KULLBACK-LEIBLER INFORMATION FUNCTION AND THE SEQUENTIAL SELECTION OF EXPERIMENTS TO DISCRIMINATE AMONG SEVERAL LINEAR MODELS

by

Steven Michael Sidik

Submitted in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

Thesis Advisor: Shelemyahu Zacks

Department of Mathematics and Statistics

CASE WESTERN RESERVE UNIVERSITY

June 7, 1972

KULLBACK-LEIBLER INFORMATION FUNCTION AND THE SEQUENTIAL SELECTION OF EXPERIMENTS TO DISCRIMINATE AMONG SEVERAL LINEAR MODELS

Abstract

by

Steven Michael Sidik

Assume that a finite set of potential linear models relating several controlled variables to an observed variable is postulated and that exactly one of these models is the true model. The problem is to sequentially design most informative experiments so that the correct model can be determined with as little experimentation as possible. We assume that the error variance of the process is known. In addition, we assume the statistician possesses prior information which can be expressed as the prior probability that each of the proposed models is indeed the correct model and prior multivariate normal distributions on the parameters of each of the postulated model equations. After each stage of sampling, the prior distributions and the observed data values are used to compute posterior probabilities of the models being the true one and posterior distributions on the parameters of the models. Then sampling is terminated if either a prespecified number of observations has been taken or if any of the posterior probabilities of the models exceeds a prespecified minimum stopping probability. Upon termination of sampling, the model with the largest posterior

probability is chosen to be the correct model. If sampling is not to be terminated, the next experiment chosen is that one in the set of allowable values of the controlled variables which maximizes the expected Kullback-Leibler information function based upon the current posterior probabilities and distributions.

An analytical study of this procedure is too complex and difficult to adequately achieve. Hence a number of Monte-Carlo simulation experiments were performed to obtain information about the performance of this adaptive design procedure. Two basic types of Monte-Carlo experiments were performed. In the first, one of the models was chosen to be used to generate the random observations using known fixed values for the parameters. Then a large number of observations were taken using the Kullback-Leibler information functions as a criterion to choose the sequence of experiments. It was found the posterior probability of the chosen model relatively rapidly approaches the value of 1.0 and then fluctuates near 1.0. The posterior mean of the parameters of the correct model also rapidly approaches the known fixed values used to generate the observations. In the second type of experiment, one of the models was chosen to be used to generate the random observations. Then for various combinations of the maximum number of observations, stopping probability, prior distributions of the parameters, and error variance of the process, a large number of repetitions of the sequential design procedure were executed. Then a probability of correct selection and average sample number were calculated based upon the

number of times the procedure chose the correct model and the number of observations taken until termination. In general, it was found that as long as the prior mean of the correct model is not too distant from the true value with respect to the means of the other models the probability of correct selection is respectably high.

ACKNOWLEDGMENTS

The author wishes to express his gratitude for the time and encouragement given him by Prof. S. Zacks.

Special thanks are due the National Aeronautics and Space Administration Lewis Research Center for their generous support of the author's research in terms of salary, tuition, typing, and computer time.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION AND LITERATURE SURVEY

The general linear model has become one of the most useful statistical tools available to the modern scientific experimenter. There have been many books and papers written about techniques for choosing the appropriate or "best" linear model to fit to a set of data already collected. In general, these have been methods of hypothesis testing to determine which of a set of specified terms in a model equation may be dropped from the model. Much work has also been done with regard to the problem of designing best or optimal experiments to estimate the parameters of specified model equations.

In this dissertation we study a sequential adaptive experimental design procedure for a related problem. Assume that a finite set of potential linear models relating certain controlled variables to an observed variable is postulated and that exactly one of these models is correct. The problem is to sequentially design most informative experiments so that the correct model equation can be determined with as little experimentation as possible. We also assume that the error variance of the process is known. In addition, we assume that the statistician possesses prior information which can be expressed by the prior probability that each of the proposed models is indeed the correct model and prior multivariate normal distributions on the parameters of the various models. We then de-

rive an adaptive procedure for designing the successive experiments using the Kullback-Leibler information function to maximize the anticipated information for discriminating among the models. That is, after each stage of sampling, the prior distributions and the observed values are used to compute posterior probabilities of the postulated models being correct and posterior distributions on the parameters of the models. Then if sampling is not to be terminated, the next experiment chosen is that which maximizes the expected Kullback-Leibler information based on the current posterior probabilities and distributions. Sampling is terminated whenever either a prespecified number of observations is finally taken or whenever any of the posterior probabilities of the models exceeds a prespecified value. Upon termination of sampling, the model with the largest posterior probability is chosen to be the correct model.

An analytical study of this procedure is too complex and difficult to adequately achieve. Hence a number of Monte-Carlo simulation experiments were performed to obtain information about the performance of this adaptive design procedure. Two basic types of Monte-Carlo experiments were performed. In the first, one of the models was chosen to be used to generate the random observations using known fixed values for the parameters. Then a large number of observations were taken using the Kullback-Leibler information as a criterion to choose the sequence of experiments. It was found the posterior probability of the chosen model relatively rapidly approaches the value of 1.0 and then fluctuates near 1.0. The

posterior mean of the parameters of the correct model also rapidly approach the known fixed values used to generate the observations. In the second type of experiment, one of the models was chosen to be used to generate the random observations. Then for various combinations of the maximum number of observations, stopping probability, prior distributions of the parameters, and error variance of the process, a large number of repetitions of the sequential design procedure were executed. Then a probability of correct selection and average sample number were calculated based upon the number of times the procedure chose the correct model and the number of observations taken until termination. In general, it was found that as long as the prior mean of the correct model is not too distant from the true value with respect to the means of the other models the probability of correct selection is respectably high.

We now briefly indicate the general organization of the dissertation. In Chapter 2 the notation used is described and the structure of the linear models is derived. Chapter 3 then develops the distribution theory which will be basic to the remainder of the dissertation. In particular, the posterior probabilities of the models, the posterior distributions of the parameters, and the Markovian nature of the sampling process are developed. Some large sample results are then derived for the situation where the sequence of experiments is specified in advance of experimenting. These results do not thus formally apply to the adaptive design procedure.

We find, however, that they do appear to be true to a surprising extent and provide some help in explaining and interpreting the Monte-Carlo results.

In Chapter 4, the Kullback-Leibler information concept is introduced and the derivation of the anticipated information as a function of the current posterior probabilities of the models and the current posterior distributions of the parameters is presented. This anticipated information is the criterion function used to define the most informative experiment. Its use is discussed both from the point of view of its relation to the expected decrease in entropy and the point of view that it results in a very simple function measuring the amount by which the expected value of the observed variable under each model is separated.

The sequential experiment selection, stopping, and model selection rules are presented in Chapter 5.

In Chapter 6, the Monte-Carlo simulation experiments are described and the results presented and discussed. Chapter 7 presents an example of application. Several appendixes are also included. Of most importance is appendix A which presents the computer program used to perform the simulation experiments.

We now turn to a discussion of works by earlier authors who have considered similar problems.

Lindley (1956) was one of the first to consider the general idea of applying information concepts to the problems of statistical inference. He modified the concept of entropy and developed a num-

ber of interesting general results on the amount of information in an experiment about the parameters of the distribution of a random variable.

Stone (1959) was one of the first to consider information concepts as applied to designing and comparing regression experiments. He used a Bayesian framework, but the problem he considers is that of parameter estimation rather than that of model selection.

Another early and more relevant paper is that of Chernoff (1959) who applied the Kullback-Leibler information function to the sequential design of experiments when the cost of experimenting is small. His results are valid for the case of two terminal decisions and a finite number of experiments and states of nature. These results have been generalized by Albert (1961) to an infinite number of states of nature and by Bessler (1960) to an infinite number of experiments and k terminal actions. Kiefer and Sacks (1963) have also provided some extensions.

The statement of Chernoff's problem and the problem considered here are not identical and we proceed by analogizing his results to the problem at hand. In the context of the current problem, he would proceed by first assuming that at each stage of sampling the model with the largest posterior probability is the correct one. Then if A denotes the space of allowable experiments, define the Kullback-Leibler (K-L) information about model j in experiment $a \in A$ when model i is true as

$$I(a,i,j) = \int \ln\left[\frac{f_i(y|a)}{f_j(y|a)}\right] f_i(y|a)dy$$

where $f_i(y|a)$ denotes the probability density of y under model i when experiment $a\epsilon A$ is performed. Let $\hat{i}$ denote the model with the highest current probability of being the correct one. Then in analogy to Chernoff, we define the optimal experiment as $a(\hat{i})$ where $a(\hat{i})$ is defined by that experiment satisfying

$$I[a(\hat{i}),\hat{i},j] = \sup_{a\epsilon A} \inf_{k\neq\hat{i}} I(a,\hat{i},k)$$

That is, Chernoff represents the problem as a game between nature and the statistician where the statistician maximizes over A and nature minimizes over the alternative models assuming $\hat{i}$ is the correct model. Chernoff also specifically derives a stopping rule which we do not discuss here.

Hunter and Reiner (1965) considered a sequential design procedure for discriminating between two model equations. Their procedure chooses the experimental conditions which, based upon maximum likelihood estimates of the parameters from the data already collected, separate the expected values of the observed variable under the two models by as much as possible.

Box and Hill (1967) discussed the use of the Kullback-Leibler information function, deriving it from considerations involving the entropy function. They consider the use of the K-L information function to sequentially discriminate among several mechanistic (nonlinear) model equations. Besides the fact that they consider

nonlinear models, their approach is different in the sense that although they do assume prior probabilities on the proposed models, and compute posterior probabilities from the observations, they assume the parameters of the model equations are known constants.

Meeter, Pirie, and Blot (1970) have done a number of computer simulations comparing the methods of Chernoff and of Box and Hill. They found that the Box-Hill procedure performed quite well on the examples in comparison to Chernoff's procedure. It is interesting to note that Chernoff seems to be the only one of these authors who defined an explicit rule for terminating sampling. Although Chernoff's procedure is known to be asymptotically optimal, it is also known to require very large sample sizes.

CHAPTER 2

STRUCTURE OF THE LINEAR MODELS

In the theory of the general linear statistical model, we are concerned with problems involving model equations relating k controlled variables (z_i; $i = 1, \ldots, k$) to an observed variable (y). The form of the model equation is required to be

$$y = \sum_{i=1}^{I} \beta_i h_i(z_1, \ldots, z_k) + \varepsilon$$

The known functions h_i are arbitrary except that they may not contain any unknown parameters. The equation is linear in the unknown parameters β_i and ε is assumed to be a random variable with expectation zero and known finite variance. We may write $x_i = h_i(z_1, \ldots, z_k)$ and henceforth express the models in terms of the x_i variables. If n observations are made upon y we let x_{ji} denote the value of x_i at which the j^{th} observation is made. Thus for the n observations the model may conveniently be written as

$$\vec{y} = M\vec{\beta} + \vec{\varepsilon} \qquad (2\text{-}1)$$

where

$$\vec{y}' = (y_1, y_2, \ldots, y_n)$$

$$M = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1I} \\ x_{21} & x_{22} & & x_{2I} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nI} \end{bmatrix}$$

$$\vec{\beta}' = (\beta_1, \beta_2, \ldots, \beta_I)$$

$$\vec{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$$

and the ε_i are uncorrelated. The matrix M is called the design matrix for the experiment consisting of the n observations. The problem of experimental design is that of choosing the x_{ji} values in some "optimal" manner.

In certain situations in practice the experimenter can postulate several possible models involving different functions of the z_i variables which correspond to several possible mechanistic or empirically based theories. These may lead to the various models containing different sets of x_i. There may be some overlapping of the x_i among the models or there may be none.

There are then two problems requiring solution. The first is that of choosing experiment designs which will enable the experimenter to decide which of the potential models is the correct one. Then, having chosen the model, the parameters must be estimated. The second problem has many solutions using a variety of standard

techniques. This dissertation concerns itself primarily with a method of designing experiments to provide information for choosing the appropriate model equation.

We assume there are L different competing model equations. These models may be combined into one large possible model equation and then the L hypothetical models are equivalent to there being L hypotheses restricting certain sets of parameters of the large model to be a priori zero. For example, we might have two controlled variables x_1 and x_2. And suppose the model equations postulated are:

$$H_1: \quad y = \beta_1^{(1)} x_1 + \varepsilon$$

$$H_2: \quad y = \beta_2^{(2)} x_2 + \varepsilon$$

$$H_3: \quad y = \beta_1^{(3)} x_1 + \beta_2^{(3)} x_2 + \varepsilon$$

where $\beta_i^{(j)}$ denotes the coefficient of controlled variable i in model equation j. The distinction must be made because although $\beta_i^{(j)}$ and $\beta_i^{(k)}$ are coefficients of variable i, their distributions need not be the same. This notation is clumsy, however, and if we implicitly accept the fact that the distributions of the $\beta_i^{(j)}$ depend upon the model, we may more simply rewrite the models as

$$H_1: \quad y = \beta_1 x_1 + \varepsilon$$

$$H_2: \quad y = \beta_2 x_2 + \varepsilon$$

$$H_3: \quad y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

We say that models 1 and 2 are <u>nested within</u> model 3. This is

equivalent to writing one model as $y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon = \vec{X}'\vec{B} + \varepsilon$ and hypothesizing

$$H_1: \ \beta_2 = 0$$

$$H_2: \ \beta_1 = 0$$

$$H_3: \ \beta_1 \neq 0, \beta_2 \neq 0$$

In this sense it is seen that the terms model and hypothesis are interchangeable and will be used interchangeably in the remainder of this dissertation. The notation we adopt is that H_ℓ claims

$$\vec{y} = M_\ell \vec{\alpha}_\ell + \vec{\varepsilon}$$

where $\vec{\alpha}_\ell$ is the appropriate $k_\ell \times 1$ vector of β's from $\vec{B}$ which appear in model ℓ and M_ℓ is the appropriate matrix of x's.

We now precisely state the three basic distributional assumptions about the parameters and random variables of the models:

(1) The vector $\vec{\varepsilon}$ follows a multivariate normal distribution with mean $\vec{0}$ and precision matrix T. T is assumed known. (The precision matrix is the inverse of the covariance matrix of the distribution.) Since T must be positive definite symmetric, we need only consider the special case where $T = \tau I$ since linear transformation of the y reduces all other cases to this one. Note that we assume τ is known. Thus $\vec{\varepsilon} \sim N(\vec{0}, \tau I)$.

(2) For each $\ell = 1, \ldots, L$ the prior distribution of $\vec{\alpha}_\ell$ is

$$\vec{\alpha}_\ell \sim N(\vec{\mu}_{\ell,0}, \Psi_{\ell,0})$$

where $\vec{\mu}_{\ell,0}$ and $\Psi_{\ell,0}$ are known.

(3) The prior probability that the ℓ^{th} model is the correct

model equation is assumed specified and denoted by $\theta_{\ell,0}$. We require $\sum_{\ell=1}^{L} \theta_{\ell,0} = 1.0$. In order to satisfy this requirement in a completely precise manner we must make the models mutually exclusive. As described so far, this need not be true. However, this is a simple problem to get around for the following reason. Each of the H_ℓ specifies that $\vec{\alpha}_\ell$ is an element of a k_ℓ dimensional subset of K-space which we denote as E_ℓ. For any pair H_j and H_k we have either (1) $E_j \supset E_k$, (2) $E_k \supset E_j$, or (3) neither space contains the other and $E_j \cap E_k$ has measure zero with respect to H_j and H_k. For case 1 define θ_j as $\theta_j = \Pr\{\vec{\alpha}_j \varepsilon E_j - E_k\}$. But E_k has zero measure with respect to H_j and hence the distribution function of $\vec{\alpha}_j$ restricted to $E_j - E_k$ is identical to the distribution function of $\vec{\alpha}_j$ over all of E_j. Thus for any practical purpose, the fact that $E_j \supset E_k$ does not affect any probability computations. Similar arguments apply to cases two and three.

We now describe the space A of allowable experiments in more detail. If the number of elements of $\vec{X}$ is K, then a choice of experiment $a \varepsilon A$ is composed of the number J of observations to take and J vectors from some subset of Euclidean K-space. The J vectors specify the values of the controlled variables x_{ji}. At the j^{th} experiment or j^{th} stage of experimenting the particular choice from A is denoted a_j.

CHAPTER 3

PREREQUISITE DISTRIBUTION THEORY

In the remainder of this dissertation, much use will be made of the distribution of the observed variable, the posterior probabilities of the models, and the posterior distributions of the parameters of the model equations. The first part of this chapter develops these distributions. The second part derives the fact that the sampling procedure is Markovian in nature and provides a notation for describing the state of the process. The third section of this chapter discusses some results on the limiting behavior of the posterior distributions when the sequence of experiments is chosen in advance. The strong restrictions that must be made to accomplish these large sample results and the fact that they do not describe the adaptive process might lead one to believe that they are not worthwhile pursuing.

We find in chapter 6, however, that there is a close analogy between these results and the behavior of the adaptive procedure, and that these results help explain and interpret the Monte Carlo simulation results.

3.1 Posterior and Marginal Distributions

Let $f_\ell(\vec{y}_{j+1} | a_{j+1}, \vec{\alpha}_\ell)$ denote the density function of the vector $\vec{y}_{j+1}$ under H_ℓ when the parameter values are given by $\vec{\alpha}_\ell$ at stage $j+1$ of sampling. Let the probability density function

of $\vec{\alpha}_\ell$ after j stages of sampling be denoted $\xi_{\ell,j}(\vec{\alpha})$. This is a preposterior density since it serves as the posterior density of $\vec{\alpha}_\ell$ after j stages of sampling and the prior density of $\vec{\alpha}_\ell$ before the $j+1^{st}$ stage of sampling occurs.

Lemma 3.1: After j stages of sampling, $\vec{\alpha}_\ell$ follows a multivariate normal distribution with mean vector $\vec{\mu}_{\ell,j}$ and precision matrix $\Psi_{\ell,j}$. That is, after j stages of sampling,

$$\vec{\alpha}_\ell \sim N(\vec{\mu}_{\ell,j}, \Psi_{\ell,j})$$

where

$$\Psi_{\ell,j} = \Psi_{\ell,j-1} + M'_{\ell,j} T M_{\ell,j}$$

$$= \Psi_{\ell,0} + \sum_{i=1}^{j} M'_{\ell,i} T M_{\ell,i} \qquad (3\text{-}1)$$

and

$$\vec{\mu}_{\ell,j} = \Psi_{\ell,j}^{-1}(M'_{\ell,j} T \vec{y}_j + \Psi_{\ell,j-1}\vec{\mu}_{\ell,j-1})$$

$$= \Psi_{\ell,j}^{-1}\left[\sum_{i=1}^{j} M'_{\ell,i} T \vec{y}_i + \Psi_{\ell,0}\vec{\mu}_{\ell,0}\right] \qquad (3\text{-}2)$$

and where $M_{\ell,i}$ denotes the design matrix specified by a_i under H_ℓ.

Proof: By Bayes theorem, if $\vec{y}_j$ is the observed vector at stage j

$$\xi_{\ell,j}(\vec{\alpha}) = \frac{f_\ell(\vec{y}_j | a_j, \vec{\alpha}) \xi_{\ell,j-1}(\vec{\alpha})}{\int f_\ell(\vec{y}_j | a_j, \vec{\alpha}^*) \xi_{\ell,j-1}(\vec{\alpha}^*) d\vec{\alpha}^*} .$$

$$\propto f_\ell(y_j | a_j, \vec{\alpha}) \xi_{\ell,j-1}(\vec{\alpha}) \qquad (3\text{-}3)$$

The symbol $\propto$ means "proportional to" and is used in the context of DeGroots (1970, p. 160) usage. Thus

$$\xi_{\ell,j}(\vec{\alpha}) \propto e^{-Q/2}$$

where (dropping subscripts)

$$Q = (M\vec{\alpha} - \vec{y})'T(M\vec{\alpha} - \vec{y}) + (\vec{\alpha} - \vec{\mu})'\Psi(\vec{\alpha} - \vec{\mu})$$

Since T and Ψ are positive definite symmetric we can write

$$Q = \vec{\alpha}'(\Psi + M'TM)\vec{\alpha} - 2(\vec{\mu}'\Psi + \vec{y}'TM)\vec{\alpha} + (\text{terms not involving } \vec{\alpha})$$

$$= [\vec{\alpha} - (\Psi + M'TM)^{-1}(M'T\vec{y} + \Psi\vec{\mu})]' \cdot (\Psi + M'TM)$$

$$\cdot [\vec{\alpha} - (\Psi + M'TM)^{-1}(M'T\vec{y} + \Psi\vec{\mu})] + (\text{terms not involving } \vec{\alpha})$$

The terms not involving $\vec{\alpha}$ may be factored out through use of the proportionality device, leaving the kernel of a multivariate normal distribution with parameters as specified by the lemma. Thus $\vec{\alpha}_\ell$ is distributed as claimed.

Q.E.D.

Owen (1970) has derived a result similar to Lemma 3.1 in the case of a two factor experiment where the factors are treatments and blocks.

We now turn to determining the distribution of $\vec{y}_{j+1}$. This is done in two stages. First we do not know which of the models is in fact the correct one. Then for any given model, we do not know the value of $\vec{\alpha}_\ell$. Let $f_\ell(\vec{y}_{j+1}|a_{j+1},\vec{\alpha})$ denote the distribution of $\vec{y}_{j+1}$ under H_ℓ when experiment $a_{j+1} \epsilon A$ is performed and $\vec{\alpha}_\ell$ is specified. Since we do not know $\vec{\alpha}_\ell$ we must average this distribution over all $\vec{\alpha}_\ell$. Let $f_\ell(\vec{y}_{j+1}|a_{j+1})$ denote the mixture of the densities $f_\ell(\vec{y}_{j+1}|a_{j+1},\vec{\alpha})$ with respect to the marginal posterior of $\vec{\alpha}_\ell$.

Lemma 3.2 The conditional distribution of $\vec{y}$ given H_ℓ and a_j is a multivariate normal distribution with mean vector $\vec{s}_{\ell,j}$ and precision matrix $R_{\ell,j}$ where

$$R_{\ell,j} = T\left[I - M_{\ell,j}(M'_{\ell,j}TM_{\ell,j} + \Psi_{\ell,j-1})^{-1}M'_{\ell,j}T\right] \tag{3-4}$$

$$\vec{s}_{\ell,j} = R^{-1}_{\ell,j}TM_{\ell,j}(M'_{\ell,j}TM_{\ell,j} + \Psi_{\ell,j-1})^{-1}\Psi_{\ell,j-1}\vec{\mu}_{\ell,j-1} \tag{3-5}$$

Proof: The required mixture distribution is given by

$$f_\ell(\vec{y}_j|a_j) = \int f_\ell(\vec{y}_j|a_j,\vec{\alpha})\xi_{\ell,j-1}(\vec{\alpha})d\vec{\alpha}$$

$$\propto \int e^{-Q/2}\, d\vec{\alpha}$$

where

$$Q = (\vec{y}_j - M_{\ell,j}\vec{\alpha})'T(\vec{y}_j - M_{\ell,j}\vec{\alpha}) + (\vec{\alpha} - \vec{\mu}_{\ell,j-1})'\Psi_{\ell,j-1}(\vec{\alpha} - \vec{\mu}_{\ell,j-1})$$

$$= \vec{\alpha}'(M'TM + \Psi)\vec{\alpha} - 2\vec{\alpha}'(M'TM + \Psi)(M'TM + \Psi)^{-1}(M'T\vec{y} + \Psi\vec{\mu})$$
$$+ \vec{y}'T\vec{y} + \vec{\mu}'\Psi\vec{\mu}$$

$$= \vec{\alpha}'(M'TM + \Psi)\vec{\alpha} - 2\vec{\alpha}'(M'TM + \Psi)(M'TM + \Psi)^{-1}(M'T\vec{y} + \Psi\vec{\mu})$$
$$+ (M'T\vec{y} + \Psi\vec{\mu})'(M'TM + \Psi)^{-1}(M'Ty + \Psi\vec{\mu})$$
$$- (M'T\vec{y} + \Psi\vec{\mu})'(M'TM + \Psi)^{-1}(M'T\vec{y} + \Psi\vec{\mu}) + \vec{y}'T\vec{y} + \vec{\mu}'\Psi\vec{\mu}$$

The first three terms yield the quadratic form

$$Q_1 = [\vec{\alpha} - (M'TM + \Psi)^{-1}(M'T\vec{y} + \Psi\vec{\mu})]' \cdot (M'TM + \Psi)$$
$$\cdot [\vec{\alpha} - (M'TM + \Psi)^{-1}(M'T\vec{y} + \Psi\vec{\mu})]$$

The remainder of Q does not involve $\vec{\alpha}$ and $e^{-Q_1/2}$ is the kernel of a multivariate normal distribution so that when $e^{-Q_1/2}$

is integrated over $\vec{\alpha}$ we remain with

$$f_{\ell,j-1}(\vec{y}_j|a_j) \propto e^{-Q_2/2}$$

where

$$Q_2 = -(M'T\vec{y} + \Psi\vec{\mu})'(M'TM + \Psi)^{-1}(M'T\vec{y} + \Psi\vec{\mu}) + \vec{y}'T\vec{y} + \vec{\mu}'\Psi\vec{\mu}$$

$$= y'[T - TM(M'TM + \Psi)^{-1}M'T]\vec{y} - 2\vec{y}TM(M'TM + \Psi)^{-1}\Psi\vec{\mu}$$

$$+ (\text{terms not involving } \vec{y})$$

$$= (\vec{y} - \vec{s})'R(\vec{y} - \vec{s}) + (\text{terms not involving } \vec{y})$$

The terms not involving y may be factored out via the proportionality device leaving

$$f_\ell(\vec{y}_j|a_j) \propto e^{-(\vec{y}-\vec{s}_{\ell,j})'R_{\ell,j}(\vec{y}-\vec{s}_{\ell,j})/2}$$

This is the kernel of a multivariate normal distribution with mean vector $\vec{s}_{\ell,j}$ and precision matrix $R_{\ell,j}$ as claimed. Thus the density of y_j given H_ℓ and a_j is given by

$$f_\ell(\vec{y}_j|a_j) = (2\pi)^{-J/2}|R_{\ell,j}|^{1/2}\exp\left\{-\frac{1}{2}(\vec{y}_j - \vec{s}_{\ell,j})'R_{\ell,j}(\vec{y}_j - \vec{s}_{\ell,j})\right\} \tag{3-6}$$

Q.E.D.

