for the initial four runs and maximising the discrimination function \(D\), Equation...
(2.8) with \( m = 2, r = 2, \) and \( n = 5 \), we obtain \((\xi_{15}, \xi_{25}) = (125, 600)\) as the best discriminatory conditions for the fifth run which yield \((\Pi_{15}, \Pi_{25}) = (0.9950, 0.0050)\). The probability ratio test indicates termination of experimentation at this point. For this example, the rate of convergence on the correct model is greatly increased by the use of response data on \( B \) as well as \( A \).

The discrimination procedure was continued for eight runs and, although the same operating conditions were found as in the single-response case, the evidence in favour of model 1 (measured in terms of the ratio \( \Pi_{10}/\Pi_{20} \)) is overwhelmingly greater in the double-response case. These results are summarised in Table 3.

In general, the availability of high-speed computers makes the discrimination procedure described here a powerful tool for the experimenter who has multiresponse data and wishes to design experiments to distinguish among several rival models.
Table 1. Initial 4 data runs

<table>
<thead>
<tr>
<th>n</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>575</td>
<td>.3961</td>
<td>.6165</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>475</td>
<td>.7232</td>
<td>.1988</td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td>475</td>
<td>.4215</td>
<td>.5984</td>
</tr>
<tr>
<td>4</td>
<td>125</td>
<td>575</td>
<td>.1297</td>
<td>.8906</td>
</tr>
</tbody>
</table>

Table 2. Discrimination results using only response A

where $\Pi_{10} = .5$ and $\Pi_{20} = .5$

<table>
<thead>
<tr>
<th>n</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>$\Pi_{1n}$</th>
<th>$\Pi_{2n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>After initial 4 runs</td>
<td></td>
<td></td>
<td>.4614</td>
<td>.5386</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>600</td>
<td>.5663</td>
<td>.4337</td>
</tr>
<tr>
<td>6</td>
<td>125</td>
<td>600</td>
<td>.8658</td>
<td>.1342</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>450</td>
<td>.9757</td>
<td>.0243</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>600</td>
<td>.9988</td>
<td>.0012</td>
</tr>
</tbody>
</table>

Table 3. Discrimination results using both responses A and B where $\Pi_{10} = .5$ and $\Pi_{20} = .5$

<table>
<thead>
<tr>
<th>n</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>$\Pi_{1n}$</th>
<th>$\Pi_{2n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>After initial 4 runs</td>
<td></td>
<td></td>
<td>.9563</td>
<td>.0437</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>600</td>
<td>.9950</td>
<td>.0050</td>
</tr>
<tr>
<td>6</td>
<td>125</td>
<td>600</td>
<td>-1.0000</td>
<td>.1245 x 10^{-4}</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>450</td>
<td>-1.0000</td>
<td>.5200 x 10^{-8}</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>600</td>
<td>-1.0000</td>
<td>.1385 x 10^{-11}</td>
</tr>
</tbody>
</table>
REFERENCES


ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation under grant GP-2755. W. J. Hill was supported by a fellowship grant from Monsanto.