Since the true model is unknown we now compute the mixture of the distributions of Lemma 3.2 with respect to the probabilities $\theta_{\ell,j}$ as

$$f(y_j|a_j) = \sum_{\ell=1}^{L} \theta_{\ell,j-1} f_\ell(\vec{y}_j\ a_j) \tag{3-7}$$

To compute the posterior probability of each model being cor-

rect after the observation y_{j+1} is obtained, we apply Bayes theorem directly to get

$$\theta_{\ell,j+1} = \frac{f_\ell(\vec{y}_{j+1}|a_{j+1})\theta_{\ell,j}}{\sum_{k=1}^{L} f_k(\vec{y}_{j+1}|a_{j+1})\theta_{k,j}} \tag{3-8}$$

3.2 Markovian Nature of Sampling Process

Consider a sequence of random variables $W_1, W_2, \ldots$ which take on values in a sample space or state space Ω. We let $\mathcal{F}$ denote the σ-field of subsets of Ω for which probabilities are defined. The sequence of random variables W_i form a Markov Process if for every $F \epsilon \mathcal{F}$ and for all $w_1, \ldots, w_n$ in Ω, and all for n, $n = 1,2,3, \ldots$ we have

$$\begin{aligned} \Pr\{W_{n+1} \epsilon F | W_1 = w_1, \ldots, W_n = w_n\} &= \Pr\{W_{n+1} \epsilon F | W_n = w_n\} \\ &= \int_F g_{n+1}(w|w_n)\,dw \end{aligned} \tag{3-9}$$

where $g_{n+1}(w|w_n)$ denotes the generalized conditional probability density function of W_{n+1}. If the conditional probabilities in equation (3-9), equivalently the g_n, do not depend upon n the transition process is called stationary. The state space in this paper can be described by a vector containing: (1) the probabilities θ_ℓ, (2) the elements of the vectors describing the current posterior means under the various H_ℓ, and (3) the lower triangular part of the current posterior precision matrices under the various H_ℓ. Thus

$$\Omega = \Big\{\theta_1, \ldots,\theta_L,\mu_1^{(1)},\mu_1^{(2)}, \ldots,\mu_2^{(1)},\mu_2^{(2)}, \ldots,\mu_L^{(1)}, \ldots,$$
$$\mu_L^{(K_L)},\psi_1^{(1,1)}, \ldots,\psi_L^{(K_L,K_L)} : \sum_{i=1}^{L} \theta_i = 1.0,\ -\infty < \mu_k^{(j)} < +\infty,$$
$$\psi_i \text{ positive definite for } i = 1, \ldots,L\Big\} \qquad (3\text{-}10)$$

For any given state $w\varepsilon\Omega$ the transition to the next state depends only upon the state w and the experiment $a\varepsilon A$ that is chosen.

This is true because a determines the posterior precision matrices regardless of the value of y, and the posterior means $\vec{\mu}_\ell(y)$ and probabilities $\vec{\theta}(y)$ are determined by equations (3-1) and (3-8) which again depend only upon w, y, and a. Thus the transition process on the states is Markovian. The process is stationary, also, since for given initial g_i the successive g_n do not depend upon n. The transition function may be described as follows. Define a mapping $T:\Omega\times Y \to \Omega$ and let $Q(\Omega\times Y)$ denote the Borel sets on $\Omega\times Y$ and $Q(\Omega)$ denote the Borel sets on Ω. Let $T^{-1}(F)$ denote the inverse image of F where $F\varepsilon Q(\Omega)$ and $T^{-1}(F)\varepsilon Q(\Omega\times Y)$. Then if w' denotes the state of the system after sampling,

$$\Pr(w'\varepsilon F|w,a) = \int_{(w,y)\varepsilon T^{-1}(F)} \sum_{i=1}^{L} \theta_i f_i(y|a,w)dy$$

3.3 Large Sample and Limiting Results

Even though this paper is concerned primarily with small sample procedures, it is interesting and informative to know the large sample or limiting behavior of the parameters and the sampling

process. Unfortunately, for the adaptive procedure this is an extremely difficult subject to study. Thus we do not study the adaptive procedure here but instead consider the experiment selection procedure under the restrictions listed below in the hope that these results will illuminate the adaptive procedure in some sense.

(1) Assume A is finite with $N(A)$ elements, and represented as

$$A = \{a^{(1)}, a^{(2)}, \ldots, a^{[N(A)]}\}$$

(2) An infinite sequence $\{a_j\}$ is specified such that as the number of experiments approaches infinity, the proportion of times that $a^{(i)}$ is performed approaches p_i with $0 < p_i < 1$ and $\Sigma p_i = 1.0$. The experiments a_j are chosen independently of each other.

(3) Assume H_{i*} is the true model and that $\vec{\mu}^*$ is the true value of the parameters in the model.

(4) Assume that only one observation is taken in each experiment $a^{(i)}$.

(5) Assume that the structure of A is such that all matrices under consideration are nonsingular.

It should be noted that the most restrictive of the above assumptions is the second. For in a true sequential decision procedure, the actual experiment chosen is a random variable depending upon the previous observations obtained. Since we are in fact studying a problem other than the one of most importance the remainder of the chapter will not be developed in rigorous detail and the results

obtained cannot be rigorously applied to the sequential procedure. It will be seen in Chapter 6, however, that fairly extensive Monte Carlo simulations seem to bear up the general conclusions reached here.

Let $k(j)$ denote the superscript of the experiment performed at stage j. Thus if $a_{10} = a^{(5)}$, then $k(10) = 5$. Also let $n(i,j)$ denote the number of times $a^{(i)}$ is performed in the sequence of experiments up to and including the j^{th} stage. Let $M_{\ell,i}$ denote the design matrix under H_ℓ when $a^{(i)}$ is chosen.

Lemma 3.3 Under the above assumptions the posterior precision matrices and mean vectors converge with probability one as $j \to \infty$ to:

$$\frac{1}{j\tau} \Psi_{\ell,j} \to \Psi_\ell = \sum_{i=1}^{N(A)} p_i M'_{\ell,i} M_{\ell,i}$$

$$\vec{\mu}_{\ell,j} \to (\Psi_\ell)^{-1} \left(\sum_{i=1}^{N(A)} p_i M'_{\ell,i} M_{i^*,i} \right) \vec{\mu}^*$$

Proof: To prove the first limit, recall from equation (3-1) that

$$\Psi_{\ell,j} = \Psi_{\ell,0} + \sum_{i=1}^{j} \tau M'_{\ell,k(i)} M_{\ell,k(i)}$$

Thus

$$\frac{1}{j\tau}\Psi_{\ell,j} = \frac{1}{j\tau}\Psi_{\ell,0} + \sum_{i=1}^{j}\frac{1}{j}M'_{\ell,k(i)}M_{\ell,k(i)}$$

$$= \frac{1}{j\tau}\Psi_{\ell,0} + \sum_{i=1}^{N(A)}\frac{n(i,j)}{j}M'_{\ell,i}M_{\ell,i}$$

As $j \to \infty$, the first term goes to zero and the factors $\frac{n(i,j)}{j} \to p_i$ by assumption.

To prove the second part note that $y_j = M_{i^*,k(j)}\vec{\mu}^* + \varepsilon_j$ where $\varepsilon_j \sim N(0,\tau)$. Using the second form of equation (3-2) we get

$$\vec{\mu}_{\ell,j} = (\Psi_{\ell,j})^{-1}\left\{\tau\sum_{i=1}^{j}M'_{\ell,k(i)}y_i + \Psi_{\ell,0}\vec{\mu}_{\ell,0}\right\} \qquad (3\text{-}11)$$

Then substituting the expression for y_j into equation (3-11) gives

$$\vec{\mu}_{\ell,j} = (\Psi_{\ell,j})^{-1}\left\{\sum_{i=1}^{N(A)} n(i,j)\tau M'_{\ell,i}M_{i^*,i}\vec{\mu}^* + \tau\sum_{i=1}^{N(A)}M'_{\ell,i}\left(\sum_{m=1}^{j}\varepsilon_m\delta_{k(m),i}\right) + \Psi_{\ell,0}\vec{\mu}_{\ell,0}\right\}$$

where $\delta_{i,j}$ denotes the Kronecker delta function. Thus

$$\vec{\mu}_{\ell,j} = \left(\frac{1}{j\tau}\Psi_{\ell,j}\right)^{-1}\left\{\sum_{i=1}^{N(A)}\frac{n(i,j)}{j}M'_{\ell,i}M_{i^*,i}\vec{\mu}^* + \sum_{i=1}^{N(A)}M'_{\ell,i}\left(\frac{1}{j}\sum_{m=1}^{j}\varepsilon_m\delta_{k(m),i}\right) + \frac{1}{j\tau}\Psi_{\ell,0}\vec{\mu}_{\ell,0}\right\}$$

From assumption 2 we know that $n(i,j) \to \infty$ as $j \to \infty$ and since the

ε_m form a sequence of independent and identically distributed random variables, the strong law of large numbers may be applied to show for $i = 1, \ldots, L$

$$\Pr\left\{\lim_{n(i,j)\to\infty} \frac{1}{n(i,j)} \sum_{m=1}^{j} \varepsilon_m \delta_{k(m),i} = 0\right\} = 1.0$$

Since $\frac{1}{j\tau} \to 0$ as $j \to \infty$ we then have

$$\vec{\mu}_{\ell,j} \simeq \Psi_\ell^{-1}\left(\sum_{i=1}^{N(A)} \frac{n(i,j)}{j} M'_{\ell,i} M_{i*,i}\right)\vec{\mu}^*$$

This sequence will not have a limit unless $\lim_{j\to\infty} \frac{n(i,j)}{j} = p_i$ exists. If such a limit exists, the lemma follows immediately.

Q.E.D.

Lemma 3.4 Under the assumptions stated, $R_{\ell,j} \to \tau$ irrespective of $\{a_j\}$ and

$$s_{\ell,j} \simeq M_{\ell,k(j)} (\Psi_\ell)^{-1}\left(\sum_{i=1}^{N(A)} p_i M'_{\ell,i} M_{i*,i}\right)\vec{\mu}^*$$

for large enough j.

Proof: From equation (3-4) and the assumptions

$$R_{\ell,j} = \tau\left\{1.0 - \tau M_{\ell,k(j)}[\Psi_{\ell,j} + \tau M'_{\ell,k(j)} M_{\ell,k(j)}]^{-1} M'_{\ell,k(j)}\right\}$$

As $j \to \infty$, $[\Psi_{\ell,j} + \tau M'_{\ell,k(j)} M_{\ell,k(j)}] \to (\infty)$ and hence its inverse $\to (0)$. But then $R_{\ell,j} \to \tau$ as claimed.

From equation (3-5)

$$s_{\ell,j} = (R_{\ell,j})^{-1}\left\{\tau M_{\ell,k(j)}[M'_{\ell,k(j)} M_{\ell,k(j)} \tau + \Psi_{\ell,j}]^{-1} \Psi_{\ell,j} \vec{\mu}_{\ell,j}\right\}$$

For large enough j, $[M'_{\ell,k(j)}M_{\ell,k(j)}\tau + \Psi_{\ell,j}]^{-1}\Psi_{\ell,j}$ is asymptotically like the identity matrix, I, so that

$$s_{\ell,j} \approx \frac{1}{\tau}\left\{\tau M_{\ell,k(j)}\vec{\mu}_{\ell,j}\right\}$$

$$\approx M_{\ell,k(j)}(\Psi_\ell)^{-1}\left[\sum_{i=1}^{N(A)} p_i M'_{\ell,i}M_{i^*,i}\right]\vec{\mu}^*$$

We note that if $\ell = i^*$, then from the definition of Ψ_ℓ, we have $s_{i^*,j} \approx M_{i^*,k(j)}\vec{\mu}^*$ as expected.

Q.E.D.

Lemma 3.5 If H_{i^*} is the true hypothesis and the model of H_{i^*} is nested within the model of H_ℓ, then under the above assumptions and assuming the parameter vectors are rearranged appropriately

$$\Pr\left\{\vec{\mu}_{\ell,j} \to \begin{pmatrix}\vec{\mu}^* \\ \vec{0}\end{pmatrix}\right\} = 1.0$$

Proof: From Lemma 3.3 we have

$$\vec{\mu}_{\ell,j} \to \Psi_\ell^{-1}\left(\sum_{i=1}^{N(A)} p_i M'_{\ell,i}M_{i^*,i}\right)\vec{\mu}^*$$

If the parameters are rearranged appropriately then $M_{\ell,i}$ may be written

$$M_{\ell,i} = (M_{i^*,i}, \tilde{M}_{\ell,i})$$

where $\tilde{M}_{\ell,i}$ denotes the design matrix corresponding to the independent variables in H_ℓ but not in H_{i^*}. Thus

$$M'_{\ell,i}M_{\ell,i} = \begin{bmatrix} M'_{i^*,i}M_{i^*,i} & M'_{i^*,i}\tilde{M}_{\ell,i} \\ \tilde{M}'_{\ell,i}M_{i^*,i} & \tilde{M}'_{\ell,i}\tilde{M}_{\ell,i}\end{bmatrix}$$

and

$$\Psi_{\ell} = \begin{bmatrix} \Psi_{i*} & \cdot & \Psi_{i*,\ell} \\ \Psi'_{i*,\ell} & & \tilde{\Psi}_{\ell} \end{bmatrix}$$

where

$$\Psi_{i*} = \sum_{i=1}^{N(A)} p_i M'_{i*,i} M_{i*,i}$$

$$\Psi_{i*,\ell} = \sum_{i=1}^{N(A)} p_i M'_{i*,i} \tilde{M}_{\ell,i}$$

and

$$\tilde{\Psi}_{\ell} = \sum_{i=1}^{N(A)} p_i \tilde{M}_{\ell,i} \tilde{M}_{\ell,i}$$

Thus from a well known identity (e.g., Graybill (1969), p. 165)

$$\Psi_{\ell}^{-1} = \begin{bmatrix} \left(\Psi_{i*} - \Psi_{i*,\ell}\tilde{\Psi}_{\ell}^{-1}\Psi_{i*,\ell}\right)^{-1} & -\Psi_{i*}^{-1}\Psi_{i*,\ell}\left(\tilde{\Psi}_{\ell} - \Psi'_{i*,\ell}\Psi_{i*}^{-1}\Psi_{i*,\ell}\right) \\ -\left(\tilde{\Psi}_{\ell} - \Psi'_{i*,\ell}\Psi_{i*}^{-1}\Psi_{i*,\ell}\right)^{-1}\Psi'_{i*,\ell}\Psi_{i*}^{-1} & \left(\tilde{\Psi}_{\ell} - \Psi'_{i*,\ell}\Psi_{i*}^{-1}\Psi_{i*,\ell}\right) \end{bmatrix}$$

Also

$$\sum_{i=1}^{N(A)} p_i M'_{\ell,i} M_{i*,i} = \sum_{i=1}^{N(A)} \begin{pmatrix} p_i M'_{i*,i} M_{i*,i} \\ p_i \tilde{M}'_{\ell,i} M_{i*,i} \end{pmatrix} = \begin{pmatrix} \Psi_{i*} \\ \Psi'_{i*,\ell} \end{pmatrix}$$

Thus

$$\Psi_{\ell}^{-1}\begin{pmatrix}\Psi_{i^*}\\ \Psi'_{i^*,\ell}\end{pmatrix} = \begin{bmatrix}\left(\Psi_{i^*} - \Psi_{i^*,\ell}\tilde{\Psi}_{\ell}^{-1}\Psi'_{i^*,\ell}\right)^{-1}\Psi_{i^*} \\ \qquad - \Psi_{i^*}^{-1}\Psi_{i^*,\ell}\left(\tilde{\Psi}_{\ell} - \Psi'_{i^*,\ell}\Psi_{i^*}^{-1}\Psi_{i^*,\ell}\right)^{-1}\Psi'_{i^*,\ell} \\ -\left(\tilde{\Psi}_{\ell} - \Psi'_{i^*,\ell}\Psi_{i^*}^{-1}\Psi_{i^*,\ell}\right)^{-1}\Psi_{i^*,\ell} \\ \qquad + \left(\tilde{\Psi}_{\ell} - \Psi'_{i^*,\ell}\Psi_{i^*}^{-1}\Psi_{i^*,\ell}\right)^{-1}\Psi_{i^*,\ell}\end{bmatrix}$$

$$= \begin{bmatrix}I\\ 0\end{bmatrix}$$

upon application of Lemma 3.6 which follows. Thus

$$\vec{\mu}_{\ell,j} \to \begin{pmatrix}I\\ 0\end{pmatrix}\vec{\mu}^* = \begin{pmatrix}\vec{\mu}^*\\ \vec{0}\end{pmatrix}$$

Q.E.D.

Lemma 3.6. (Problem 2.9 of Rao (1965)) If A and D are matrices possessing inverses, then

$$(A + BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1}$$

Proof: By direct multiplication we only need show

$$I = (A + BDB')(A^{-1} - A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1}$$

$$= I - B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1} + BDB'A^{-1}$$

$$\qquad - BDB'A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1}$$

$$= I - B[-(B'A^{-1}B + D^{-1})^{-1} + D - DB'A^{-1}B(B'A^{-1}B + D^{-1})^{-1}]B'A^{-1}$$

$$= I - B[D - [I + DB'A^{-1}B][B'A^{-1}B + D^{-1}]^{-1}]B'A^{-1}$$

$$= I - B[D - D[D^{-1} + B'A^{-1}B][D^{-1} + B'A^{-1}B]^{-1}]B'A^{-1}$$

$$= I - B[D - D]B'A^{-1}$$

$$= I$$

Q.E.D.

To apply this result to Lemma 3.5 simply set

$$A = \Psi_{i*}$$

$$B = \Psi_{i*,\ell}$$

$$D = -\tilde{\Psi}_{\ell}^{-1}$$

We now turn to consideration of the limiting behavior of $\theta_{\ell,j}$.

Computer simulations for both nested and non-nested cases indicate that for any k where H_{i*} is not nested in H_k, $\theta_{k,j} \to 0.0$ fairly rapidly and steadily. If H_{i*} is nested in H_k then it seems that $\theta_{k,j} \to 0.0$. The rate is initially rapid but then becomes very slow and it behaves in a very erratic manner. These points are discussed in some detail in Chapter 6.

It should be reiterated and these discussions have assumed the sequence $\{a_j\}$ to be specified and fixed for the sequence of experiments. In a sequential decision problem the sequence $\{a_j\}$ is not fixed, but $k(j)$ is in fact a random variable whose distribution depends upon $k(i)$ for $i < j$ and the y_i for $i < j$.

CHAPTER 4

ENTROPY FUNCTIONS AND THE KULLBACK-LEIBLER INFORMATION FUNCTION

When comparing a number of experiments to determine which is the optimal one to perform, one must define optimal. In this dissertation, that experiment which yields the largest expected K-L information is defined as the optimal experiment. In particular, let $I(w,a)$ denote the expected K-L information as a function of the experiment a and the current state w of the process. This function will be specified explicitly later. In this chapter, we first describe how the K-L information arises from attempting to reduce the entropy of the probabilities of the models. We then develop an expression for $I(w,a)$ and finally discuss the operational meaning of the use of $I(w,a)$ from a heuristic point of view.

4.1 Development of the K-L Information Function

The problem under consideration here is that we must choose one of a set of postulated model equations. For each model we have the posterior probability $\theta_{\ell,j}$ that it is the correct one. We would like to choose experiments which cause the posterior probability of the correct model to increase most rapidly. An indirect method of accomplishing this is to choose experiments which most rapidly decrease the entropy of the set of probabilities $\theta_{\ell,j}$. The entropy is defined as

$$\mathcal{E}(w) = -\sum_{\ell=1}^{L} \theta_{\ell,j} \ln(\theta_{\ell,j})$$

It can be verified that the entropy attains a maximum when all the probabilities are equal and attains a minimum when any one of the probabilities is one and the rest are zero.

Box and Hill (1967) proposed the use of the expected decrease between the entropy at the current stage of sampling and the anticipated entropy at the next stage of sampling as the criterion for selection of experiments. They found, however, that the entropy function is quite intractable analytically and applied a well-known inequality to show the expected K-L information function provides an upper bound on the reduction of entropy. Let $\theta_i(\vec{y}|w,a)$ denote the posterior probability of model i if the value $\vec{y}$ is observed when the state was w. Let $w(\vec{y})$ denote the state of the process after observing the value $\vec{y}$ when it was in state w. Then the anticipated entropy is given by

$$E\left\{\mathcal{E}[w(\vec{y}),a]\right\} = -\int \left\{\sum_{\ell=1}^{L} \theta_\ell(\vec{y}|w,a) \ln[\theta_\ell(\vec{y}|w,a)]\right\} f(\vec{y}|w,a)\, d\vec{y}$$

Thus if the current state of the sampling process is $w\varepsilon\Omega$, and the experiment $a\varepsilon A$ is performed, the expected decrease in entropy, $R(w,a)$, is then defined as

$$R(w,a) = \mathcal{C}(w) - E\{\mathcal{C}[w(\vec{y}),a]\}$$

$$= -\sum_{i=1}^{L}\theta_i \ln(\theta_i) + \int\left\{\sum_{i=1}^{L}\theta_i(\vec{y}|w,a)\ln[\theta_i(\vec{y}|w,a)]\right\}\cdot$$

$$\left\{\sum_{k=1}^{L}\theta_k f_k(\vec{y}|w,a)\right\}d\vec{y}$$

$$= -\sum_{i=1}^{L}\theta_i \ln(\theta_i) + \int\sum_{\ell=1}^{L}\theta_\ell f_\ell(\vec{y}|w,a)\ln\left[\frac{\theta_\ell f_\ell(\vec{y}|w,a)}{\sum_{k=1}^{L}\theta_k f_k(\vec{y}|w,a)}\right]d\vec{y}$$

$$\leq \int\sum_{\ell=1}^{L}\theta_\ell\left\{\sum_{i=1}^{L}\theta_i f_\ell(\vec{y}|w,a)\ln\left[\frac{f_\ell(\vec{y}|w,a)}{f_i(\vec{y}|w,a)}\right]\right\}d\vec{y} \qquad (4\text{-}1)$$

by application of the following inequality (Kullback (1969), p. 15)

$$\sum_{i=1}^{L}\theta_i f_\ell(\vec{y}|w,a)\ln\left[\frac{f_\ell(\vec{y}|w,a)}{f_i(\vec{y}|w,a)}\right] \geq f_\ell(\vec{y}|w,a)\ln\left[\frac{f_\ell(\vec{y}|w,a)}{\sum_{k=1}^{L}\theta_k f_k(\vec{y}|w,a)}\right]$$

Let

$$I(w,a,i,j) = \int f_i(\vec{y}|w,a)\ln\left[\frac{f_i(\vec{y}|w,a)}{f_j(\vec{y}|w,a)}\right]d\vec{y} \qquad (4\text{-}2)$$

We note $I(w,a,i,j)$ is defined as the expected amount of information in the observations from experiment a for discriminating against H_j in favor of H_i. Let $\mathcal{I}(w,a)$ denote the matrix whose i,j element is $I(w,a,i,j)$. Then the inequality (4-1) may be written as

$$R(w,a) \leq \vec{\theta}' \Omega(w,a) \vec{\theta} = I(w,a) \tag{4-3}$$

Meeter et al. (1970) proposed the following heuristic argument in favor of using $I(w,a)$. If one knew that H_i were indeed the correct hypothesis and wished to maximize the information about H_k for $k \neq i$, then it would be natural to maximize

$$\sum_{k \neq i} \theta_k I(w,a,i,k)$$

But since H_i is assumed correct only with probability θ_i, it is equally natural to multiply the foregoing expression by θ_i and sum over i. But in doing this, one does end up with $I(w,a)$.

4.2 Evaluation of K-L Information Function

From equation (3-6) we have (if $\vec{y}$ is $J \times 1$) that the density of $\vec{y}$ under H_ℓ is given by

$$f_\ell(\vec{y}|a) = (2\pi)^{-J/2} |R_\ell|^{1/2} e^{-1/2(\vec{y}-\vec{s}_\ell)' R_\ell (\vec{y}-\vec{s}_\ell)}$$

Hence

$$\frac{f_m(\vec{y}|a)}{f_n(\vec{y}|a)} = |R_m|^{1/2} |R_n|^{-1/2} \frac{e^{-1/2(\vec{y}-\vec{s}_m)' R_m (\vec{y}-\vec{s}_m)}}{e^{-1/2(\vec{y}-\vec{s}_n)' R_n (\vec{y}-\vec{s}_n)}}$$

Moreover

$$\begin{aligned} \ln\left[\frac{f_m(\vec{y}|a)}{f_n(\vec{y}|a)}\right] = \frac{1}{2} (\ln|R_m| - \ln|R_n|) \\ - \frac{1}{2} (\vec{y} - \vec{s}_m)' R_m (\vec{y} - \vec{s}_m) \\ + \frac{1}{2} (\vec{y} - \vec{s}_n)' R_n (\vec{y} - \vec{s}_n) \end{aligned} \tag{4-4}$$

and

$$I(w,a,m,n) = \int \ln\left[\frac{f_m(\vec{y}|a)}{f_n(\vec{y}|a)}\right] f_m(\vec{y}|a)\,d\vec{y}$$

$$= E\left\{\ln\left[\frac{f_m(\vec{y}|a)}{f_n(\vec{y}|a)}\right]\right\} \qquad (4\text{-}5)$$

where the expectation is taken under the assumption $\vec{y} \sim N(\vec{s}_m, R_m)$.

Note that $I(w,a,m,m) = 0.0$ for $m = 1, \ldots, L$.

Lemma 4.1 If $\vec{y} \sim N(\vec{c},R)$ and R is positive definite, and A is symmetric, then

$$E\{\vec{y}'A\vec{y}\} = tr(AR^{-1}) + \vec{c}'A\vec{c}$$

Proof By theorem 10.3.2 of Graybill (1969)

$$E\{(\vec{y} - \vec{c})'A(\vec{y} - \vec{c})\}$$

$$= \frac{|R|^{1/2}}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} (\vec{y} - \vec{c})'A(\vec{y} - \vec{c}) e^{-1/2(\vec{y}-\vec{c})'R(\vec{y}-\vec{c})}\, d\vec{y}$$

$$= tr(AR^{-1})$$

But

$$E\{(\vec{y} - \vec{c})'A(\vec{y} - \vec{c})\} = E\{\vec{y}'A\vec{y}\} - \vec{c}'A\vec{c}$$

The lemma follows immediately.

Q.E.D.

Applying the lemma to the expectations of the quadratic forms in equation (4-4) we see:

1. $\vec{y} \sim N(\vec{s}_m, R_m) \Rightarrow E\{(\vec{y} - \vec{s}_m)'R_m(\vec{y} - \vec{s}_m)\} = tr(R_m R_m^{-1}) = J$

2. $\vec{y} - \vec{s}_n \sim N(\vec{s}_m - \vec{s}_n, R_m) \Rightarrow E\{(\vec{y} - \vec{s}_n)'R_n(\vec{y} - \vec{s}_n)\}$

$$= tr(R_n R_m^{-1}) + (\vec{s}_m - \vec{s}_n)'R_n(\vec{s}_m - \vec{s}_n)$$

Thus

$$I(w,a,m,n) = \frac{1}{2}\left\{\ln\left|R_m\right| - \ln\left|R_n\right|\right\} - \frac{1}{2}J + \frac{1}{2}\,\mathrm{tr}(R_n R_m^{-1}) + \frac{1}{2}(\vec{s}_m - \vec{s}_n)' R_n (\vec{s}_m - \vec{s}_n) \qquad (4\text{-}6)$$

$$I(w,a,m,n) + I(w,a,n,m) = -J + \frac{1}{2}\left[\mathrm{tr}\left(R_n R_m^{-1}\right) + \mathrm{tr}\left(R_m R_n^{-1}\right)\right] + \frac{1}{2}\left[(\vec{s}_m - \vec{s}_n)' R_n (\vec{s}_m - \vec{s}_n) + (\vec{s}_n - \vec{s}_m)' R_m (\vec{s}_n - \vec{s}_m)\right]$$

$$= -J + \frac{1}{2}\left[\mathrm{tr}\left(R_n R_m^{-1}\right) + \mathrm{tr}\left(R_m R_n^{-1}\right)\right] + \frac{1}{2}\left[(\vec{s}_m - \vec{s}_n)'(R_m + R_n)(\vec{s}_m - \vec{s}_n)\right] \qquad (4\text{-}7)$$

$$I(w,a) = \sum_{n=2}^{L}\sum_{m=1}^{n-1}\theta_n\theta_m\left[I(w,a,m,n) + I(w,a,n,m)\right]$$

$$= \sum_{n=2}^{L}\sum_{m=1}^{n-1}\theta_n\theta_m\left\{-J + \frac{1}{2}\left[\mathrm{tr}\left(R_n R_m^{-1}\right) + \mathrm{tr}\left(R_m R_n^{-1}\right)\right] + \frac{1}{2}\left[(\vec{s}_m - \vec{s}_n)'(R_m + R_n)(\vec{s}_m - \vec{s}_n)\right]\right\}$$

$$= -J\sum_{n=2}^{L}\sum_{m=1}^{n-1}\theta_m\theta_n + \frac{1}{2}\sum_{n=1}^{L}\theta_n\,\mathrm{tr}\left[\left(\sum_{m\neq n}\theta_m R_m\right)R_n^{-1}\right] + \frac{1}{2}\sum_{n=2}^{L}\sum_{m=1}^{n-1}\theta_n\theta_m(\vec{s}_m - \vec{s}_n)'(R_m + R_n)(\vec{s}_m - \vec{s}_n) \qquad (4\text{-}8)$$

The last form of this equation appears to be the most convenient for computing purposes.

4.3 Intuitive Analysis

Looking at the computing form of equation (4-8) it can be seen that there are three terms. The first term is $-J \sum_{n=2}^{L} \sum_{m=1}^{n} \theta_m \theta_n$. The value of this term does not depend upon a and hence has no effect upon the choice of a. From this consideration we note that computing the value of this term would not be beneficial if only one more stage of experimentation is available.

The third term of the sum is a weighted sum of the quadratic forms

$$(\vec{s}_m - \vec{s}_n)'(R_m + R_n)(\vec{s}_m - \vec{s}_n)$$

Thus this term is in effect a separating function in the sense that these quadratic forms will be maximized when the pairs of expected values of $\vec{y}$ under the various hypotheses are as far apart as possible in comparison to the precisions of $\vec{y}$. If the precisions R_m and R_n are large then $\vec{s}_m$ and $\vec{s}_n$ do not need to be far apart to provide much information whereas if these precisions are small then the expected values $\vec{s}_m$ and $\vec{s}_n$ must be further apart to provide the same information. The weighting factors are the products $\theta_n \theta_m$. Thus when θ_n and θ_m are both small, $\theta_n \theta_m$ is very small and the information due to the separation of $\vec{s}_n$ and $\vec{s}_m$ is discounted somewhat. If θ_n and θ_m are large then the information due to separation of $\vec{s}_n$ and $\vec{s}_m$ is given more importance. Thus this third term causes experiments to be chosen which separate the expected values of $\vec{y}$ under the respective hypotheses which are still in serious contention for being chosen.

It is interesting to note that some authors (Hunt and Reiner (1965), e.g.) have proposed criteria for selection of experiments involving only distances between expected values. In a later paper, Box and Hill (1967) proposed that the distances as such are not important, but the distances weighted by some function of the variability about the expected values are important. It is seen here that the expected K-L information function does just that.

The second term in equation (4-8) is $\frac{1}{2}\sum_{n=1}^{L}\theta_n \operatorname{tr}\left[\left(\sum_{m\neq n}\theta_m R_m\right)R_n^{-1}\right]$. This can be thought of as a weighted sum of ratios of precisions. If only one y value is to be observed, this component becomes

$$\frac{1}{2}\sum_{n=1}^{L}\theta_n \frac{\sum_{m\neq n}\theta_m R_m}{R_n} \qquad (4\text{-}9)$$

It would be interesting to see when this term is maximized. Upon taking partial derivatives of equation (4-9), setting to zero, and simplifying, one arrives at the following set of simultaneous nonlinear equations.

$$\sum_{k=1}^{L}\theta_i\left(\frac{R_k^2 - R_i^2}{R_i}\right) = 0 \qquad i = 1, \ldots, L$$

It can be immediately seen that one solution to this system is $R_1 = R_2 = \ldots = R_L$. This solution implies that the experiments should tend to give the same precision for the expected value of $\vec{y}$

under each hypothesis. This term is not considered any further here.

In summary, it can be seen that the expected K-L information function in this case is basically a rather simple separating function. One would be hard pressed to construct a much simpler separating function which has more intuitive appeal. If multivariate observations are permitted, then it might be possible to delete the second term of equation (4-8) to save a good deal of computing.

CHAPTER 5

THE SEQUENTIAL DECISION PROCEDURE

Three components are required for a sequential adaptive decision procedure; (1) a rule which determines if sampling should be terminated or continued, (2) a rule which specifies the experiment to be performed given the current state of the system, and (3) a rule which selects the model equation which will be claimed to be true when sampling is terminated. The first part of this chapter discusses the experiment selection rule and the second section presents the stopping and model selection rules.

5.1 Experiment Selection Rule

The procedure adopted for this dissertation is the so-called myopic procedure. This rule simply chooses as the next experiment that one which maximizes the anticipated K-L information for the next stage only.

We assume that an upper limit, J_{MAX}, to the number of observations is specified. This number may be infinite. An allocation of the observations to the stages of sampling is described by a $J_{MAX} \times 1$ vector $\vec{n}$, where n_i gives the number of observations at stage i. The question arises as to how the observations should be allocated. That is, should all J_{MAX} be taken at once, strictly one-at-a-time, or in different sized groups. As the first step in answering this, let A_j denote the set of experiments in A which

specify that j observations should be taken. For any given state $w\varepsilon\Omega$, let $a_j^*(w)$ denote the element of A_j such that

$$I[w,a_j^*(w)] = \sup_{a_j \varepsilon A_j} I(w,a_j)$$

Lemma 5.1 For any $w\varepsilon\Omega$, and i,j such that $i > j$ we have $I[w,a_i^*(w)] \geq I[w,a_j^*(w)]$.

Proof: We introduce the following notation. Let $y_k(a_i^*)$, $k = 1, \ldots,i$ denote the random variables observed under $a_i^*(w)$ and $y_k(a_j^*)$, $k = 1, \ldots,j$ denote the random variables observed under a_j^*. Define another experiment $\tilde{a}_i \varepsilon A_i$ by choosing the first j observations according to a_j^* and the remaining $i - j$ observations according to the last $i - j$ of a_i^*. This leads to the random variables

$$\tilde{y}_k(\tilde{a}_i) = \begin{cases} y_k(a_j^*) & k = 1, \ldots,j \\ y_k(a_i^*) & k = j + 1, \ldots,i \end{cases}$$

Because $I(w,a,m,n)$ is positive definite and is additive for independent observations

$$I(w,\tilde{a}_i,m,n) \geq I(w,a_j^*,m,n)$$

Thus

$$I(w,\tilde{a}_i) = \vec{\theta}'[I(w,\tilde{a}_i,m,n)]\vec{\theta}$$

$$\geq \vec{\theta}'[I(w,a_j^*,m,n)]\vec{\theta} = I(w,a_j^*)$$

But by definition $I(w,a_i^*) \geq I(w,\tilde{a}_i)$ and hence

$$I(w,a_i^*) \geq I(w,a_j^*)$$

Q.E.D.

The lemma simply proves that an experiment with more observations will be expected to provide more information than one with fewer observations. In determining an allocation one should also consider the cost of experimenting. In particular, if we assume that each observation has a constant cost associated with it, then it is reasonable to choose the experiment which maximizes

$$\frac{1}{j} I(w,a_j) \qquad j = 1, \ldots, J_{MAX}$$

Thus prior to stage k let $m = \sum_{i=1}^{k-1} n_i$ and assume $m < J_{MAX}$. The optimal experiment is the element $a^* \epsilon A$ which for the current state w_{k-1} yields

$$j = 1, \ldots, J_{MAX}^{MAX} - m \left\{ \max_{a \epsilon A_j} \frac{1}{j} I(w_{k-1},a) \right\}$$

If sampling has not been terminated by the rules developed in Chapter 5.2, then we stop when $\Sigma n_i = J_{MAX}$ and select the model according to the rules in Chapter 5.2.

5.2 Stopping and Model Selection Rules

We now discuss the problems of determining which of the postulated models is the true one and determining when the results of the experiments are sufficiently informative to stop sampling and make the choice.

Box and Hill (1967) suggested that for their procedure, experimenting be terminated whenever one model is clearly superior to

the others. This is obviously a reasonable statement but it is in need of formal definition before it can be used as a stopping and selection rule. We propose general stopping and selection rules and a modified version which might be used in certain instances involving nested models.

(1) Stopping rule: Let θ_m be some specified value $1/L < \theta_m \leq 1.0$. Let J_{MAX} denote the maximum number of observations permitted. Then terminate sampling whenever either $\max_{i=1,L}\{\theta_i\} \geq \theta_m$ or J_{MAX} observations have been taken, whichever occurs first.

(2) Model selection rule: Upon termination choose the correct model to be H_j* where $\theta_j* = \max_{i=1,L}\{\theta_i\}$.

We now present a modified stopping and selection procedure for use with nested models which may be of some value when θ_m is very near 1.0 and/or when J_{MAX} is relatively large. The reason for presenting a modified procedure arises from the large sample results of Chapter 3 and the Monte-Carlo results of Chapter 6. First, if H_i* denotes the unknown true model, it is not known whether $\theta_{i*,j} \to 1.0$ or not. From the Monte-Carlo results it seems that the typical behavior of $\theta_{i*,j}$ for nested models is to fairly rapidly increase to something near 1.0 and then fluctuate, possibly slowly approaching 1.0. Thus, if θ_m is very near 1.0 it may be that extremely large samples would be required. Thus we would like to reduce the average sample size without seriously detracting from the probability of choosing the correct model.

To introduce the modified procedure consider the following example:

$$H_1: \quad y = \beta_1 x_1 + \varepsilon$$

$$H_2: \quad y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

If H_1 is actually the true model, then the posterior distribution of (β_1, β_2) under H_2 should approach a point distribution with $\mu_2^{(2)} = 0$ and $\mu_1^{(2)}$ equal to the unknown value of the parameter. However, $\theta_{1,j}$ may not approach 1.0. Assume some small positive constant γ is specified. Then after each stage of sampling, test if

$$d = \left[\mu_1^{(1)} - \mu_1^{(2)}\right]^2 + \left[\mu_2^{(2)}\right]^2 \leq \gamma$$

If $d \leq \gamma$ then drop model 2 from contention and replace θ_1 by $\theta_1 + \theta_2$. Then apply the previously described stopping and selection rules. In this simple example, the dropping of model 2 would automatically cause sampling to be terminated. This would of course not necessarily be true in more general situations.

To generalize the procedure some additional notation and concepts must be introduced. We use the symbol $\supset$ to denote inclusion. Thus $H_i \supset H_j$ means that the model of H_j is nested within the model of H_i. The set of models $\{H_i\}$ is a partially ordered set under the partial ordering relation $\supset$. In the theory of partially ordered sets a <u>chain</u> is defined to be a partially ordered set such that for any two elements (H_1 and H_2 say) of the set either $H_1 \supset H_2$ or $H_2 \supset H_1$. For the purposes of this dissertation we define a <u>string</u> of elements from the partially ordered set as a sub-

set of elements such that the subset forms a chain. A maximal string is constructed from any string by adding all the elements of $\{H_i\}$ to the string which can be added without causing the enlarged set to lose the property of being a chain.

To formulate the modified stopping and selection procedure we first construct all of the maximal strings that can be constructed from the set $\{H_i\}$ and order the elements of the strings using the relation $\supset$.

For example, suppose $L = 5$ and the five models are as specified below:

Model number	Model equation
1	$y = \beta_0 + \beta_1 x_1 + \varepsilon$
2	$y = \beta_0 + \beta_3 x_3 + \varepsilon$
3	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
4	$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon$
5	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

The maximal strings are easily verified as being

$$(H_1 \subset H_3 \subset H_5)$$

$$(H_1 \subset H_4 \subset H_5)$$

$$(H_2 \subset H_4 \subset H_5)$$

In each of these strings the maximal element is H_5.

The modified procedure consists of computing for the first

maximal string the squared distance of the posterior mean vector of the maximal element from the posterior mean vector of the submaximal or next largest submodel of the string. If this quantity is less than some prespecified value γ, the maximal model is dropped from the set $\{H_i\}$. The posterior probability of the maximal element is added to the probability of the next element of the string until either only one model remains or there is no need to drop models. Before considering the next maximal string, all models which have been dropped must also be deleted from the remaining strings. The above procedure is then repeated for each maximal string in turn.

Once this has been completed and all models which can be dropped because they reduce to models with fewer parameters have been dropped, the same stopping and selection rules proposed for the non-nested case are applied.

Note: The procedure just described is not necessarily the best or the most natural one to use for combining models. For example, an alternative to the distance of the means might be to combine models when the probability distribution of the maximal element is sufficiently concentrated about the mean of the submaximal element. This would have the advantage of using the information contained in the precisions of the distributions also.

CHAPTER 6

COMPUTER SIMULATION RESULTS

The purpose of this chapter is to report and discuss the results of a number of Monte-Carlo simulation studies of the sequential procedure proposed in Chapter 5. The chapter is divided into four major sections. The first section describes the general simulation procedure and presents a brief description of the algorithm used. A computer program based on this algorithm is described in further detail in appendix A. The second section presents and discusses the results of a number of simulations performed to gain further information about the large sample behavior first discussed in Chapter 3. The primary concern is with the posterior probabilities and the posterior means of the parameter distributions after a large number of observations. The third section presents and discusses the results of some simulation studies of the proposed sequential procedure when the number of observations is limited and when the stopping rule of Chapter 5 is used. The primary concern is with the probability of the procedure actually selecting the correct model and the average sample size required until termination. The last section of the chapter presents an overall discussion of the results.

6.1 General Simulation Procedure

The sequential procedure proposed in Chapter 5 consisted of

(1) an experiment termination rule, (2) an experiment selection rule, and (3) a model selection rule. Because of the mathematical complexity of the posterior distributions involved it was not feasible to analytically examine how well these rules work. The general procedure by which the Monte Carlo simulation technique was used to study performance is outlined in the following algorithm.

1. Input:

   $\vec{\mu}_{\ell,0}$ the prior means of the parameters of the models

   $\Psi_{\ell,0}$ the prior precision matrices of the parameters of the models

   $\theta_{\ell,0}$ the prior probabilities of the models being correct

   N the number of simulations

   θ_m stopping probability

   J_{MAX} maximum number of observations

   i^* the model chosen to generate the observed variable

   $\vec{\mu}^*$ values of the parameters of the true model

2. $n \leftarrow 0$

3. $PCS \leftarrow 0$

4. $\bar{N}_i \leftarrow 0$ (for $i = 1, J_{MAX}$)

5. $j \leftarrow 0$

6. $j \leftarrow j + 1$

7. Determine optimal acA as described in Chapter 4. Denote as a^* and let M_{a^*} denote design matrix for model i^* when a^* is chosen. (All simulations in this dissertation

consider strictly one-at-a-time sampling for simplicity.)

8. $y_j \leftarrow M_{a^*}^{\vec{\mu}^*}$

9. Generate a pseudo-random observation ε_j from a $N(0,\tau)$ distribution. (Described in detail in appendix A)

10. $y_j \leftarrow y_j + \varepsilon_j$

11. For $\ell = 1, \ldots, L$ compute $\theta_{\ell,j}$, $\Psi_{\ell,j}$, and $\vec{\mu}_{\ell,j}$ from y_j and $\theta_{\ell,j-1}$, $\Psi_{\ell,j-1}$, and $\vec{\mu}_{\ell,j-1}$ as described in Chapter 3.

12. Find k such that $\theta_{k,j} = \underset{i}{\text{MAX}}\{\theta_{i,j}\}$

13. If $j \geq J_{MAX}$ or $\theta_{k,j} \geq \theta_m$ go to 14. Otherwise go to 6.

14. $\bar{N}_j \leftarrow \bar{N}_j + 1$

15. If $k = i^*$; PCS $\leftarrow$ PCS + 1

16. $n \leftarrow n + 1$

17. If $n \geq N$ go to 18. Otherwise go to 5.

18. PCS $\leftarrow$ PCS/N

19. $\text{ASN} \leftarrow \left(\sum_{i=1}^{J_{MAX}} i\bar{N}_i\right)/N$

20. Stop

Upon stopping, the value of PCS is the observed probability of correctly choosing i^* as the true model for the prior distributions specified when in fact the true value of the parameters is given by $\vec{\mu}^*$. ASN gives the average sample number upon termination.

The above algorithm can be easily used for either large sample or small sample studies. For example, for large sample studies set $\theta_m = 1.0$, $N = 1$, and J_{MAX} to some large number, say 100 or 500.

For small sample studies set $\theta_m < 1.0$, J_{MAX} to some small number, and N to some larger number, say 500 or 1000.

6.2 Large Sample Studies

In this section we examine the large sample properties of the posterior probabilities of the models and the posterior means of the parameter distributions. Three sets of problems are studied. First, two sets of nested polynomial models are studied. The posterior probabilities of each model, the posterior means of the parameter distributions, and the proportion of times each of the allowable values of the independent variable is chosen as optimal are tabulated for simulations of 100 and 500 observations. Second, one set of nested factorial models is studied for three different prior distributions on the models. And third, one set of non-nested factorial models is simulated. For the last two, the posterior probabilities and means of the parameter distributions are tabulated.

6.2.1 Polynomial Model Studies

Two sets of nested polynomial models are considered which have the following general form:

$$H_\ell: \quad y = \sum_{j=0}^{\ell-1} \beta_j x^j + \varepsilon, \quad \ell = 1, L$$

Two values of L are studied, and for each of these choices, two choices of H_i^* are made. The values of τ, $\theta_{\ell,0}$, and $\Psi_{\ell,0}$ are specified as

$$\tau = 100.0$$

$$\Psi_{\ell,0} = I$$

$$\theta_{\ell,0} = \frac{1}{L}$$

for all simulations. The values of $\vec{\mu}_{\ell,0}$ are tabulated at the tops of figures 1 and 2 and the resulting functions are graphed on the interval x [-1,+1] at the bottoms of the respective figures. For L = 4, the two choices of H_{i*} are H_2 and H_3. For L = 6, the two choices of H_{i*} are H_3 and H_5. For simplicity, the actual values of the parameters used to generate the data were chosen to be $\vec{\mu}_{i*,0}$ for each of the four cases.

For these simulations, the definition of A was arbitrarily taken to be

$$A = \{a^{(i)}: \ i = 0, \ldots, 9\}$$

where

$$a^{(i)}: \ x = -1 + \frac{2i}{9}$$

Note that sampling is strictly one observation per stage.

The simulation results are summarized in table 1 and given in further detail in tables 2 through 9. For each choice of L and i^*, five simulations of 100 observations and five simulations of 500 observations were performed. For these simulations, the sample paths of the $\theta_{\ell,j}$ were printed out and the choice of $a^{(i)}$ at each stage were printed. The posterior means of the parameter distributions were printed only after the last stage. Tables 2, 4, 6, and 8 give the posterior probabilities after 100 observations and the first 100 out of 500 observations. The proportions p_i of using $a^{(i)}$ are also given. Tables 3, 5, 7, and 9 give the same informa-

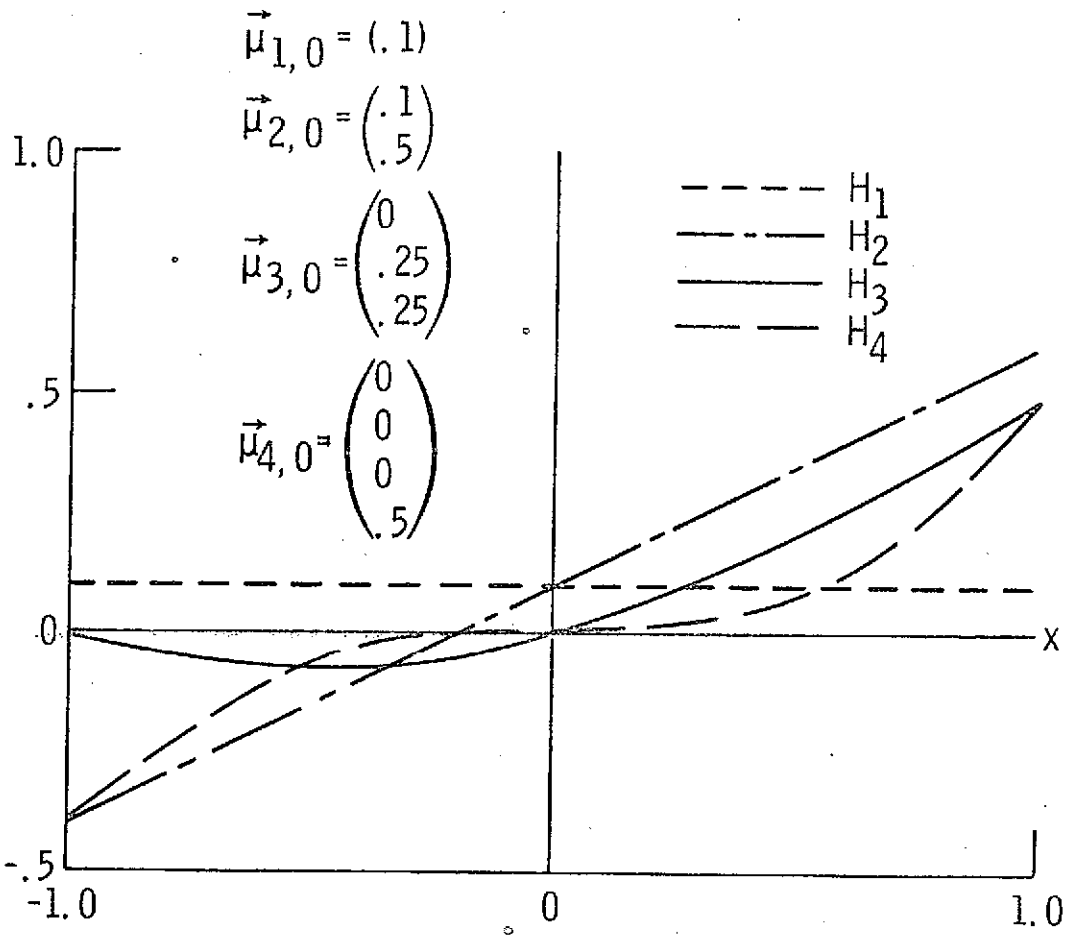


Figure 1. - Tabulations of the prior means of the parameters and graphs of the resulting functions over the interval $[-1, +1]$ for large sample polynomial study one.

$$\vec{\mu}_{1,0} = (0) \qquad \vec{\mu}_{5,0} = \begin{pmatrix} 0 \\ .5 \\ 0 \\ 0 \\ .1 \end{pmatrix}$$

$$\vec{\mu}_{2,0} = \begin{pmatrix} 0 \\ .5 \end{pmatrix}$$

$$\vec{\mu}_{3,0} = \begin{pmatrix} 0 \\ .25 \\ .25 \end{pmatrix}$$

$$\vec{\mu}_{4,0} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ .5 \end{pmatrix} \qquad \vec{\mu}_{6,0} = \begin{pmatrix} 0 \\ .5 \\ 0 \\ 0 \\ 0 \\ .1 \end{pmatrix}$$

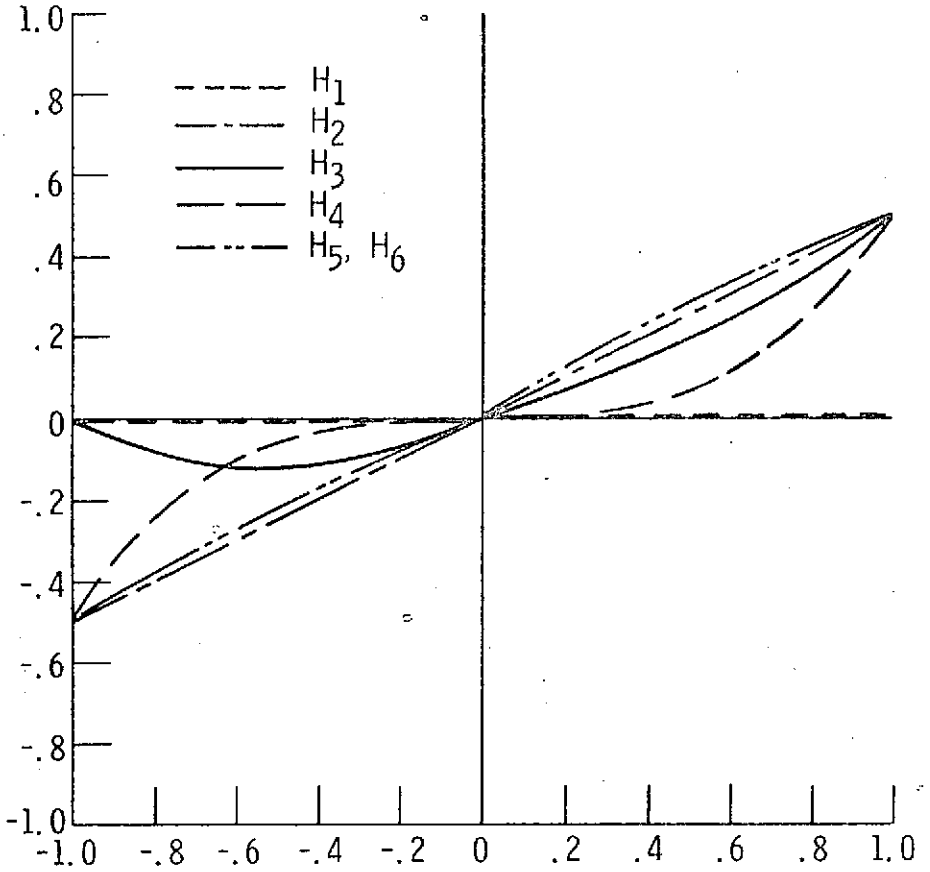


Figure 2. - Tabulations of the prior means of the parameters and graphs of the resulting functions over the interval $[-1, +1]$ for large sample polynomial study two.

tion for the 500 observation simulations.

Figures 3 and 4 present typical sample paths for the posterior probability of the correct model. In figure 3, the value of $\theta_{2,j}$ is plotted for the first 250 observations of the third simulation for $L = 4$ and $i^* = 2$. In figure 4, the value of $\theta_{3,j}$ is plotted for the first 250 observations of the first simulation for $L = 4$ and $i^* = 3$. These figures illustrate the typical behavior of $\theta_{i^*,j}$. It fairly rapidly rises to a value of about 0.85 to 0.95 and then slowly and erratically oscillates. This is suspected to be because of the nested nature of the model equations. It was because of this behavior that the modified selection rule of Chapter 5 was first introduced. Consideration of the posterior means of the parameter distributions will also provide some information concerning this modified rule.

For $L = 4$, consideration of tables 2, 3, 4, and 5 show that as j increases, $\vec{\mu}_{i,j} \to \begin{pmatrix} \vec{\mu}_{i^*,j} \\ \vec{0} \end{pmatrix}$ for $i > i^*$. This is in accord with the conclusions of Chapter 3. For $L = 6$ and $i^* = 3$ we again see the same close agreement with Chapter 3 as evidenced by tables 6 and 7. However, for $i^* = 5$, an entirely different situation arises. To understand this we should note that the model used to generate the sequential observations is

$$y = 0.5\, x + 0.1\, x^4 + \varepsilon$$

This function can be very closely approximated by a model of the form

$$y = ax + bx^2 + \varepsilon$$

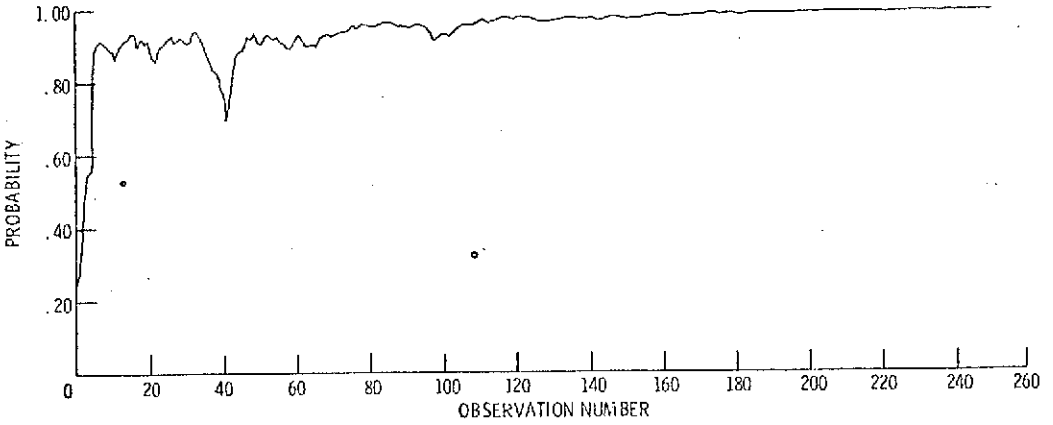


Figure 3. - The sample path of $\theta_{2,j}$ for L = 4, i° = 2 for the first 250 observations of simulation no. 3. A well behaved path for nested models.

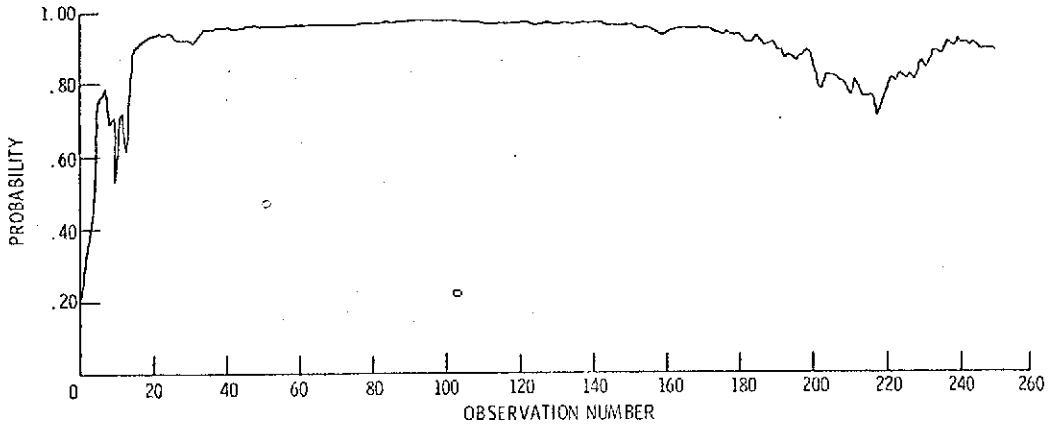


Figure 4. - The sample path of $\theta_{3,j}$ for L = 4, i° = 3 for the first 250 observations of simulation no. 1. A typical path for nested models.

over the range of x values considered. And in fact we note that there is a marked preference for choosing the lower degree model as indicated by $\theta_{3,j}$ becoming close to 1.0. It is also interesting to note the behavior of $\vec{\mu}_{i,j}$ for $i > 3$. We do not see that $\vec{\mu}_{i,j} \to \begin{pmatrix} \vec{\mu}_{3,j} \\ \vec{0} \end{pmatrix}$ as might be expected when H_3 is so close to being true, except for the case of $i = 4$. For $\vec{\mu}_5$ we note that the average posterior mean of the coefficient of x^3 is quite close to zero and the sum of the posterior means of the coefficients of x^2 and x^4 is quite close to 0.1. For $\vec{\mu}_6$ we note that the sums of the posterior means of the coefficients of x^2 and x^4 is close to 0.1 and the sum of the posterior means of the coefficients of x, x^3, and x^5 is close to 0.5. From these simulation studies it is not clear whether this behavior is simply because 500 observations is not a sufficiently large number to discriminate well between such nearly equivalent functions or if this behavior will persist no matter how large the number of observations.

We now turn to a discussion of the observed proportions of times the $a^{(i)}$ were chosen as the optimal experiments. From tables 2 and 3 which present the results of $L = 4$ and $i^* = 2$ we see that the largest p_i are for p_0, p_4, p_5, and p_9. These correspond to $x = -1$, $x = -1/9$, $x = +1/9$, and $x = +1$. Because of the discretization of the interval $(-1,+1)$ we might assume that the asymptotically most informative experiments were $x = -1$, $x = 0$, and $x = +1$. From tables 4 and 5 we see the largest p_i are p_0, p_2, p_7, and p_9 corresponding to $x = -1$, $x = -5/9$, $x = +5/9$, and

$x = +1$. The relationship of these proportions and x points to the experimental designs which are optimal from other considerations might be interesting. For example, Kiefer and Wolfowitz (1959) consider optimal designs for regression problems of a somewhat different nature. The comparison of the current results with such other works is currently being pursued but will not be reported in this dissertation.

6.2.2 Nested Factorial Models

A second set of simulation studies were made using the following models

$$H_1: \quad y = \beta_0 + \varepsilon$$

$$H_2: \quad y = \beta_0 + \beta_1 x_1 + \varepsilon$$

$$H_3: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$H_4: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

with $\tau = 2.0$, $\Psi_{\ell,0} = I$, and $i^* = 3$. The prior means $\vec{\mu}_{\ell,0}$ were chosen as

$$\vec{\mu}'_{1,0} = (0)$$

$$\vec{\mu}'_{2,0} = (0,1)$$

$$\vec{\mu}'_{3,0} = (0,1,-1) = \left(\vec{\mu}^*\right)'$$

$$\vec{\mu}'_{4,0} = (0,1,-1,0)$$

Three sets of $\theta_{\ell,0}$ were chosen:

1. $\theta_{1,0} = 0.1$

$\theta_{2,0} = 0.2$

$\theta_{3.0} = 0.3$

$\theta_{4,0} = 0.4$

2. $\theta_{\ell,0} = 0.25 \qquad \ell = 1,4$

3. $\theta_{1,0} = 0.4$

$\theta_{2,0} = 0.3$

$\theta_{3,0} = 0.2$

$\theta_{4,0} = 0.1$

The experiment space A is defined as $A = \{(x_1,x_2): x_i = \pm 1\}$. Note that experimenting is strictly one-at-a-time. The sequential selection procedure of Chapter 5 was used for five simulations of 500 observations each. The results are presented in table 10. We note that the posterior values are again in close agreement with Chapter 3 and the results of the polynomial models. There does not seem to be a pronounced effect upon the posterior probabilities of the models from changing the prior distribution although there does appear to be slightly higher posterior values of $\theta_{3,500}$ when the $\theta_{\ell,0}$ distribution is skewed toward the lower values. A possible explanation for this is that when this distribution is skewed toward the high values, the procedure is choosing experiments primarily to discriminate between H_3 and H_4. Since H_3 is true, the model of

H_4 will rapidly become close to that of H_3 and the resulting experiments will not be very informative. When the prior probabilities are larger for the lower degree polynomials, however, the procedure chooses experiments primarily to discriminate between H_1 and H_2. These experiments should then more rapidly tend to prove H_1 and H_2 to be inadequate.

6.2.3 Non-Nested Factorial Models

In this study, the following non-nested models were studied.

$$H_1: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$H_2: \quad y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon$$

$$H_3: \quad y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

The values of the parameters are chosen as

$$\left.\begin{aligned} i^* &= 1 \\ \vec{\mu}_{\ell,0} &= (1,1,1)' \\ \Psi_{\ell,0} &= I \\ \theta_{\ell,0} &= \frac{1}{3} \end{aligned}\right\} \quad \ell = 1, \ldots, L$$

$$\tau = 0.01, 1, 100$$

$$\vec{\mu}^* = (1,1,1)'$$

The experiment space A was assumed to allow only one observation at a time with $x_i = \pm 1$. Five simulations were performed for each value of τ. For $\tau = 100$ it took only three observations for θ_3 to become 1.0 (within the accuracy of the computer). For $\tau = 1.0$ the number of observations required for the final posterior probabilities to reach 1.0 are tabulated in table 11. For $\tau = 0.01$,

1000 observations were taken and the resulting posterior probabilities are given in table 11. Again the results are in general agreement with Chapter 3.

6.3 Small Sample Performance Studies

In this section we examine the performance of the proposed sequential procedure as measured by the PCS and ASN values. First, two studies are presented of the problem of discriminating among the three models

$$H_1: \quad y = \beta_1 x_1 + \varepsilon$$

$$H_2: \quad y = \beta_2 x_2 + \varepsilon$$

$$H_3: \quad y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

The first study assumes H_3 is true and the second study assumes H_2 is true. The experiment space A is defined as

$$A = \{(x_1, x_2): \ x_i = \pm 1; \text{ one-at-a-time sampling}\}$$

Then we consider the problem of choosing among the four nested models.

$$H_1: \quad y = \beta_0 + \varepsilon$$

$$H_2: \quad y = \beta_0 + \beta_1 x_1 + \varepsilon$$

$$H_3: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$H_4: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

where A is as in the first two studies.

The primary reasons for concentrating on these simple models are that the parameter spaces are of a low enough dimension that

they can be visualized and they are small enough that extensive simulation studies would not require inordinate amounts of computer time. The fact that the parameter spaces can be visualized allows the effect of varying prior means upon PCS and ASN to be more easily grasped. Note that the modified stopping and selection rule was not used.

6.3.1 Study One - H_3 Assumed True

We study discriminating among

$$H_1: \quad y = \beta_1 x_1 + \varepsilon$$

$$H_2: \quad y = \beta_2 x_2 + \varepsilon$$

$$H_3: \quad y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$A = \{(x_1, x_2): \quad x_i = \pm 1; \text{ one-at-a-time sampling}\}$$

where

$$\Psi_{\ell,0} = I \qquad \theta_{\ell,0} = \frac{1}{3}$$

$$\vec{\mu}_{1,0} = (1.0) \qquad \vec{\mu}_{2,0} = (1.0)$$

and

$$\vec{\mu}^{*} = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$$

Then a number of simulation experiments were performed for each combination of:

$$\tau = 0.50, 1.0, 2.0$$

$$\theta_m = 0.70, 0.80, 0.90$$

$$J_{MAX} = 8, 16$$

$$\vec{\mu}_{3,0} = \begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix}, \begin{pmatrix} 0.50 \\ 0.50 \end{pmatrix}, \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}, \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}$$

The experiments for $J_{MAX} = 8$ used 1500 simulations and for $J_{MAX} = 16$ used 1000 simulations.

The choice of prior means deserves some comment. Figure 5 illustrates the points in (β_1, β_2) coordinate space corresponding to the prior means. The points corresponding to $\vec{\mu}_{1,0}$ and $\vec{\mu}_{2,0}$ are as close to $\vec{\mu}^*$ as possible since $\vec{\mu}_{1,0}$ is restricted to the horizontal axis and $\vec{\mu}_{2,0}$ to the vertical. The four choices for $\vec{\mu}_{3,0}$ then span a range about $\vec{\mu}^*$ and hence the resulting PCS and ASN values will indicate the importance of mis-specified prior means.

Tables 12 and 13 present the observed PCS and ASN values for the combinations of θ_m, τ, and $\vec{\mu}_{3,0}$. These results are also plotted as parametric surfaces in figures 6 through 9.

In general, the results are about what should be expected. The PCS increases with τ and ASN decreases with τ. PCS increases as $\vec{\mu}_{3,0}$ gets closer to $\vec{\mu}^*$. We also note that in most cases, PCS increases with θ_m for fixed τ and $\vec{\mu}_{3,0}$. There is, however, a distinct dropping off of PCS with θ_m along the peaks of the surfaces. There does not seem to be any ready explanation for this.

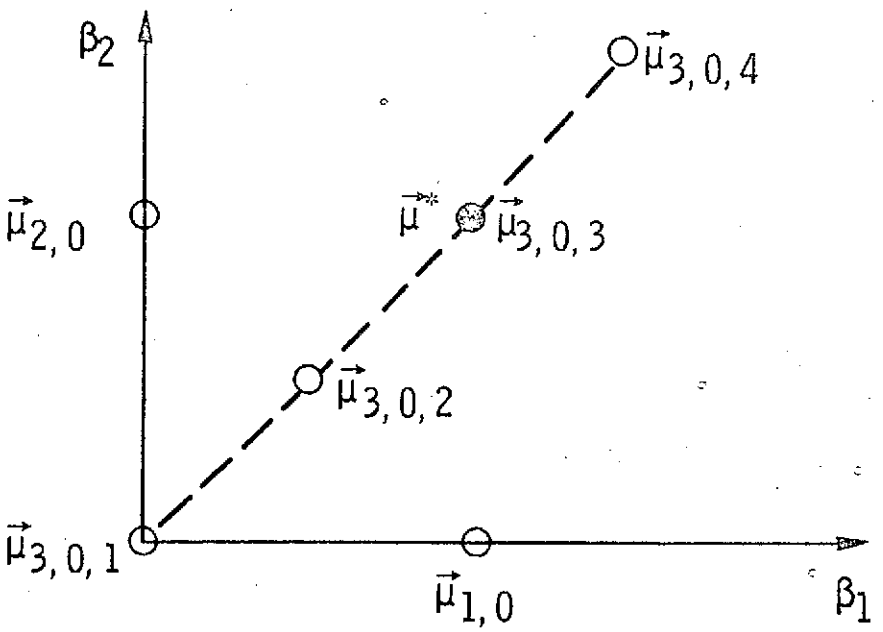


Figure 5. - Illustration of prior means for performance simulation experiment one.

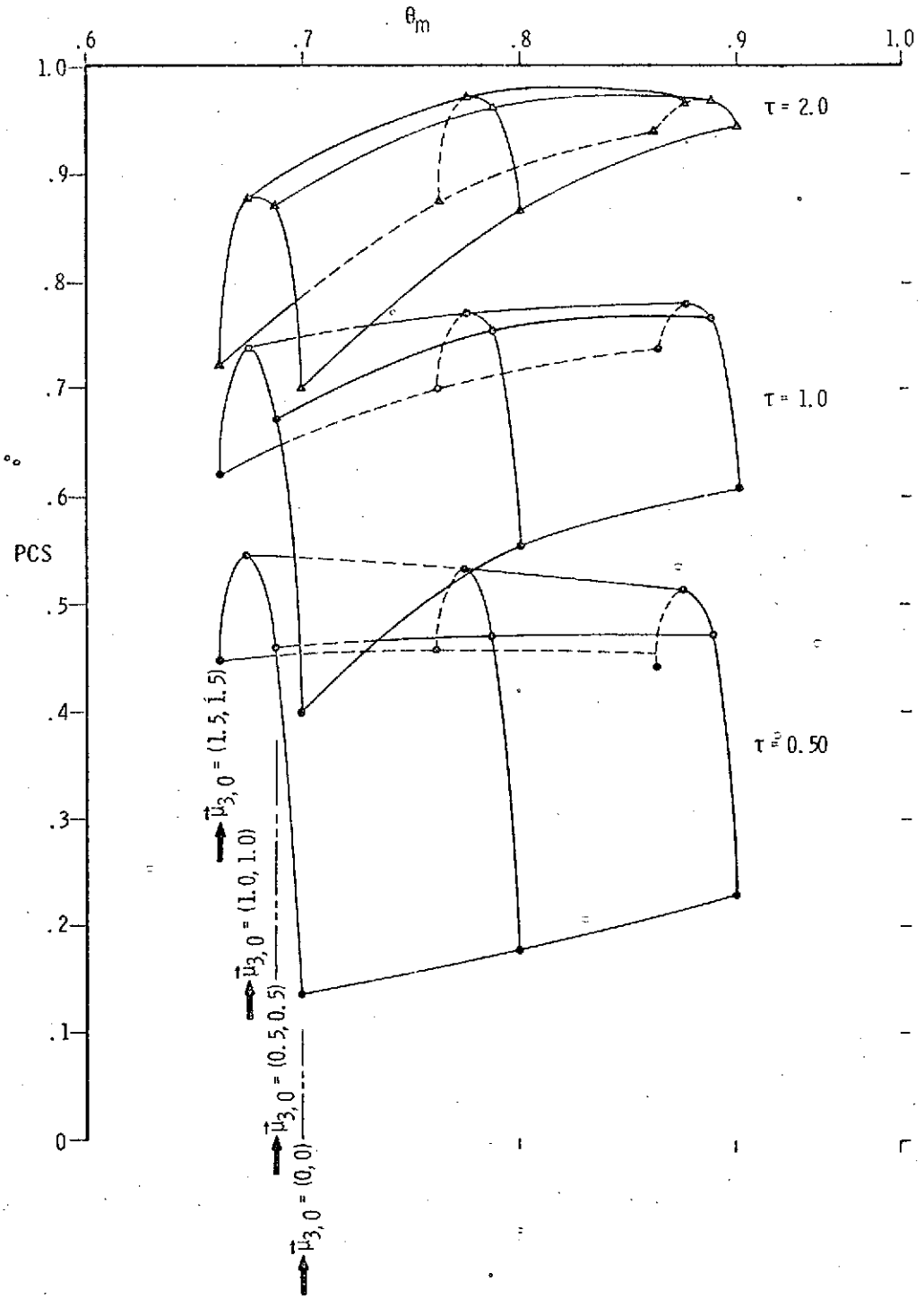


Figure 6. - Probability of correct selection (PCS) as a function of θ_m, τ, $\vec{\mu}_{3,0}$ for $J_{max} = 8$ and H_3 true. Small sample performance simulation experiment one.

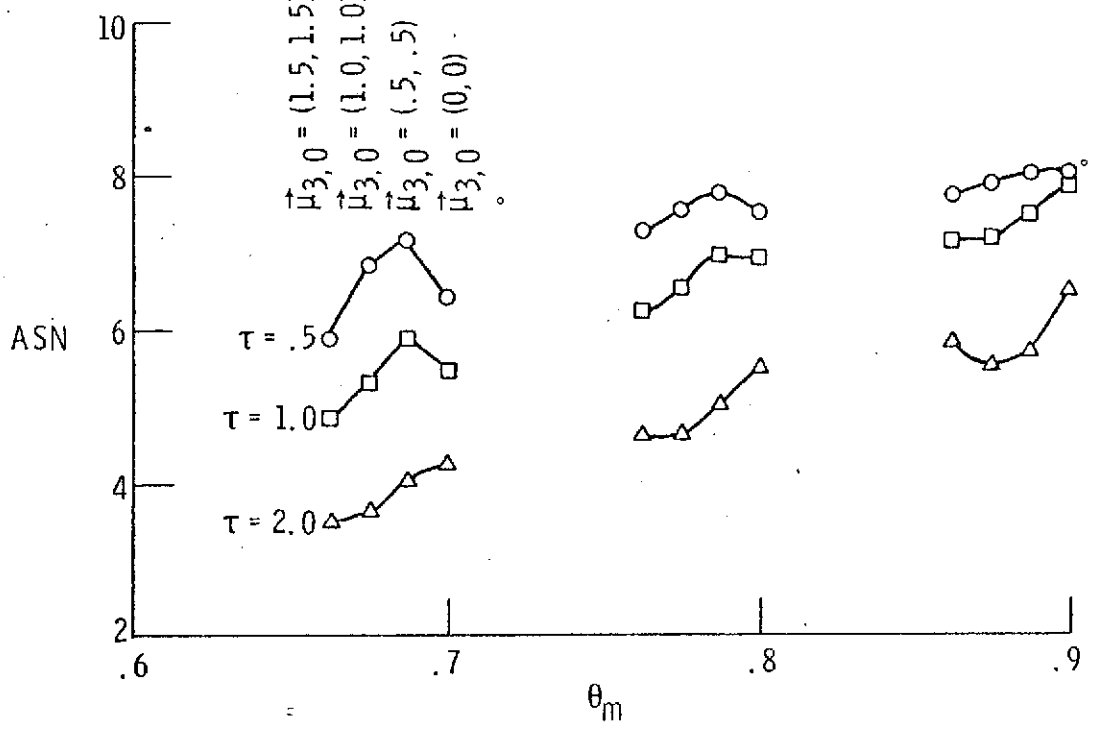


Figure 7. - Average sample number (ASN) as a function of θ_m, τ, $\vec{\mu}_{3,0}$ for $J_{max} = 8$ and H_3 true. Small sample performance simulation experiment one.

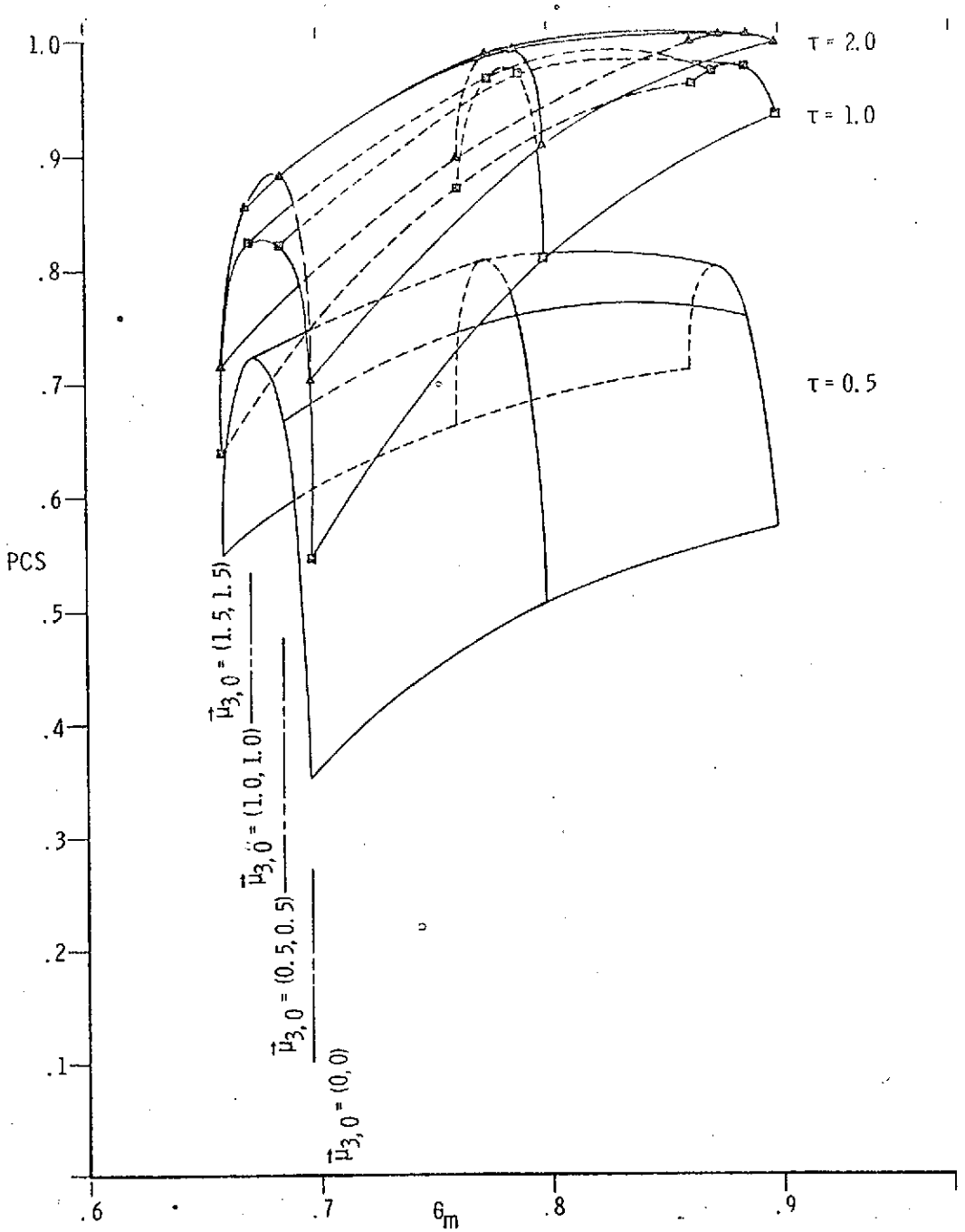


Figure 8. - Probability of correct selection (PCS) as a function of θ_m, τ, $\vec{\mu}_{3,0}$ for $J_{max} = 16$ and H_3 true. Small sample performance simulation experiment one.

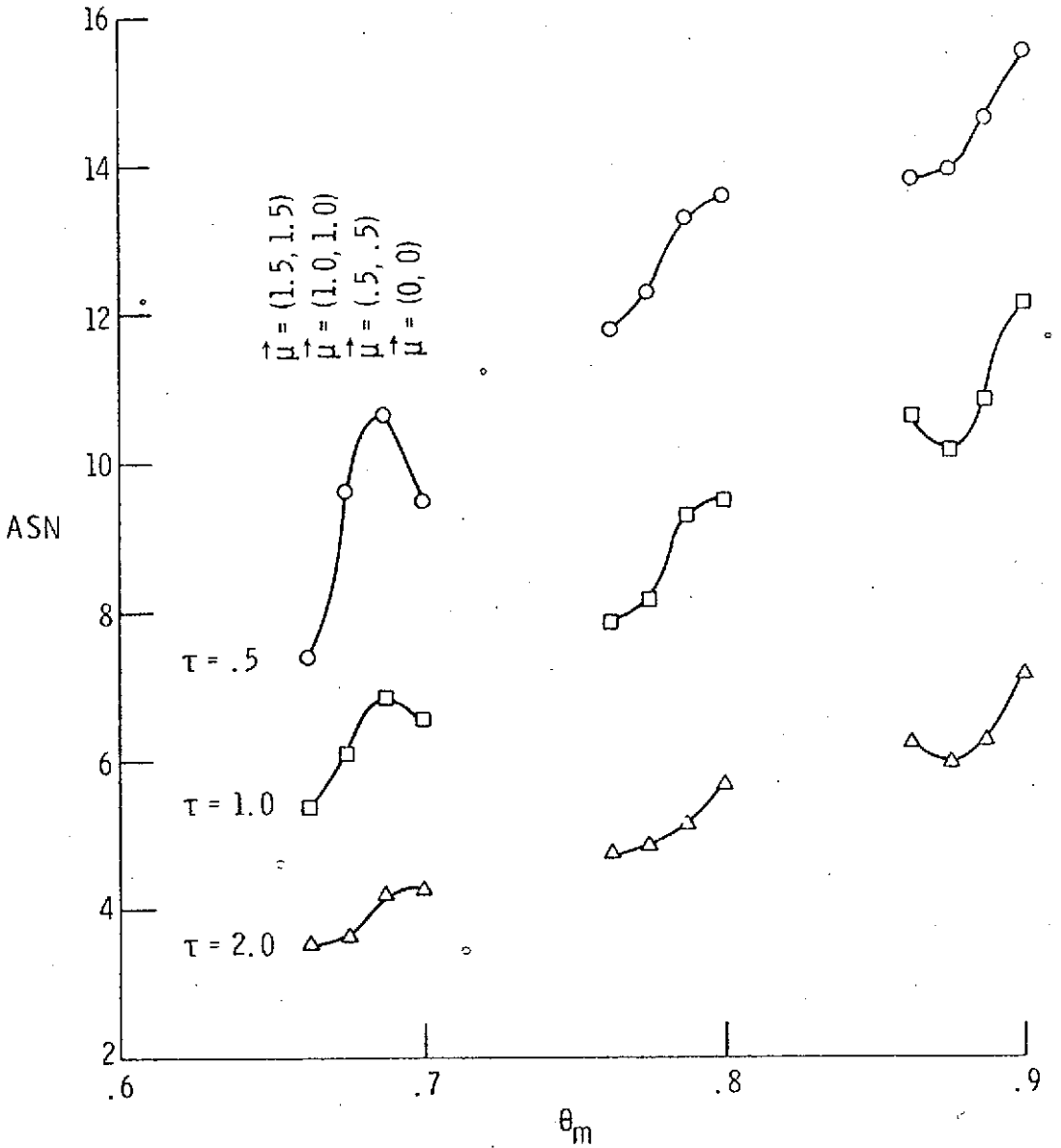


Figure 9. - Average sample number (ASN) as a function of θ_m, τ, $\vec{\mu}_{3,0}$ for $J_{max} = 16$ and H_3 true. Small sample performance simulation experiment one.

6.3.2 H_2 Assumed True

A much less extensive study of this case was made than the case of H_3 assumed true. The same model equations were postulated and we assume

$$\left.\begin{aligned} \Psi_{\ell,0} &= I \\ \theta_{\ell,0} &= \frac{1}{3} \end{aligned}\right\} \quad \ell = 1, \ldots, L$$

$$\vec{\mu}_{1,0} = (0.0)$$

$$\vec{\mu}_{3,0} = (0.0,\ 1.0)'$$

$$\vec{\mu}^{*} = (1.0)$$

The values of τ, θ_m, and $\vec{\mu}_{2,0}$ which were simulated are tabulated in table 14 along with the simulation results. Figure 10 illustrates the prior means. Only one level of J_{MAX} (=8) was considered. Also, only 500 simulations were performed for each of these cases. The PCS results are also graphed as a parametric surface in figure 11. The results are generally the same as for H_3 true.

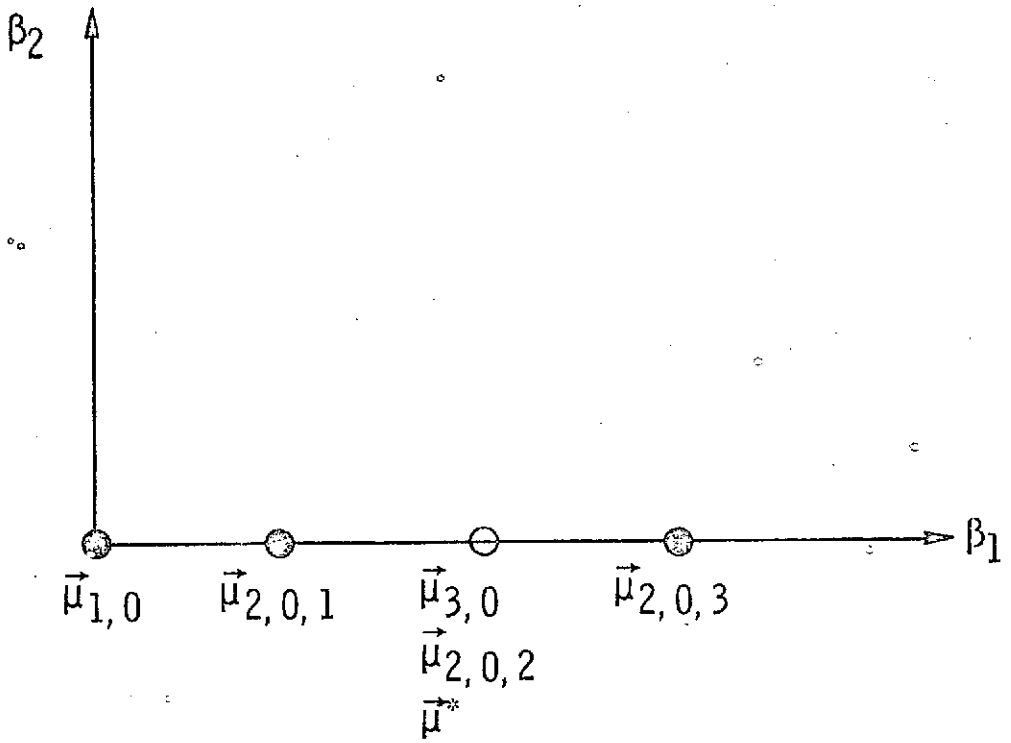


Figure 10. - Illustration of prior means for small sample performance simulation experiment two.

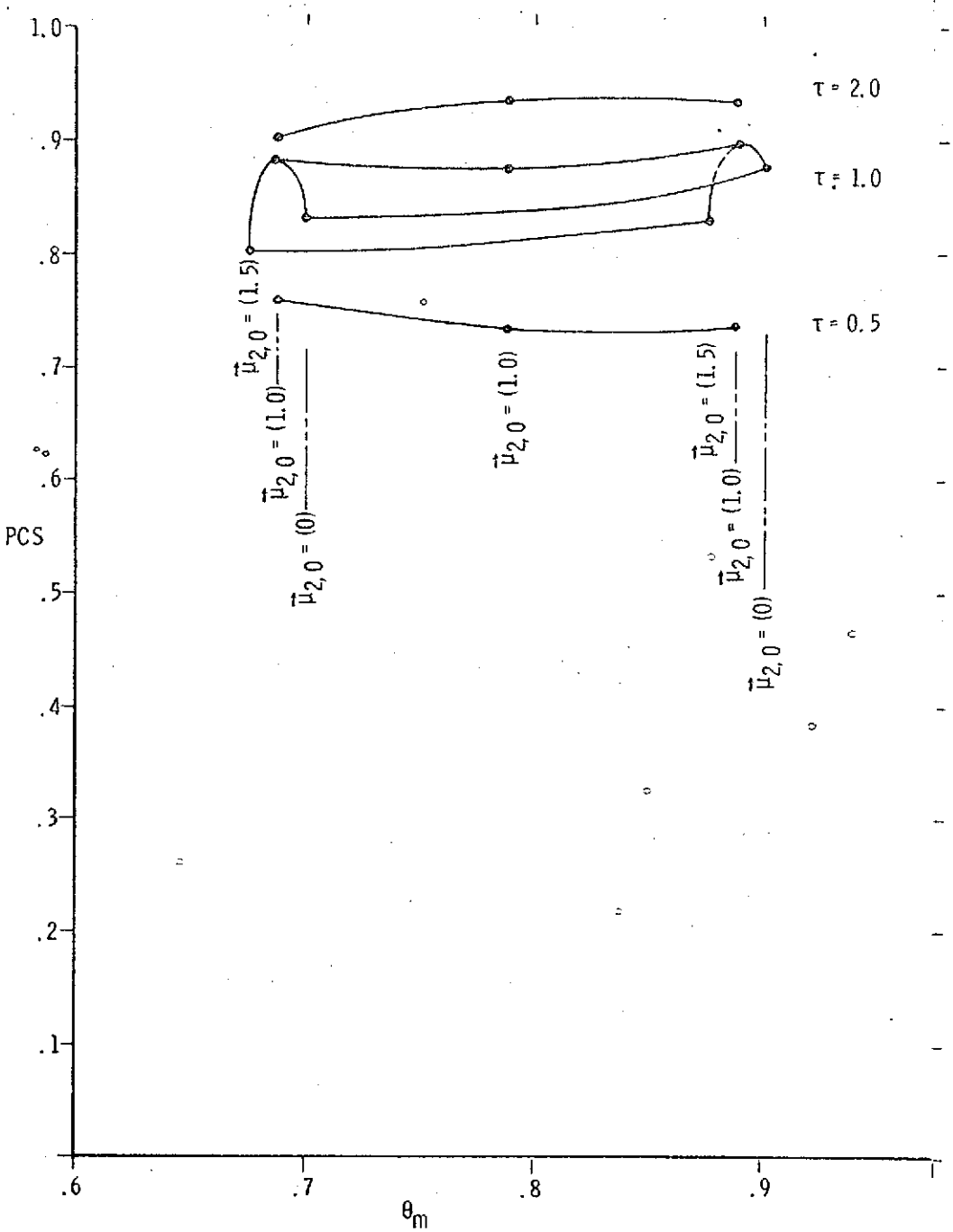


Figure 11. - Probability of correct selection (PCS) as a function of θ_m, τ, $\vec{\mu}_{2,0}$ for $J_{max} = 8$ and H_2 true. Small sample performance simulation experiment two.

6.3.3 A Four Model Problem

In this section we study the ability of the sequential procedure to choose the correct model from the following set of completely nested model equations.

$$H_1: \quad y = \beta_0 + \varepsilon$$

$$H_2: \quad y = \beta_0 + \beta_1 x_1 + \varepsilon$$

$$H_3: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$H_4: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

The prior distributions are defined by

$$\theta_{\ell,0} = \frac{1}{4}$$

$$\Psi_{\ell,0} = I$$

$$\vec{\mu}_{1,0} = (0), \ \vec{\mu}_{2,0} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \vec{\mu}_{4,0} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

And

$$\vec{\mu}_{3,0} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0.5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0.5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0.5 \end{pmatrix}$$

The equation used to generate the observations was that of H_3 with values of the parameters given by

$$\vec{\mu}^* = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

The value of τ used was 1.0.

For fixed values of β_3, the values of the prior parameter means for H_1 and H_2 and the four prior means for H_3 can be plotted in 3-space as in figure 12. For each of the four choices of $\vec{\mu}_{3,0}$, three values of θ_m (=0.7, 0.8, 0.9) for $J_{MAX} = 8$ were used and the resulting PCS and ASN values for the 12 combinations are given in table 15. PCS is also plotted in figure 13. For all cases, 1000 simulations of the procedure were performed.

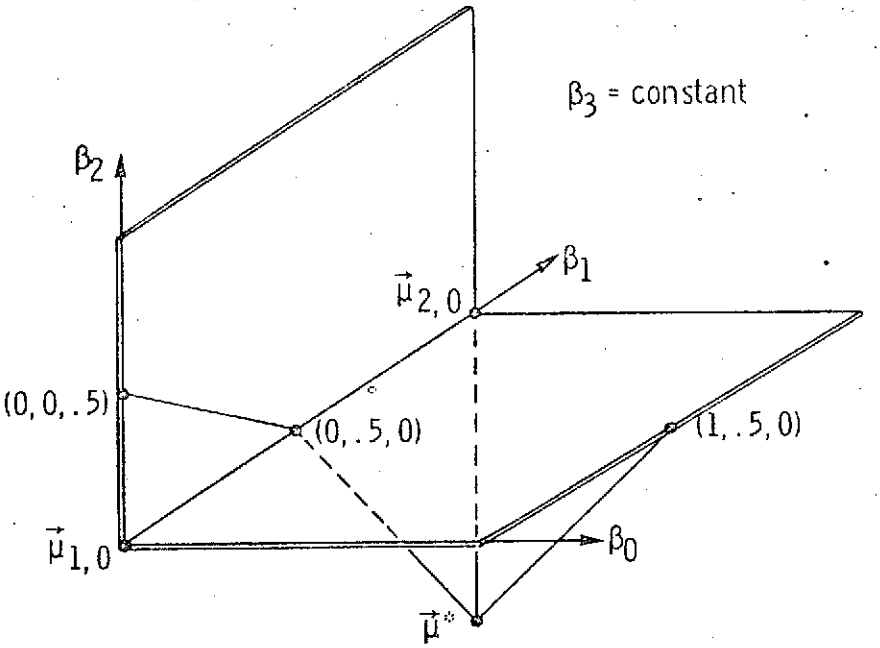


Figure 12. - Illustration of the prior means when β_3 is held constant. Small sample performance simulation experiment three.

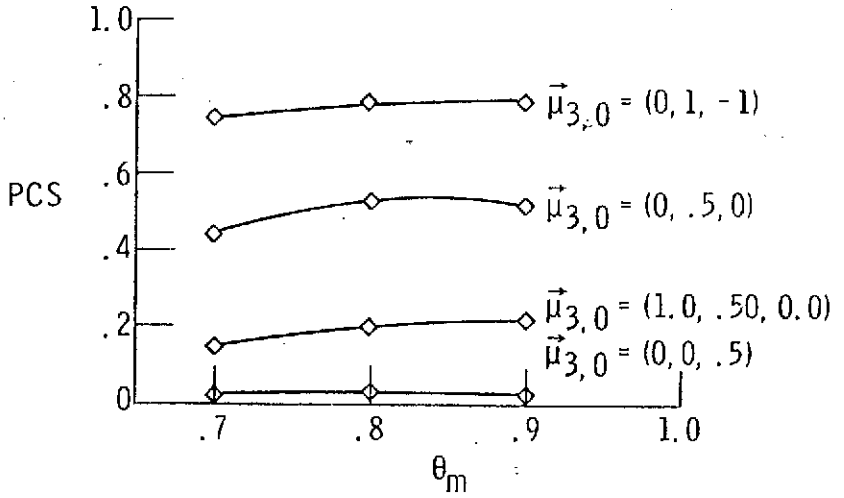


Figure 13. - Probability of correct selection (PCS) as a function of θ_m and $\vec{\mu}_{3,0}$ for $J_{max} = 8$ and H_3 true. Small sample performance simulation experiment three.

6.4 Discussion of Results

We now make some general observations concerning the results of the simulation experiments.

First, consider the large sample results. In the context of the fact that sequential procedures are primarily developed in the hope that reliable decisions can be made with small samples rather than large samples, these results are not of primary importance. It is interesting and informative to know, however, that the procedures are consistent. Since the study of limiting posterior distributions resulting from sequentially chosen experiments is known to be an extremely difficult and delicate problem, simulation experiments may be helpful by indicating to researchers what large sample behavior is likely to be true. In the problems studied in this paper it seems quite likely that when non-nested models are encountered, the posterior probability of the true hypothesis has a limiting value of unity. It also seems most likely that the limiting posterior mean of the true hypothesis does indeed equal the values of the unknown parameters generating the data.

When nested models are encountered, however, the results are not as enlightening. It appears that if the posterior probability of the correct hypothesis does not achieve a limit of unity, it at least attains a large value and then randomly fluctuates about that value. There is indication that the conjecture of Box and Hill that for these nested models there is a distinct preference by the

sequential procedure to choose the model with the smaller number of parameters as true. For instance, the polynomial study $L = 6$, $i^* = 5$ indicates that if a model with more parameters is true but can be approximated closely by one with fewer parameters, there is a preference for the smaller model.

In examining the small sample performance simulation experiments, it is seen that PCS drops off fairly rapidly as the distance of the prior mean of the correct model from the true values of the parameters increases. This supports the conjecture of Chernoff and Meeter et al. that there may often be "initial bungling." It should be noted, however, that in all cases studied, the prior means of the competing models were all set to be as close to the true model parameter values as could be done. Thus, in a sense, these experiments can be considered to be presenting the most unfavorable situation possible to the sequential procedure. In actual application it might be more reasonable to assume that the prior distributions of all the models are mis-specified to the same extent. This problem of "initial bungling" should also indicate that the statistician should have the prior precision matrices of the parameter distributions be as vague as the prior information permits.

One approach studied by Kiefer and Sacks (1963) was to plan small initial experiments as a basis for gaining information to plan a large second experiment. An alternative not studied in this dissertation, but which seems worthy of investigation, would be to set

a lower limit, say J_{MIN}, as the minimum number of observations taken before a stopping rule is applied. The sequential procedure would use the same rule as developed for selection of experiments but large posterior probabilities on the models would be ignored until a sufficient number of observations are taken to avoid the consequences of initial bungling. This also makes sense from the point of view of obtaining parameter estimates. Surely an experimenter would not be content to terminate sampling with two or three observations even if the resulting probabilities are overwhelmingly in favor of one hypothesis unless he had extremely good prior information.

The last topic to discuss is the modified stopping and selection procedure introduced in Chapter 5. This was not applied to any of the simulation experiments performed in this dissertation. The large sample simulation results indicate that when θ_m and/or J_{MAX} are large, then this modified procedure may be of value. For the problems considered here, it is seen that even for nested models, the unmodified procedure performs quite well when J_{MAX} is small.

CHAPTER 7

EXAMPLE OF APPLICATION

This chapter first presents a general outline of the situations in which the results of this paper may be applied. Following this an example from the literature is presented. The purpose of this example is to illustrate how the information available from previous experimentation can be translated to the information required for the application of the sequential procedure developed herein.

A Bayesian framework is used in this paper because in a great many applications there does exist some prior information which can be incorporated. The Bayesian approach to statistical inference is the most natural and satisfying method of incorporating prior knowledge. This prior knowledge may arise in several ways.

For example, when expensive or large experiments are contemplated, there is often available data from pilot studies, the literature of the field, or poorly designed prior experiments. Typically, some type of regression analysis is performed on this data but there is so little data that practically no conclusions can be drawn, only recommendations for further experimentation. The resulting equations, however, provide a very convenient starting point for the application of Bayesian methods.

In other situations, an experimenter has a great deal of ex-

perience in experiments that are similar and involve factors with which he has some previous experience. In these cases it may be safe to extrapolate his acquired knowledge from the similar but different experiments to the current experiment. If so, this may indicate some characteristics of the model equation.

A third possibility might arise for example in the carrying over of laboratory results to a production process or out-of-laboratory process. In the laboratory greater control can be exerted on many variables and typically only a small number of variables may be investigated. Often one or more mechanistic models are available. When the process is taken out of the laboratory, there will be less control over other variables and they must be accounted for by adding them to the model. Thus the experimenter is faced with the situation of having a partly mechanistic model and a partly empirical model. If the mechanistic model is sufficiently smooth in the region of interest, factorial or polynomial models can be applied in these cases and prior information might indicate which interactions or terms are most likely to exist.

The example we consider is studied in Lloyd and Lipow (1962) and Draper and Smith (1966). In these books the data presented in table 16 is used to illustrate some topics in the design of experiments and multiple linear regression analysis. The dependent variable y is the chamber pressure in rocket engines put on test. The four controlled variables are

z_1 = temperature of cycle (starting)

z_2 = vibration level

z_3 = shock by dropping (temperature)

z_4 = static fire temperature

We first postulate the model equation given by equation (7-1).

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 z_2 + \beta_4 z_2^2 + \beta_5 z_3 + \beta_6 z_3^2 + \beta_7 z_4 + \beta_8 z_1 z_2 + \beta_9 z_1 z_3 + \beta_{10} z_1 z_4 + \beta_{11} z_2 z_3 + \beta_{12} z_2 z_4 + \beta_{13} z_3 z_4 + \varepsilon \qquad (7\text{-}1)$$

The results of a multiple linear regression analysis of the model are summarized in table 17. The terms of the model are ordered in table 17 in decreasing order of descriptive significance level. The experiment is highly saturated with respect to equation (7-1) in the sense that 14 parameters are estimated from the data from 18 distinct combinations of levels of the independent variables. There are also quite a few high correlations among the terms of equation (7-1) and hence high correlations among the estimated parameters. The power of the resulting t-tests may be somewhat low under these circumstances. From examination of the various descriptive significance levels, the model equations tabulated in table 18a can be considered reasonable. The prior means of the distributions are also given in table 18b. The prior precision matrices may be derived by multiplying τ times the submatrices of order 3, 6, 9, and 15 of the matrix given in table 18c. How these prior distributions were determined is now described.

In multiple linear regression, under the usual normality

assumptions, the parameter estimates from the model

$$\vec{y} = M\vec{\beta} + \vec{\varepsilon}$$

are given by

$$\hat{\vec{\beta}} = (M'M)^{-1}M'\vec{y}$$

and we know

$$\hat{\vec{\beta}} \sim N\left(\vec{\beta}, \frac{1}{\sigma^2} M'M\right)$$

Thus, for the first three models of table 18, the prior means and precision matrices would reasonably be the $\hat{\vec{\beta}}$ and $\frac{1}{\sigma^2} M'M$ derived by least squares analysis using the appropriate subset of data from table 16. This is how the values of $\vec{\mu}_{\ell,0}$, $\Psi_{\ell,0}$ for $\ell = 1,2,3$ were derived. For $\ell = 4$ and the data of table 16, the full equation is not estimable because there are not three levels of z_4 to estimate a coefficient of z_4^2. Thus least squares estimates were computed and $M'M$ computed for the first 14 terms of model 4. Then an essentially diffuse prior was specified with respect to β_{14} by setting the prior mean to zero and adding the last row of the matrix in table 18c to $M'M$. The diagonal term was arbitrarily chosen to make the matrix nonsingular yet not comparable to any of the other diagonal elements in magnitude.

To complete the information required, τ must be specified and $\theta_{\ell,0}$ chosen. From the data of table 17 an unbiased estimate of σ^2 is 1.85 as computed from the replicated points. Thus we may use $\tau = \frac{1}{1.85} = 0.541$. To determine the $\theta_{\ell,0}$ it will be helpful to examine the F-ratios for lack-of-fit for the first three models in

table 18. These are $F = 5.295$, $F = 2.038$, and $F = 1.871$. These statistics are significantly large at about the 0.90, 0.80, and 0.75 levels, respectively. Based upon this, the following values of $\theta_{\ell,0}$ seem reasonable

$$\theta_{1,0} = 0.10$$

$$\theta_{2,0} = 0.30$$

$$\theta_{3,0} = 0.30$$

$$\theta_{4,0} = 0.30$$

REFERENCES

Albert, A. E. (1961): The sequential design of experiments for infinitely many states of nature. Ann. Math. Statist. 32: 774-799.

Bessler, S. (1960): Theory and Application of the Sequential Design of Experiments, k-actions and Infinitely Many Experiments. Part I - Theory. Part II - Applications. Technical reports No. 55 and No. 56. Applied Mathematics and Statistics Laboratories, Stanford University.

Box, G. E. P. and Hill, W. J. (1967): Discrimination Among Mechanistic Models. Technometrics 9: 57-72.

Chernoff, H. (1959): Sequential Design of Experiments. Annals Math. Statist. 30: 755-770.

Coveyou, R. R. and Macpherson, R. D. (1967): Fourier Analysis of Uniform Random Number Generators. J. Assoc. Computing Mach. 14: 100-119.

DeGroot, M. H. (1970): Optimal Statistical Decisions. McGraw-Hill N.Y.

Draper, N. R. and Smith, H. (1966): Applied Regression Analysis. John Wiley & Sons.

Fisz, M. (1963): Probability Theory and Mathematical Statistics. John Wiley & Sons.

Graybill, F. (1969): Introduction to Matrices with Applications in Statistics. Wadsworth Publ. Co.

Hunter, W. G. and Reiner, A. M. (1965): Designs for Discriminating Between Two Rival Models. Technometrics 7: 307-324.

Kiefer, J. and Sacks, J. (1963): Asymptotically Optimum Sequential Inference and Design. Ann. Math. Statist. 36: 705-750.

Kiefer, J. and Wolfowitz, J.: Optimum Designs in Regression Problems. Ann. Math. Statist. 30: 271-294.

Kullback, S. (1968): Information Theory and Statistics. Dover Publications. (Reprint of 1959 Wiley edition)

Lindley, D. V. (1956): On the Measure of the Information Provided by an Experiment. Ann. Math. Statist. 27: 986-1005.

Lloyd, D. K. and Lipow, M. (1962): Reliability: Management, Methods, and Mathematics. Prentice-Hall.

Meeter, D., Pirie, W. and Blot, W. (1970): A Comparison of Two Model-Discriminating Criteria. Technometrics, 12: 457-470.

Owen, R. J. (1970): The Optimum Design of a Two-Factor Experiment Using Prior Information. Ann. Math. Statist. 41: 1917-1934.

Rao, C. R. (1965): Linear Statistical Inference and Its Applications. Wiley & Sons.

Sidik, S. M. and Holms, A. G. (1971): Optimal Design Procedures for Two-Level Fractional Factorial Experiments Given Prior Information About Parameters. NASA TN D-6527.

Stone, M. (1959): Application of a Measure of Information to the Design and Comparison of Regression Experiments. Ann. Math. Statist. 30: 55-70.

Taussky, O. and Todd, J. (1956): Generating and Testing of Pseudo-Random Numbers. Publ. in Symposium on Monte-Carlo Methods. Wiley & Sons.

APPENDIX A

COMPUTER PROGRAM FOR SIMULATION STUDIES

The general flow of operations and computations performed by the program is described by the algorithm given in Chapter 6.1. The input required to perform these computations is first described. How these computations are achieved is described briefly by giving the major functions of the subprograms constituting the complete program. A complete FORTRAN listing is given.

INPUT

The program, as presented here, can only accommodate polynomial models over the interval [-1,+1] and two-level factorial models. This can be changed by writing one new subroutine (MFORM) to handle more general models. The program identifies the parameters by their integer subscripts and computes the x_{ij} values for the M-matrices according to the following convention:

1. For polynomial models, the subscript i indicates the parameter which is the coefficient of x^i.

2. For factorial models, the coefficients are assumed ordered in the standard order according to the description in Sidik and Holms (1971). The treatment combinations are also assumed to be in standard order and the independent variable values are constructed as described in Sidik and Holms (1971).

3. The order of the models as specified for input are written

such that the parameter subscripts are in increasing order.

The specific input cards are now described below and illustrated by the input for a case run in Chapter 6.3.1. The problem input is specified by the models and parameter values:

$$H_1: \quad y = \beta_1 x_1 + \varepsilon$$

$$H_2: \quad y = \beta_2 x_2 + \varepsilon$$

$$H_3: \quad y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\left.\begin{aligned} \Psi_{\ell,0} &= I \\ \theta_{\ell,0} &= \frac{1}{3} \end{aligned}\right\} \ell = 1,2,3 \qquad \begin{aligned} \theta_m &= 0.70 \\ \tau &= 0.5 \end{aligned}$$

$$J_{MAX} = 8$$

1500 simulations

$$\vec{\mu}_{1,0} = (1.0)$$

$$\vec{\mu}_{2,0} = (1.0)$$

$$\vec{\mu}_{3,0} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \qquad \vec{\mu}^* = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$$

Start random number generator with 041 574 501 221.

1. IDENTIFICATION (13A6)

   One card for Hollerith input description of the problem.

2. NAMELIST INPUT ($NAML)

   Most of the control parameters are included in a NAMELIST input set. The list of parameters and their purpose follows:

NFAC — For factorial problems, this supplies the number of factors. For polynomial problems, this variable need not be supplied.

NHYP — Number of hypothesized models.

NSYM — Number of simulations to run.

MXTRTS — J_{MAX}

NTREND — Not used for this report. Set to zero.

LEVOUT — An output control variable. Certain basic output is automatically printed. Extra intermediate output can be printed by setting LEVOUT to an integer between 1 and 7. For performance studies set to 0. For large sample runs set to 2. For debugging set to 7.

LTRUE — Supplies i^*, the correct model subscript.

TAU — τ

CSTOP — θ_m

TMAX — An upper limit on execution time. If this limit is exceeded, the program dumps for a restart.

IFSTRT — Set T for supplied starting value for random number sequence. Set F if sequence is to start with initialization value. (See description of subroutines RAND and SAND for further information.)

RESTRT — Is this problem a restart of a case terminated by exceeded time? T or F.

POLY T implies polynomial model. F implies factorial model

NX For polynomial models, the x space is restricted to the interval -1 to +1. NX specifies the number of points used to discretize the interval into equal increments.

3. FORMAT FOR PRECISION MATRICES (13A6, A2)

For each set of model equations supply one set of 4A, 4B, 4C, 4D, and 4E.

4A. NUMBER OF PARAMETERS IN MODEL, PRIOR PROBABILITY OF MODEL (I6, F12.6)

4B. PARAMETER SUBSCRIPTS (13I6)

4C. PRIOR MEANS (12F6.0)

4D. TRUE VALUE OF PARAMETERS (12F6.0).

This card should be supplied only for the set corresponding to the correct model.

4E. PRIOR PRECISION MATRIX

Only the upper triangular half of Ψ is specified with the order being

$$\Psi_{11}, \Psi_{12}, \Psi_{22}, \Psi_{13}, \text{ etc.}$$

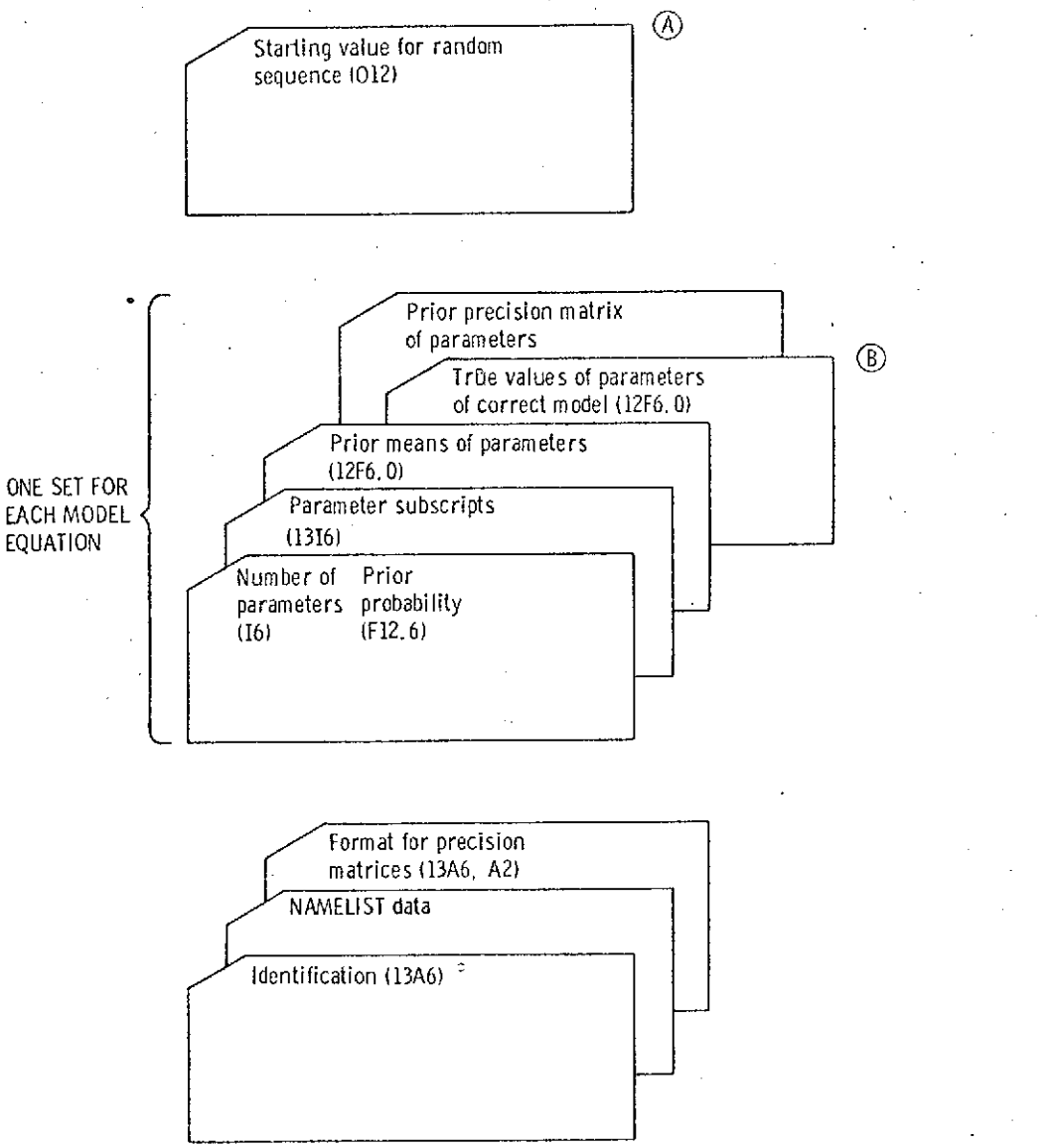
5. STARTING VALUE FOR PSEUDO-RANDOM SEQUENCE (O12)

A graphical illustration of a data deck is given in figure A1 and a FORTRAN sheet giving the sample input is given in table A1.

TITLE: TABLE A1 | PROJECT NUMBER: | ANALYST: | SHEET ____ OF ____

STATEMENT NUMBER	CONT	FORTRAN STATEMENT	IDENTIFICATION
		SMALL SAMPLE PERFORMANCE STUDY FOR THREE MODELS	
		$NAML NFAC=2, NHYP=3, NSYM=1500,	
		MXTRTS=8, NTREND=0, LEVOUT=0,	
		LTRUE=3, TAU=.5, CSTOP=.70,	
		TMAX=10.0, IFSTRT=T, RESTRT=F,	
		POLY=F, NX=0 $	
		(3F1.0)	
	1	.33333333	
	1		
		1.0 0 0	
		1	
	1	.33333333	
	2		
		1.0	
		1	
	2	.33333333	
	1	2	
		1.0 1.0	
		101	
		04157 4 501221	

NASA-C-836 (REV. 9-14-59)



A - This card is optional and its presence depends upon NAMELIST data.

B - Supplied only for true model.

Figure A-1. - Graphical illustration of card input for simulation program.

SUBPROGRAMS AND THEIR MAJOR FUNCTIONS

1. SEQDES. This is the main program. It reads the input cards, exercises general control over the other subroutines. It also outputs the final results.

2. ACOMB. Called once by SEQDES at the beginning of each case. This subroutine scans the lists of parameter subscripts for each model and constructs from them a new list of subscripts in ascending order which contains the parameters appearing in at least one model.

3. SVSTRT (GTSTRT). A double entry subprogram. The entry SVSTRT is used once at the beginning of each case to save the prior probabilities, means, and precision matrices. Then after each simulation, entry GTSTRT is used to re-initialize the working probability, mean, and precision vectors to the original values.

4. MFORM. There are two versions, MFORMX and MPOLY. The first is for factorial runs and the second for polynomial models. These subroutines construct the M-matrices required for each model when given the experiment choice (a level of x for polynomials, a treatment combination for factorials). It is called by INFX and YGEN.

5. INFX (UPDATE). Entry INFX accepts the current state of sampling as defined by $\{\theta_{\ell,j}, \vec{\mu}_{\ell,j}, \Psi_{\ell,j}; \ell = 1,L\}$ and determines the expected K-L information for a specified experiment. Repeated calls then are made by SEQDES to determine the optimal experiment. Entry UPDATE is called by YGEN and accepts the current state values

and an observed random response to compute the posterior distributions.

6. RANDUM. Called once for each random observation taken. Generates a random observation, ε, from $N(0,\tau)$.

7. YGEN. Called once for each random observation. This routine accepts the $N(0,\tau)$ variate generated by RANDUM and calls MFORM to compute M, where M is the design matrix appropriate for the correct model and the experiment chosen as optimal. Then

$$y = M\vec{\mu}^{*} + \varepsilon$$

is computed. After generating y, posterior values for θ_ℓ are computed. Sampling for that simulation trial is then terminated if any θ_ℓ exceeds θ_m. If sampling is not terminated, YGEN calls UPDATE to compute the posterior $\vec{\mu}_\ell(y)$ and $\Psi_\ell(y)$, and the sampling procedure is continued.

8. COUNTX. This subroutine is called by YGEN whenever a $\theta_\ell \geq \theta_m$ or by SEQDES whenever J_{MAX} observations have been taken. It counts the number of times each model is chosen and records the distribution of sample sizes. These are then output after all simulations have been computed.

9. INVXTX. Inverts a symmetric matrix stored in the lower symmetric storage mode.

10. TRIOUX. Outputs a lower triangular matrix.

11. MTVEC. Computes the product vector resulting from the multiplication of a vector and a symmetric matrix.

12. RAND (SAND). This pair of entries provide the pseudo-random sequence of uniform random variates. SAND is an initializing entry which must be called before any calls to RAND. When RAND is called, it computes the next value in the random sequence from the current value. The return argument of RAND is the floating point uniform random variate. The input and return argument of SAND is an argument which saves the integer value of the random variate. The generator is the multiplicative-congruential type obtained by taking the low order 36 bits of the product $r_{r-1} \times k$ where

r_{r-1} is the previous random number

r_0 is 1 and

k is 5^{15}.

The properties of this generator are discussed by Taussky and Todd (1956) and Coveyou and Macpherson (1967).

13. BCREAD (X1,X2). BCDUMP (X1,X2). These routines, respectively, read and punch cards in absolute binary at the rate of 22 words per card. The data read or punched begins at the location in core of the variable X1 and ends at the location in core of the variable X2.

```
0000100       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT,PERMT(4)
0000200       COMMON/XRAND/Y,EPS,UNIF,U
0000300       COMMON/ERR/TAU,SIGMA
0000400       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)      -
0000500      X      ,NPARAM(10),MUD(11),PMU(1000)
0000600       COMMON/PS/PINDEX(11),PREC(2000)
0000700       INTEGER PINDEX
0000800       COMMON/CNTRLS/NHYP,NALL,NFAC,NFULL,NFULM1,TRTMNT,NTREND,NOBS,LTRUE
0000900       INTEGER TRTMNT
0001000       COMMON/PREMNC/NAVG(1000),PCS,XMSE,XMST,INITAL ,IHCNT(10),MSOFAR
0001100       COMMON/INF/XINF
0001200       COMMON/PROBS/THETA(10),CSTOP
0001300 C
0001400 C
0001500 C
0001600       COMMON/MM/OESM(1000)
0001700       COMMON/YDIST/S(10),R(10),SV(10),RV(10)
0001800       COMMON/XTRA/AA(1000),B(2000),C(2000),D(10)
0001900 C
0002000 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0002100 C
0002200       DATA ENDCRD/4HENDCRD/
0002300       REAL IDENT
0002400       EQUIVALENCE (X1,IX),(KS,SK)
0002500       DIMENSION IDENT(14)
0002600       DIMENSION INPERMT(14)
0002700       LOGICAL IFSTRT,RESTRT
0002800       LOGICAL KPGEN
0002900       LOGICAL POLY
0003000       NAMELIST/NAML/NFAC,NHYP,NSYM,MXTRTS,NTREND,LEVOUT,LTRUE,TAU,      -
0003100      X  CSTOP,TMAX,IFSTRT,RESTRT                                        -
0003200      X ,POLY,NX                                                         -
0003400 C
0003500 C*************************************************************************
0003600 C
0003700       KPGEN=.FALSE.
0003800       IGOTO=1
0003900       CALL TIME1(TSTRT)
0004000     1 READ(IUNIN,5040) IDENT
0004100       IF(IDENT(1).EQ.ENDCRD) STOP
0004200       WRITE(IUNOUT,6010) IDENT
0004300 C
0004400       READ(IUNIN,NAML)
0004500       WRITE(IUNOUT,NAML)
0004600       NFULL=2**NFAC
0004700       NFULM1=NFULL-1
0004800       IF(POLY) NFULL=NX
0004900       SIGMA= 1.0/SQRT(TAU)
0005000       READ(IUNIN,5040) INPERMT
0005100       MUD(1)=1
0005200       PINDEX(1)=1
0005300       DO 190 N= 1,NHYP
0005350       IHCNT(N)=0
0005400       WRITE(IUNOUT,6365)
0005500       READ(IUNIN,5020) NPARAM(N),THETA(N)
0005700       MUD(N+1)=MUD(N)+ NPARAM(N)
0005800       PINDEX(N+1)=PINDEX(N)+ (NPARAM(N)*(NPARAM(N)+1))/2
0005900       MULO = MUD(N)
0006000       MUHI= MUD(N+1)-1
```

```
0006100       IPLO= PINDEX(N)
0006200       IPHI= PINDEX(N+1)-1
0006300       READ(IUNIN,5010)(ALPH(L),L=MULO,MUHI)
0006400       READ(IUNIN,5050)(ALPHMU(L),L=MULO,MUHI)
0006500       IF (N.NE.LTRUE)  GO TO 100
0006600       NR = NPARAM(N)
0006700       WRITE(IUNOUT,6360)
0006800       READ(IUNIN,5050) (REALMU(L),L=1,NR)
0006900       WRITE (IUNOUT,6355) (REALMU(L),L=1,NR)
0007000       WRITE (IUNOUT,6360)
0007100   100 CONTINUE
0007900       READ(IUNIN,INPFMT)(PREC(L),L=IPLO,IPHI)
0008000       WRITE(IUNOUT,6050) N,NPARAM(N),THETA(N)
0008200       WRITE(IUNOUT,6060)(ALPH(L),ALPHMU(L),L=MULO,MUHI)
0008300       WRITE(IUNOUT,6090)
0008400       CALL TRIANS(PREC(IPLO),NPARAM(N),B,PFMT,ALPH(MULO))
0008500 C
0008600       CALL MTVEC(PREC(IPLO),ALPHMU(MULO),NPARAM(N),PMU(MULO))
0008700       IF(LEVOUT.GE.5)WRITE(IUNOUT,6060)(ALPH(L),PMU(L),L=MULO,MUHI)
0008800   190 CONTINUE
0008900 C
0009000 C
0009100 C
0009200       M = NPARAM(LTRUE)
0009300       JJ = MUD(LTRUE)-1
0009400       XMSE=0.0
0009500       DO 195 J=1,M
0009600       JJ = JJ+1
0009700   195 XMSE = XMSE + (ALPHMU(JJ)-REALMU(J))**2
0009800       XMST = SQRT(XMSE*TAU)
0009900       XMSE = SQRT(XMSE)
0010000       WRITE (IUNOUT,6100) XMSE,XMST
0010100       XMSE = 0.0
0010200       XMST = 0.0
0010300       DO 197  I=1,MXTRTS
0010400   197 NAVG(I) = 0
0010500       PCS = 0.0
0010600       CALL SAND(INITAL)
0010700       IF(KPG(N) INITAL=NTLSV
0010800       IF(IFSTRT) READ(IUNIN,5090) INITAL
0010900       IF(.NOT.RESTRT) GO TO 200
0011000       WRITE(IUNOUT,6425)
0011100       CALL BCREAD(NAVG(1),MSOFAR)
0011200 C
0011300 C
0011400 C
0011500 C*****************************************************************************
0011600 C
0011700   200 CONTINUE
0011800        WRITE(IUNOUT,6500) INITAL
0011900  6500 FORMAT(26H STARTING VALUE FOR RAND= O12)
0012000       CALL ACOMB
0012100       CALL SVSTRT
0012200 C
0012300 C*****************************************************************************
0012400 C
0012500       MMM=1
0012600       IF(RESTRT) MMM=MSOFAR+1
0012700       DO 800 M=MMM,NSYM
```

```
0012800         MSOFAR=M
0012900         CALL GTSTRT
0013000         DO 700  NTR=1,MXTRTS
0013100         NORS= NTR
0013200         SINF= 0.0
0013300         DO 600  ITRT=1,NFULL
0013400         TRTMNT= ITRT-1
0013500         CALL MFORM
0013600         CALL INFMTN
0013700         IF(XINF.LE.SINF) GO TO 600
0013800         SINF= XINF
0013900         ISV= TRTMNT
0014000         DO 590 N=1,NHYP
0014100         SV(N)=S(N)
0014200         RV(N)=R(N)
0014300     590 CONTINUE
0014400     600 CONTINUE
0014500         TRTMNT= ISV
0014600         IF (LEVOUT.GE.3) WRITE(IUNOUT,6210) ISV,SINF
0014700         CALL RNDM
0014800         CALL YGEN(1750)
0014900     700 CONTINUE
0015000     750 CALL COUNT
0015100         CALL TIMER(TNOW)
0015200         TPRNT=(TNOW-TSTRT)/3600.
0015300         IF(TPRNT.LT.TMAX) GO TO 800
0015400         IGOTO=2
0015500         GO TO 810
0015600     800 CONTINUE
0015700 C
0015800 C
0015900 C
0016000     810 CONTINUE
0016100         WRITE(IUNOUT,6240) TPRNT
0016200         XXMSE=XMSE/FLOAT(MSOFAR)
0016300         XXMST=XXMSE*SQRT(TAU)
0016400         ASN=0.0
0016500         DO 850 I=1,MXTRTS
0016600     850 ASN=ASN+FLOAT(NAVG(I)*I)
0016700         ASN=ASN/FLOAT(MSOFAR)
0016800         WRITE(IUNOUT,6400) (NAVG(I),I=1,MXTRTS)
0016810         WRITE(IUNOUT,6600) (IHCNT(I),I=1,NHYP)
0016820 6600    FORMAT(1H 10I10)
0016900  6400 FORMAT(1H 10I12)
0017000         PPCS=PCS/FLOAT(MSOFAR)
0017100         WRITE(IUNOUT,6300) ASN,PPCS,XXMSE,XXMST,INITAL
0017200         NTLSV=INITAL
0017300         KPGEN=.TRUE.
0017400         GO TO (1,1000), IGOTO
0017500    1000 CALL RCOUMP(NAVG(1),MSOFAR,0)
0017600         WRITE(IUNOUT,6450) MSOFAR
0017700         STOP
0017800 C
0017900 C*******************************************************************************
0018000 C
0018100    5010 FORMAT(13I6)
0018200    5020 FORMAT(I6,3F12.6,10I1)
0018300    5040 FORMAT(13A6,A2)
0018400    5050 FORMAT(12F6.0)
```

```
0018500  5080 FORMAT(6I6,3F6.0,2L1)
0018600  5090 FORMAT(012)
0018700 C
0018800 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0018900  6010 FORMAT(1H1,13A6,1A2)
0019000  6050 FORMAT(11HFOR MODEL I3/16H NO. PARAMETERS I5/13H PRIOR PROB. G14.-
0019100      X5)
0019300  6060 FORMAT(51HKTHE PARAMETERS IN THE MODEL AND THEIR MEANS ARE-- //   -
0019400      X (( I10,G14.5 ))
0019500  6090 FORMAT(44HKTHE PRECISION MATRIX OF THE PARAMETERS IS--)
0019600  6100 FORMAT( 49HKINITIAL DISTANCE OF PRIOR MEAN FROM TRUE VALUE  G14.5/-
0019700      X   18X,16HDIVIDED BY SIGMA,15X,G14.5)
0019800  6210 FORMAT(1H 5G18.8)
0019900  6240 FORMAT(24H CURRENT EXECUTION TIME  F10.3)
0020000  6300 FORMAT(6HKASN= G14.2/6H PCS= G14.6/12H AVG.DIST.= G14.6/          -
0020100      X  14H NORMALIZED = G14.6/2H* 012)
0020200  6355 FORMAT(1H 8G16.8)
0020300  6360 FORMAT(1H 40(2H +))
0020400  6365 FORMAT(1H 80(1H*))
0020500  6425 FORMAT( 37H THIS IS A RESTART OF A PREVIOUS CASE)
0020600  6450 FORMAT(55H THIS CASE WAS TERMINATED BY CLOCK. DUMPING FOR RESTART/-
0020700      X ,29H NO. SIMULATIONS COMPLETED =  I6)
0020800       END
```

```
0000100       SUBROUTINE ACOMB
0000200       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT,PRMT(4)
0000300       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)      -
0000400      X      ,NPARAM(10),MUD(11),PMU(1000)
0000500       COMMON/CNTRLS/NHYP,NALL,NFAC,NFULL,NFULMI,TRTMNT,NTREND,NOBS,LTRUE
0000600 C
0000700 C*************************************************************************
0000800 C
0000900       NALL = NPARAM(1)-NTREND
0001000       IF(NALL.LE.0) GO TO 110
0001100       DO 100 K =1,NALL
0001200       IIALPH(K)= IALPH(K)
0001300   100 CONTINUE
0001400   110 IF(NHYP.LE.1) RETURN
0001500 C
0001600       DO 800  N=2,NHYP
0001700       MAXI= MUD(N+1)-1-NTREND
0001800       KK=0
0001900       K=MUD(N)-1
0002000   150 KK=KK+1
0002100   200 K=K+1
0002200       IF(KK.GT.NALL) GO TO 420
0002300       IF(K.GT.MAXI) GO TO 800
0002400   250 IF(IALPH(K)-IIALPH(KK))300,150,400
0002500   300 KS=NALL+2
0002600       DO 350 J=KK,NALL
0002700       KS=KS-1
0002800   350 IIALPH(KS)=IIALPH(KS-1)
0002900       IIALPH(KK)=IALPH(K)
0003000       NALL=NALL+1
```

```
0003100       GO TO 150
0003200   400 KK=KK+1
0003300       IF(KK-NALL) 250,250,450
0003400   420 IF(K.GT.MAXI) GO TO 800
0003500   450 NMORE=MAXI-K+1
0003600       KO=K-1
0003700       DO 500  J=1,NMORE
0003800       KS1=NALL +J
0003900       KS2=KO+J
0004000       IIALPH(KS1) = IALPH(KS2)
0004100   500 CONTINUE
0004200       NALL=NALL+NMORE
0004300   800 CONTINUE
0004400       IF(LEVOUT.GE.7)WRITE(IUNOUT,1000) NALL,(IIALPH(I),I=1,NALL)
0004500  1000 FORMAT(1H (10/(1H 25I5))
0004600       RETURN
0004700       END
```

```
0000100       SUBROUTINE SVSTRT
0000200       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)
0000300      X      ,NPARAM(10),MUD(11),PMU(1000)
0000400       COMMON/PS/PINDEX(11),PREC(2000)
0000500       INTEGER PINDEX
0000600       COMMON/CNTRLS/NHYP,NALL,NEAC,NFULL,NFULM1,TRTMNT,NTREND,NOBS,LTRUE
0000700       COMMON/PROBS/THETA(10),CSTOP
0000800       DIMENSION XLPHMU(1000),XPREC(2000),XTHET(10),XPMU(1000)
0000900       M= MUD(NHYP+1)-1
0001000       DO 20 J=1,M
0001100       XLPHMU(J) = ALPHMU(J)
0001200       XPMU(J)=PMU(J)
0001300    20 CONTINUE
0001400       M = PINDEX(NHYP+1)-1
0001500       DO 40 J=1,M
0001600       XPREC(J) = PREC(J)
0001700    40 CONTINUE
0001800       DO 60 J=1,NHYP
0001900       XTHET(J) = THETA(J)
0002000    60 CONTINUE
0002100       RETURN
0002200       ENTRY GTSTRT
0002300       M= MUD(NHYP+1)-1
0002400       DO 120 J=1,M
0002500       ALPHMU(J)=XLPHMU(J)
0002600       PMU(J)=XPMU(J)
0002700   120 CONTINUE
0002800       M=PINDEX(NHYP+1)-1
0002900       DO 140 J=1,M
0003000       PREC(J)= XPREC(J)
0003100   140 CONTINUE
0003200       DO 160 J=1,NHYP
0003300       THETA(J) = XTHET(J)
0003400   160 CONTINUE
0003500       RETURN
0003600       END
```

```
0000100 C   THIS IS FOR TWO-LEVEL FACTORIALS ONLY
0000150       SUBROUTINE MFORM
0000200       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT,PERMT(4)
0000300       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)
0000400      X      ,NPARAM(10),MUD(11),PMU(1000)
0000500       INTEGER TRTMNT
0000600       COMMON/CNTRLS/NHYP,NALL,NFAC,NFULL,NFULMI,TRTMNT,NTREND,NOBS,LTRUE
0000700       COMMON/MM/DESM(1000)
0000800 C
0000900 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0001000 C
0001100       DIMENSION LASTA(10)
0001200       EQUIVALENCE (KS,SK),(IX,XI)
0001300 C*************************************************************************
0001400 C
0001500       DO 5 N=1,NHYP
0001600       LASTA(N)=0
0001700     5 CONTINUE
0001800 C
0001900       DO 1000  I=1,NALL
0002000       IPARAM = IIALPH(I)
0002100       IF(IPARAM.NE.0) GO TO 40
0002200       A= +1.0
0002300       GO TO 500
0002400 C
0002500    40 CONTINUE
0002600       ITR=TRTMNT
0002700       DX= 1.0
0002800       DO 150  J=1,NFAC
0002900       XI=AND(MASK,IPARAM)
0003000       IX= IX+1
0003100       GO TO (130,100), IX
0003200   100 SK=AND(MASK,ITR)
0003300       KS= KS+1
0003400       GO TO (110,130), KS
0003500   110 DX= -DX
0003600   130 IPARAM= IPARAM/2
0003700       ITR=ITR/2
0003800   150 CONTINUE
0003900       A= DX
0004000 C
0004100 C
0004200   500 CONTINUE
0004300       IF(LEVOUT.GE.7) WRITE(IUNOUT,6000) IIALPH(I),A
0004400 C
0004500       DO 950  K=1,NHYP
0004600       IF(LASTA(K)-NPARAM(K) + NTREND)  520,950,950
0004700   520 IX= MUD(K)+LASTA(K)
0004800       IF(IIALPH(I) - IALPH(IX)) 950,550,9999
0004900   550 CONTINUE
0005000       LASTA(K)= LASTA(K) +1
0005100       DESM(IX)= A
0005200   950 CONTINUE
0005300  1000 CONTINUE
0005400       IF(NTREND.LE.0) GO TO 1060
0005500       A= NOBS
0005600       B= 1.0E0
0005700       DO 1050  J=1,NTREND
0005800       B= B*A
```

```
0005900       KS= J-NTREND
0006000       DO 1040  K=1,NHYP
0006100       KS=KS+NPARAM(K)
0006200       DESM(KS)= B
0006300  1040 CONTINUE
0006400  1050 CONTINUE
0006500  1060 CONTINUE
0006600       ITR=MUD(NHYP+1)-1
0006700       IF (LEVOUT.GE.7) WRITE(IUNOUT,6010) (DESM(I),I=1,ITR)
0006800       RETURN
0006900 C
0007000  9999 CALL EXIT
0007100       RETURN
0007200  6000 FORMAT(1H I6,F6.0)
0007300  6010 FORMAT(6H DESM=  /(6X,15F8.0))
0007400 C
0007500       END
```

```
0000100 C THIS IS FOR POLYNOMIALS ONLY
0000150       SUBROUTINE MEORM
0000200       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT,PFMT(4)
0000300       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)
0000400      X      ,NPARAM(10),MUD(11),PMU(1000)
0000500       INTEGER TRTMNT
0000600       COMMON/CNTRLS/NHYP,NALL,NEAC,NFULL,NFULM1,TRTMNT,NTREND,NOBS,LTRUE
0000700       COMMON/MM/DESM(1000)
0000800 C
0000900 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0001000 C
0001100       DIMENSION LASTA(10)
0001200       EQUIVALENCE (KS,SK),(IX,XI)
0001300 C*************************************************************************
0001400 C
0001500       DX=2.0/FLOAT(NFULL-1)
0001600       X=-1.0+FLOAT(TRTMNT)*DX
0001650       IF(ABS(X).LE.1.0E-4) X=0.0
0001700       DO 5 N=1,NHYP
0001800       LASTA(N)=0
0001900     5 CONTINUE
0002000 C
0002100       DO 1000  I=1,NALL
0002200       IPARAM = IIALPH(I)
0002300       A=X**IPARAM
0002400 C
0002500 C
0002600   500 CONTINUE
0002700       IF(LEVOUT.GE.7) WRITE(IUNOUT,6000) IIALPH(I),A
0002800 C
0002900       DO 950  K=1,NHYP
0003000       IF(LASTA(K)-NPARAM(K) + NTREND)  520,950,950
0003100   520 IX= MUD(K)+LASTA(K)
0003200       IF(IIALPH(I) - IALPH(IX)) 950,550,9999
0003300   550 CONTINUE
0003400       LASTA(K)= LASTA(K) +1
```

```
0003500       DESM(IX)= A
0003600   950 CONTINUE
0003700  1000 CONTINUE
0003800       IF(NTREND.LE.0) GO TO 1060
0003900       A= NOBS
0004000       B= 1.0E0
0004100       DO 1050  J=1,NTREND
0004200       B= B*A
0004300       KS= J-NTREND
0004400       DO 1040  K=1,NHYP
0004500       KS=KS+NPARAM(K)
0004600       DESM(KS)= B
0004700  1040 CONTINUE
0004800  1050 CONTINUE
0004900  1060 CONTINUE
0005000       ITR=MUD(NHYP+1)-1
0005100       IF (LEVOUT.GE.7) WRITE(IUNOUT,6010) (DESM(I),I=1,ITR)
0005200       RETURN
0005300 C
0005400  9999 CALL EXIT
0005500       RETURN
0005600  6000 FORMAT(1H I6,F6.0)
0005700  6010 FORMAT(6H DESM=  /(6X,15F8.0))
0005800 C
0005900       END
```

```
0000100       SUBROUTINE INFMTN
0000200       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT,PERT(4)
0000300       COMMON/CNTRLS/NHYP,NALL,NEAC,NFULL,NFULM1,TRTMNT,NTREND,NOBS,LTRUE
0000400       INTEGER TRTMNT
0000500       COMMON/XRAND/Y,EPS,UNIF,U
0000600       COMMON/ERR/TAU,SIGMA
0000700       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)
0000800      X      ,NPARAM(10),MUD(11),PMU(1000)
0000900       COMMON/PS/PINDEX(11),PREC(2000)
0001000       INTEGER PINDEX
0001100       COMMON /INF/XINF
0001200       COMMON/PROBS/THETA(10),CSTOP
0001300       COMMON/MM/DESM(1000)
0001400       COMMON/YDIST/ S(10),R(10),SV(10),RV(10)
0001500       COMMON/XTRA/AA(1000),B(2000),C(2000),D(10)
0001600       DIMENSION DESMU(10),TRACE(10)
0001700       LOGICAL UPDT
0001800 C
0001900 C*******************************************************************************
0002000 C
0002100       GO TO 5
0002200       ENTRY UPDATE
0002300       UPDT=.TRUE.
0002400       GO TO 6
0002500     5 UPDT=.FALSE.
0002600     6 CONTINUE
0002700       DO 10  K=1,NHYP
0002800       DESMU(K)=0.0
```

```
0002900    10 CONTINUE
0003000 C
0003100 C*****************************************************************************
0003200 C
0003300       DO 520 N=1,NHYP
0003400       IF(LEVOUT.GE.5) WRITE(IUNOUT,6005) N,NOBS
0003500       MS= MUD(N)
0003600       JE= NPARAM(N)
0003700 C
0003800 C
0003900       IB=0
0004000       DO 100  J=1,JE
0004100       KS1=MS+J-1
0004200       IF(UPDT) GO TO 25
0004300       DESMU(N)=DESMU(N)+DESM(KS1)*ALPHMU(KS1)
0004400    25 CONTINUE
0004500       DO 100 JJ=1,J
0004600       IB= IB+1
0004700       KS2 = MS+JJ-1
0004800       B(IB)= DESM(KS1)*DESM(KS2)
0004900   100 CONTINUE
0005000 C
0005100 C*****************************************************************************
0005200 C
0005300       IF(LEVOUT.GE.6) CALL TRIANG(B,JE,8,PFMT,IALPH(MS))
0005400       KS1=PINDEX(N+1)-PINDEX(N)
0005500       KS2= PINDEX(N)
0005600       DO 120  K=1,KS1
0005700       B(K)= TAU*B(K)+PREC(KS2)
0005800       C(K)=B(K)
0005900       KS2= KS2+1
0006000   120 CONTINUE
0006100       CALL INVXTX(B,JE)
0006200       IF(LEVOUT.GE.6) CALL TRIANG(B,JE,8,PFMT,IALPH(MS))
0006300 C
0006400 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0006500       IF(.NOT.UPDT) GO TO 199
0006600       KS1=PINDEX(N+1)
0006700       IPLO = PINDEX(N)
0006800       KS1=KS1-IPLO
0006900       MUH1=MUD(N+1)-1
0007000       KS2=IPLO
0007100       DO 130 J=1,KS1
0007200       PREC(KS2)=C(J)
0007300        KS2=KS2+1
0007400   130 CONTINUE
0007500       KS1=MS-1
0007600       DO 150  J=1,JE
0007700       KS1= KS1+1
0007800       AA(J)= Y*DESM(KS1)*TAU + PMU(KS1)
0007900   150 CONTINUE
0008000       CALL MTVEC(B,AA,JE,ALPHMU(MS))
0008100       CALL MTVEC(PREC(IPLO),ALPHMU(MS),JE,PMU(MS))
0008200       IF(LEVOUT.LT.3) GO TO 520
0008300       IF(LEVOUT.GE.5) GO TO 160
0008400       WRITE(IUNOUT,6020)(ALPHMU(L),L=MS,MUH1)
0008500       IF(LEVOUT.LT.4) GO TO 520
0008600       WRITE(IUNOUT,6090)
0008700       CALL TRIANG(PREC(IPLO),JE,8,PFMT,IALPH(MS))
```

```
0008800        GO TO 520
0008900    160 CONTINUE
0009000        WRITE(IUNOUT,6060)(ALPH(L),ALPHMU(L),PMU(L),L=MS,MUHT)
0009100        WRITE(IUNOUT,6090)
0009200        CALL TRIANG(PREC(IPLO),JF,8,PFMT,IALPH(MS))
0009300        GO TO 520
0009400 C
0009500 C      ***************************************************************************
0009600 C
0009700 C
0009800    199 CALL MTVEC(B,DESM(MS),JE,AA)
0009900        DO 300  KS2=1,JE
0010000    280 AA(KS2)= TAU*AA(KS2)
0010100    300 CONTINUE
0010200        IF(LEVOUT.GE.5) WRITE(IUNOUT,6010)(AA(J),J=1,JE)
0010300 C
0010400 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0010500 C
0010600        R(N)= 0.0
0010700        MMS=MS-1
0010800        DO 360  J=1,JE
0010900        MMS= MMS+1
0011000        R(N)= R(N) + AA(J)*DESM(MMS)
0011100    360 CONTINUE
0011200        R(N)= TAU*(1.0-R(N))
0011300 C
0011400 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0011500 C
0011600        MMS=MS
0011700        C=0.0
0011800        DO 500  J=1,JE
0011900        C=C+AA(J)*PMU(MMS)
0012000        MMS= MMS+1
0012100    500 CONTINUE
0012200        S(N)=C/R(N)
0012300        IF(LEVOUT.GE.5) WRITE(IUNOUT,6020) R(N),S(N)
0012400    520 CONTINUE
0012500 C
0012600        IF(UPDT) RETURN
0012700 C**  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **  **
0012800 C
0012900        DO 1000 N=1,NHYP
0013000        C= 0.0
0013100        DO 580  J=1,NHYP
0013200        IF(N.EQ.J)  GO TO 580
0013300        C=C+THETA(J) *R(J)
0013400    580 CONTINUE
0013500        TRACE(N)=C/R(N)
0013600        IF(LEVOUT.GE.5) WRITE(IUNOUT,6035) TRACE(N)
0013700   1000 CONTINUE
0013800 C
0013900 C***************************************************************************
0014000 C
0014100        XINF = THETA(1) * TRACE(1)
0014200        DO 1500  N=2,NHYP
0014300        JE= N-1
0014400        XINF = XINF + THETA(N) * TRACE(N)
0014500        DO 1450  J=1,JE
0014600        C= S(N) - S(J)
```

```
0014700       B=C*C*(R(NI+R(J))
0014800       XINF = XINF + THETA(N) *THETA(J) *R
0014900  1450 CONTINUE
0015000  1500 CONTINUE
0015100       IF (LEVOUT.GE.5) WRITE(IUNOUT,6020) XINF
0015200       RETURN
0015300 C
0015400 C     ******************************************************************
0015500 C
0015600 C
0015700 C*************************************************************************
0015800 C     FORMATS
0015900 C
0016000  6005 FORMAT(3H N= I6,7H NOBS= I6)
0016100  6010 FORMAT(2H A/(1H 10G12.4))
0016800  6020 FORMAT(1H 10G12.4)
0016300  6035 FORMAT(7H TRACE= G12.4)
0016400  6060 FORMAT(51HK THE PARAMETERS IN THE MODEL, AND THEIR MEANS ARE--   //-
0016500     X  ((1H I10,2G14.5)))
0016600  6090 FORMAT(44HKTHE PRECISION MATRIX OF THE PARAMETERS IS--   )
0016700       END
```

```
0000100       SUBROUTINE RNDM
0000200       COMMON/XRAND/ Y,EPS,UNIF,U
0000300       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT
0000400       COMMON/ERR/TAU,SIGMA
0000500       DATA A0/2.515517/,A1/.802853/,A2/.010328/,
0000600      X     B1/1.432788/,B2/.189269/,B3/.001308/,
0000700      X      UMIN/.0000001/
0000800       U=U
0000900       CALL RAND(UNIF)
0001000       U=UNIF
0001100       IF(UNIF.GT..50) U=1.0-UNIF
0001200       IF(U.LT.UMIN) GO TO 100
0001300       T2=ALOG(1.0/(U*U))
0001400       T=SQRT(T2)
0001500       EPS=T-(A0+A1*T+A2*T2)/(1.0+B1*T+B2*T2+B3*T*T2)
0001600       IF (LEVOUT.GE.7) WRITE(IUNOUT,6000) U,T,T2,EPS
0001700       IF(UNIF.LT..50) EPS=-EPS
0001800       EPS=EPS*SIGMA
0001900       RETURN
0002000   100 EPS=-1.0E15
0002100       IF(UNIF.GT..50) EPS=-EPS
0002200       RETURN
0002300  6000 FORMAT(6H RAND   /1H 5G16.8)
0002400       END
```

```
0000100       SUBROUTINE YGEN(*)
0000200       COMMON/UNITS/IUNIN,IUNOUT,MASK,LEVOUT,PERMT(4)
0000300       COMMON/ALPH/IALPH(1000),IEALPH(512),ALPHMU(1000),REALMU(512)
0000400      X      ,NPARAM(10),MUD(11),PMU(1000)
0000500       COMMON/CNTRLS/NHYP,NALL,NFAC,NFULL,NFULM1,TRTMNT,NTREND,NOBS,LTRUE
0000600       INTEGER TRTMNT
0000700       COMMON/MM/DESM(1000)
0000800       COMMON/XRAND/Y,EPS,UNIF,U
0000900       COMMON/YDIST/S(10),R(10),SV(10),RV(10)
0001000       COMMON/PROBS/THETA(10),CSTOP
0001100       COMMON/PREMNC/NAVG(1000),PCS,XMSE,XMST,INITAL ,IHCNT(10),MSOFAR
0001400       EQUIVALENCE (SK,KS),(IX,XI)
0001600 C
0001700 C
0001800        CALL MFORM
0001900       Y=EPS
0002000       M = NPARAM(LTRUE)
0002100       JJ = MUD(LTRUE)-1
0002200       DO 550 J=1,M
0002300       JJ = JJ +1
0002400       Y = Y + DESM(JJ)*REALMU(J)
0002500   550 CONTINUE
0002600       SUM = 0.0
0002700       DO 700 N=1,NHYP
0002800       C=Y-SV(N)
0002900       Q=RV(N)*C*C
0003000       S(N)=SV(N)
0003100       R(N)=RV(N)
0003200       Q=-.50*Q
0003300       IF(ABS(Q)-60.0) 660,670,670
0003400   660 IF(THETA(N)-1.0E-18) 670,670,680
0003500   670 THETA(N) = 0.0
0003600       GO TO 700
0003700   680 A = ALOG(THETA(N)) + Q
0003800       IF(ABS(A).GE.70.0) GOTO 670
0003900   690 THETA(N) = THETA(N) * SQRT(R(N)) * EXP(Q)
0004000       SUM = SUM + THETA(N)
0004100   700 CONTINUE
0004200       I=0
0004300       DO 710 N=1,NHYP
0004400   710 THETA(N)=THETA(N)/SUM
0006100       DO 920 II=1,NHYP
0006200        IF(THETA(II)-CSTOP) 920,910,910
0006300   910 I=1
0006400       GO TO 1000
0006500   920 CONTINUE
0006600  1000 CONTINUE
0006700       IF(LEVOUT.GE.2) WRITE(IUNOUT,6010) TRTMNT,(THETA(N),N=1,NHYP)
0006800  6010 FORMAT(1H I10,10F10.4)
0006900       IF(LEVOUT.GE.5) WRITE(IUNOUT,6000) Y,EPS,(S(N),R(N),THETA(N),C(N),-
0007000      X  N = 1, NHYP)
0007100  6000 FORMAT(  3H Y=G12.5,5H EPS=G12.5, 12H S,R,THETA,C /                -
0007200      X (2X,4G16.6))
0007300       CALL UPDATE
0007400       IF(I.EQ.1) RETURN 1
0007500       RETURN
0007600       END
```

```
0000100       SUBROUTINE COUNT
0000200       COMMON/UNITS/ IUNIN,IUNOUT,MASK,LEVOUT
0000300       COMMON/ALPH/IALPH(1000),IIALPH(512),ALPHMU(1000),REALMU(512)
0000400      X      ,NPARAM(10),MUD(11),PMU(1000)
0000500       COMMON/CNTRLS/NHYP,NALL,NFAC,NFULL,NFULM1,TRTMNT,NTREND,NOBS,LTRUE
0000600       COMMON/PROBS/THETA(10),CSTOP
0001000       COMMON/PRFMNC/NAVG(10001,PCS,XMSE,XMST,INITAL ,IHCNT(10),MSOFAR
0001100       COMMON/ERR/TAU,SIGMA
0001300       XM = 0.0
0001400       NAVG(NOBS) = NAVG(NOBS) +1
0001500       M = NPARAM(LTRUE)
0001600       JJ = MUD(LTRUE)-1
0001700       DO 5  J=1,M
0001800       JJ = JJ +1
0001900       XM = XM+(ALPHMU(JJ)-REALMU(J))**2
0002000     5 CONTINUE
0002100       XMSE = SQRT(XM) + XMSE
0002200       XMST = SQRT(XM*TAU)
0002300       XM = SQRT(XM)
0002400       IF (LEVOUT.GT.1) WRITE(IUNOUT,6000) NOBS,XM,XMST
0002500       IMIN = 0
0004200       CMAX=0.0
0004300       DO 910 I=1,NHYP
0004400       IF(THETA(I)-CMAX) 910,910,905
0004500   905 CMAX=THETA(I)
0004600       IMIN=I
0004700   910 CONTINUE
0004800   945 CONTINUE
0004900       IF (IMIN.EQ.LTRUE) PCS = PCS+ 1.0
0004950       IHCNT(IMIN)=IHCNT(IMIN)+1
0005000       IF (LEVOUT.LT.2) RETURN
0005100       DO 950  N=1,NHYP
0005200       MULO = MUD(N)
0005300       MUHI = MUD(N+1)-1
0005400       WRITE (IUNOUT,6010) THETA(N) ,(ALPHMU(L),L=MULO,MUHI)
0005500   950 CONTINUE
0005600       RETURN
0005700  6000 FORMAT(22H ******OBSERVATION NO. I4,7H XMSE= G11.3,7H XMST= G11.3 -
0005800      X  ,6H****** )
0005900  6010 FORMAT(1H F10.6 , 10G12.4/(27X10G12.4))
0006000       END
```

```
0000100       SUBROUTINE INVXTX(A,NN)
0000200 C     ASSUMES THE MATRIX A IS SYMMETRIC AND POSITIVE DEFINITE, AND ONLY
0000300 C     THE  UPPER TRIANGLE IS STORED AS A ONE-DIMENSIONAL ARRAY IN THE
0000400 C     ORDER A(1,1), A(1,2), A(2,2), A(1,3), A(2,3), A(3,3), ..., A(N,N).
0000500 C     NN IS THE ORDER N OF THE INPUT MATRIX A.
0000700 C
0000800       DIMENSION A(1)
0001500       D=1.0
0001600       N = NN
0001700       ITR1 = 0
0001900       DO 145 K=1,N
0002000 C
0002100       ITR1 = ITR1+K-1
```

```
0002200       KP1 = K+1
0002300       KM1 = K-1
0002400       KK = ITR1+K
0002600       PV =  1.000/A(KK)
0002700 C
0002800       ITR2 = 0
0002900       IF (K-1) 150,80,50
0003000 C
0003100 C     REDUCE TOP PART OF TRIANGLE, LEFT OF PIVOTAL COLUMN
0003200    50 DO 60 J=1,KM1
0003300       ITR2 = ITR2+J-1
0003400       KJ = ITR1+J
0003500       F = A(KJ)*PV
0003600       DO 60 I=1,J
0003700       IJ = ITR2+I
0003800       IK = ITR1 + I
0003900    60 A(IJ) = A(IJ) + A(IK)*F
0004000 C
0004100       IF (K-N) 70,120,150
0004200 C
0004300 C     REDUCE REST OF TRIANGLE, RIGHT OF PIVOTAL COLUMN
0004400    70 ITR2 = ITR1
0004500    80 DO 110 J=KP1,N
0004600       ITR3 = ITR1
0004700       ITR2 = ITR2+J-1
0004800       KJ = ITR2+K
0004900       F = A(KJ)*PV
0005000       DO 100 I=1,J
0005100       IF (I-K) 90,100,95
0005200    90 IJ = ITR2+I
0005300       IK = ITR1 + I
0005400       A(IJ) = A(IJ) - A(IK)*F
0005500       GO TO 100
0005600    95 IJ = ITR2 + I
0005700       ITR3 = ITR3 + I - 1
0005800       IK = ITR3 + K
0005900       A(IJ) = A(IJ) - A(IK)*F
0006000   100 CONTINUE
0006100   110 CONTINUE
0006200 C
0006300 C     DIVIDE PIVOTAL ROW-COLUMN BY PIVOT, INCLUDING APPROPRIATE SIGNS
0006400   120 ITR2 = ITR1
0006500       DO 140 I=1,N
0006600       IF (I-K) 125,130,135
0006700   125 IK = ITR1+I
0006800       A(IK) = -A(IK)*PV
0006900       GO TO 140
0007000 C     (REPLACE PIVOT BY RECIPROCAL)
0007100   130 A(KK) = PV
0007200       GO TO 140
0007300   135 ITR2 = ITR2+I-1
0007400       KI = ITR2+K
0007500       A(KI) =  A(KI)*PV
0007600   140 CONTINUE
0007700 C
0007800   145 CONTINUE
0007900 C
0008000   150 RETURN
0008100       END
```

```
0000100       SUBROUTINE TRIANG(A,NN,NKOL,FORMAT,IDOUT)
0000200       DIMENSION A(1),FORMAT(1),IDOUT(1)
0000300       COMMON/UNITS/IUNIN,IUNOUT
0000400     1 FORMAT(1HK)
0000500       N = NN
0000600       NCOL = NKOL
0000700       KLUMPS = N/NCOL
0000800 C
0000900       KEEPTR = 0
0001000       K1 = 1
0001100       K2 = NCOL - 1
0001200       K3 = NCOL
0001300       IF (KLUMPS .EQ. 0)  GO TO 120
0001400 C
0001500       DO 90 KLUMP=1,KLUMPS
0001600       ITR1 = KEEPTR
0001700       I = -1
0001800       ILO = (KLUMP-1)*NCOL + ITR1 + 1
0001900       DO 30 K=K1,K2
0002000       I = I + 1
0002100       ITR1 = ITR1 + K - 1
0002200       ILO  = ILO  + K - 1
0002300       IHI = ILO + I
0002400    30 WRITE(IUNOUT,FORMAT) IDOUT(K),(A(J),J=ILO,IHI)
0002500       KEEPTR = ITR1 + K2
0002600       DO 60 K=K3,N
0002700       ITR1 = ITR1 + K - 1
0002800       ILO  = ILO  + K - 1
0002900       IHI = ILO + NCOL - 1
0003000    60 WRITE(IUNOUT,FORMAT)IDOUT(K),(A(J),J=ILO,IHI)
0003100       K1 = K1 + NCOL
0003200       K2 = K2 + NCOL
0003300       K3 = K3 + NCOL
0003400    90 WRITE(IUNOUT,1)
0003500 C
0003600   120 ITR1 = KEEPTR
0003700       IF (K) .GT. N)  GO TO 180
0003800       I = -1
0003900       ILO = KLUMPS*NCOL + ITR1 + 1
0004000       DO 150 K=K1,N
0004100       I = I + 1
0004200       ITR1 = ITR1 + K - 1
0004300       ILO  = ILO  + K - 1
0004400       IHI = ILO + I
0004500   150 WRITE(IUNOUT,FORMAT) IDOUT(K),(A(J),J=ILO,IHI)
0004600 C
0004700   180 RETURN
0004800       END
```

```
0000100       SUBROUTINE MTVEC(A,B,NN,C)
0000200       DIMENSION A(1),B(1),C(1)
0000300       N=NN
0000400       DO 500 J=1,N
0000500       C(J)=0.0
0000600       KADD =((J-1)*J)/2
0000700       DO 200 K=1,J
0000800       KADD=KADD+1
0000900       C(J)=C(J) +A(KADD)*B(K)
0001000   200 CONTINUE
0001100       K1= J+1
0001200       IF(K1-N) 250,250,500
0001300   250 DO 300 K=K1,N
0001400       KADD= KADD+K-1
0001500       C(J)= C(J)+ A(KADD)*B(K)
0001600   300 CONTINUE
0001700   500 CONTINUE
0001800       RETURN
0001900       END
```

APPENDIX B

LIST OF SYMBOLS

A	the space of allowable experiments
A_j	the space of allowable experiments requiring exactly j observations
a	element of A
$a^{(i)}$	the i^{th} experiment in A
a_j	experiment in A performed at the j^{th} stage of sampling
$\vec{B}$	vector of parameters appearing in combined model equations
$E\{X\}$	expectation of the random variable X
$E\{X\|Y\}$	conditional expectation of the random variable X given the value of Y
$\mathcal{E}(w)$	entropy of the probabilities at state w
$\mathcal{E}[w(\vec{y}),a]$	entropy of the posterior probabilities if system is in state w and the value $\vec{y}$ is observed
F	element of $\mathfrak{F}$
$\mathfrak{F}$	sigma field of Borel sets over Ω
$f_\ell(\vec{y}\|a,\vec{\alpha})$	density function of $\vec{y}$ under model ℓ when $\vec{\alpha}$ is given and experiment a is to be performed.
$f_\ell(\vec{y}\|a)$	marginal density function of $\vec{y}$ under model ℓ when experiment a is to be performed.

$f_\ell(\vec{y}_{j+1}\|a,\vec{\alpha})$	density function of $\vec{y}_{j+1}$ under model ℓ when a and $\vec{\alpha}$ are given
$f_\ell(\vec{y}_{j+1}\|a)$	marginal density function of $\vec{y}_{j+1}$ under model ℓ when a is given
$g_i(w\|w_{i-1})$	density function of w given w_{i-1}
H_ℓ	denotes hypothesis ℓ about the form of the model equation
$h_i(z_1,\ldots,z_k)$	function of controlled variables defining x_i
$I(w,a)$	expected information in experiment a when state of system is w
$I(w,a,i,j)$	expected information for discriminating in favor of H_i against H_j in experiment a when state of system is w
$\hat{i}$	denotes subscript of hypothetical model with largest posterior probability
i^*	true model equation number
J_{MAX}	upper limit on total number of observations
$k(j)$	superscript of experiment performed at stage j of sampling
L	number of model equations or hypotheses postulated
M	design matrix
M_ℓ	design matrix for model ℓ
$N(A)$	number of elements in A
$N(\vec{\mu},T)$	normal distribution with mean vector $\vec{\mu}$ and precision matrix T

$\vec{n}$	vector of n_i
n_i	number of observations taken at stage i
$n(i,j)$	number of times experiment $a^{(i)}$ is performed in j stages
$Pr\{X\}$	probability of event X
$Pr\{X\|Y\}$	probability of event X given event Y
p_i	limiting proportion of times $a^{(i)}$ performed in an infinite sequence of experiments
Q, Q_1, Q_2	denote quadratic forms
$R_{\ell,j}$	precision matrix of distribution of y_j under model ℓ
$R(w,a)$	expected reduction in entropy if experiment a is performed and state is w
$\vec{s}_{\ell,j}$	mean vector of distribution of y_j under model ℓ
T	precision matrix of distribution of ε
W	random variable defined over Ω
w	element of Ω. An observed value
X	vector of x_i
x_i	value of $h_i(z_1, \ldots, z_k)$
$x_{i,j}$	value of $h_i(z_1, \ldots, z_k)$ at j^{th}
$y, \vec{y}$	observed variable
z_i	controlled variable i
$\vec{\alpha}_\ell$	vector of parameters in model equation ℓ
β_i	coefficient of x_i
$\beta_i^{(\ell)}$	coefficient of x_i in model ℓ

Y	defined on page 41
δ_{ij}	Kronecker delta function
$\vec{\varepsilon}$	vector of observation errors
θ_i	probability model i is correct
$\theta_{i,j}$	posterior probability that model i is correct after j stages of sampling
θ_m	stopping probability
$\vec{\mu}_\ell$	mean vector of distribution of parameters in model ℓ.
$\xi_{\ell,j}(\vec{\alpha})$	density function of parameters in model ℓ after j stages of sampling
τ	precision of distribution of ε
$\Psi_{\ell,j}$	precision matrix of distribution of parameters of model ℓ after j stages of sampling
Ω	state space of process. Defined on page 19
$\Omega \times Y$	direct product of state space and observed variable space
$\vec{0}$	vector of zeros
$\propto$	proportional to
$\lvert\ \rvert$	determinant of a matrix
$\simeq$	approximately equal to
$\sim$	distributed as
$\supset$	includes

APPENDIX C

TABLES

TABLE 1. - SUMMARY OF SIMULATION RESULTS PRESENTED IN TABLES 2 THROUGH 9

Model number	Parameter	L = 4, i^* = 2		L = 4, i^* = 3		L = 6, i^* = 3		L = 6, i^* = 5	
		100 obs	500 obs	100 obs	500 obs	100 obs	500 obs	100 obs	500 obs
1	θ_1	0	0	0	0	0	0	0	0
2	θ_2	.966	.983	0	0	0	0	.010	0
3	θ_3	.032	.016	.922	.962	.860	.877	.902	.941
4	θ_4	.002	.001	.078	.038	.109	.107	.062	.023
5	θ_5					.022	.012	.017	.029
6	θ_6					.009	.004	.009	.006
1	β_0	0.0971	0.0891	0.1322	0.1377	0.1116	0.1357	0.0278	-0.0240
2	β_0	0.1010	0.0951	0.1313	0.1384	0.1253	0.1286	0.0316	0.0382
	β_1	.4981	.5025	.2485	.2522	.2411	.2544	.5114	.4977
3	β_0	0.1013	0.0941	-0.0070	-0.0067	-0.0038	-0.0015	-0.0249	-0.0099
	β_1	.4981	.5025	.2478	.2525	.2493	.2505	.5086	.5032
	β_2	-.0007	.0021	.2578	.2634	.2157	.2544	.1179	.0981
4	β_0	0.1010	0.0943	-0.0070	-0.0335	-0.0035	-0.0015	-0.0237	-0.0097
	β_1	.5029	.4915	.2497	.2416	.2572	.2429	.5168	.5046
	β_2	-.0004	.0018	.2579	.2634	.2645	.2544	.1168	.0976
	β_3	-.0076	.0112	-.0026	.0146	-.0089	.0102	-.0083	-.0017
5	β_0					0.0040	0.0015	-0.0157	0.0001
	β_1					.2564	.2440	.5120	.5009
	β_2					.2222	.2577	-.0020	.0227
	β_3					.0032	.0088	-.0048	.0024
	β_4					.0360	-.0015	.0908	.0730
6	β_0					0.0030	-0.0037	-0.0180	-0.0021
	β_1					.2390	.2559	.5057	.4638
	β_2					.2291	.2619	.0603	.0376
	β_3					.0596	-.0373	.0438	.1340
	β_4					.0301	-.0050	.0553	.0598
	β_5					-.0534	.0345	-.0413	-.0969

The column headings give the values of L and i^* and the number of observations. The row headings present the parameters whose average posterior values are given. The probabilities listed for 100 observations are the averages after five simulations of 100 observations and the values after the first 100 observations of the 500 observation simulations. The averages of the posterior parameter means are based only upon the five full simulations of 100 and 500 observations, respectively. The posterior probabilities for 500 observations are based upon five simulations of 500 observations each.

TABLE 2. - L = 4, $i^* = 2$

Model	Param	After 100 observations					After first 100 of 500 observations				
1	θ_1	0	0	0	0	0	0	0	0	0	0
2	θ_2	.973	.979	.974	.976	.975	.976	.976	.931	.977	.923
3	θ_3	.025	.019	.024	.023	.024	.023	.023	.063	.022	.071
4	θ_4	.002	.001	.002	.001	.002	.001	.001	.006	.001	.006
1	β_0^*	0.0795	0.1017	0.0906	0.1271	0.0865	*	*	*	*	*
2	β_0	0.1187	0.1017	0.0753	0.1032	0.1059	*	*	*	*	*
	β_1	.5192	.5022	.4935	.4892	.4865					
3	β_0	0.1263	0.1021	0.0682	0.1067	0.1033	*	*	*	*	*
	β_1	.5191	.5021	.4935	.4894	.4866					
	β_2	-.0152	-.0008	.0141	-.0069	.0055					
4	β_0	0.1239	0.1022	0.0688	0.1059	0.1041	*	*	*	*	*
	β_1	.4858	.5044	.4771	.5186	.5288					
	β_2	-.0126	-.0010	.0133	-.0051	.0032					
	β_3	.0351	-.0024	.0170	-.0350	-.0527					
	p_0	0.25	0.23	0.23	0.17	0.19	0.18	0.23	0.27	0.21	0.25
	p_1	0	0	0	0	0	0	0	0	0	0
	p_2	.05	.05	.03	.17	.20	.24	.06	.04	.02	.01
	p_3	.02	.03	.02	.07	0	.02	.01	.01	.03	0
	p_4	.35	.20	.14	.01	.17	.04	.26	.43	.11	.25
	p_5	.06	.20	.27	.14	.06	.12	.14	.01	.19	.24
	p_6	.01	0	.06	.09	.02	.03	.01	.01	.17	0
	p_7	.05	.05	0	.13	.21	.19	.06	.02	.01	.01
	p_8	0	0	0	0	0	0	0	0	0	0
	p_9	.21	.24	.25	.22	.15	.19	.23	.21	.26	.24

*Not recorded.

The values of the posterior probabilities and parameter means after ten simulations, of 100 observations each, of the sequential selection procedure. The last five columns are data from the first 100 observations of the 500 observation simulations tabulated in table 3. The posterior means were not recorded for these cases. Also listed are the proportions p_i of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 3. - L = 4, $i^* = 2$

Model	Param	After 500 observations				
1	θ_1	0	0	0	0	0
2	θ_2	.991	.985	.990	.991	.957
3	θ_3	.009	.015	.009	.009	.040
4	θ_4	0	0	0	0	.003
1	β_0	0.0875	0.0599	0.0905	0.0757	0.1317
2	β_0	0.0921	0.0984	0.0999	0.0942	0.0909
	β_1	.4964	.5010	.5028	.5014	.5108
3	β_0	0.0923	0.1032	0.0984	0.0937	0.0827
	β_1	.4964	.5010	.5028	.5014	.5108
	β_2	-.0005	-.0096	.0029	.0012	.0163
4	β_0	0.0923	0.1022	0.0985	0.0935	0.0850
	β_1	.4948	.4886	.5043	.4882	.4817
	β_2	-.0004	-.0086	.0028	.0014	.0139
	β_3	.0016	.0126	-.0017	.0141	.0293
	p_0	0.234	0.264	0.236	0.236	0.228
	p_1	0	0	0	0	.002
	p_2	.050	.010	.056	.044	0
	p_3	.008	0	.004	.008	0
	p_4	.280	.422	.302	.306	.076
	p_5	.110	.060	.070	.088	.418
	p_6	.018	0	.026	.076	.006
	p_7	.072	.010	.088	.038	.002
	p_8	0	0	0	0	0
	p_9	.228	.226	.218	.204	.268

The values of the posterior probabilities and parameter means after 5 simulations, of 500 observations each, of the sequential selection procedure. Also listed are the proportions p_i of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 4. - L = 4, $i^* = 3$

Model	Param	After 100 observations					After first 100 of 500 observations				
1	θ_1	0	0	0	0	0	0	0	0	0	0
2	θ_2	0	0	0	0	0	0	0	0	0	0
3	θ_3	.788	.828	.967	.966	.966	.962	.916	.966	.941	.920
4	θ_4	.212	.172	.033	.034	.034	.038	.084	.034	.059	.080
1	β_0	0.1140	0.1548	0.1260	0.1292	0.1368	*	*	*	*	*
2	β_0	0.1183	0.1515	0.1255	0.1286	0.1324	*	*	*	*	*
	β_1	.2452	.2533	.2385	.2575	.2482					
3	β_0	-0.0043	-0.0089	-0.0159	0.0104	-0.0162	*	*	*	*	*
	β_1	.2467	.2510	.2389	.2552	.2470					
	β_2	.2263	.2939	.2683	.2216	.2788					
4	β_0	-0.0045	-0.0091	-0.0158	0.0105	-0.0162	*	*	*	*	*
	β_1	.1838	.3095	.2395	.2633	.2523					
	β_2	.2268	.2945	.2681	.2215	.2788					
	β_3	.0833	-.0779	-.0009	-.0107	-.0070					
	p_0	0.18	0.17	0.17	0.17	0.17	0.18	0.17	0.17	0.17	0.17
	p_1	0	0	0	0	0	0	0	0	0	0
	p_2	.32	.31	.30	.30	.29	.32	.30	.30	.31	.30
	p_3	0	.01	0	.02	0	0	0	0	.01	.02
	p_4	0	0	.03	0	.02	0	.03	.02	0	0
	p_5	.03	.01	.02	.03	.04	.02	.02	.02	.01	.05
	p_6	0	0	.01	.02	0	.01	0	0	0	.02
	p_7	.30	.32	.30	.28	.30	.30	.31	.31	.32	.28
	p_8	0	0	0	0	0	0	0	0	0	0
	p_9	.17	.18	.17	.18	.18	.17	.17	.18	.18	.16

The values of the posterior probabilities and parameter means after 10 simulations, of 100 observations each, of the sequential selection procedure. The last 5 columns are data from the first 100 observations of the 500 observation simulations tabulated in table 5. The posterior means for these 5 cases were not tabulated. Also listed are the proportions p_i of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 5. - L = 4, $i^* = 3$

Model	Param	After 500 observations				
1	θ_1	0	0	0	0	0
2	θ_2	0	0	0	0	0
3	θ_3	.953	.982	.953	.969	.954
4	θ_4	.047	.018	.047	.031	.046
1	β_0	0.1368	0.1385	0.1383	0.1394	0.1354
2	β_0	0.1376	0.1398	0.1391	0.1387	0.1369
	β_1	.2561	.2463	.2608	.2550	.2426
3	β_0	-0.0227	0.0085	-0.0069	-0.0028	-0.0096
	β_1	.2566	.2469	.2611	.2546	.2434
	β_2	.2906	.2392	.2648	.2548	.2677
4	β_0	-0.0227	0.0085	-0.0069	-0.0028	-0.0096
	β_1	.2355	.2385	.2399	.2715	.2225
	β_2	.2905	.2392	.2649	.2548	.2678
	β_3	.0281	.0112	.0281	-.0223	.0277
	p_0	0.178	0.178	0.178	0.178	0.178
	p_1	0	0	0	0	0
	p_2	.322	.320	.320	.318	.318
	p_3	0	0	0	.002	.004
	p_4	0	.006	.004	0	0
	p_5	.004	.004	.004	.002	.010
	p_6	.002	0	0	0	.004
	p_7	.318	.318	.318	.320	.312
	p_8	0	0	0	0	0
	p_9	.176	.174	.176	.180	.174

The values of the posterior probabilities and parameter means after 5 simulations, of 500 observations each, of the sequential selection procedure. Also listed are the proportions p_i of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 6. - L = 6, $i^* = 3$

Model	Param	After 100 observations					After first 100 of 500 observations				
1	θ_1	0	0	0	0	0	0	0	0	0	0
2	θ_2	0	0	0	0	0	0	0	0	0	0
3	θ_3	.9554	.9564	.7630	.9098	.9432	.2733	.9534	.9471	.9449	.9554
4	θ_4	.0378	.0365	.1756	.0738	.0469	.5534	.0400	.0444	.0457	.0378
5	θ_5	.0052	.0055	.0451	.0130	.0075	.1168	.0050	.0070	.0075	.0052
6	θ_6	.0016	.0017	.0162	.0034	.0024	.0564	.0016	.0015	.0019	.0016
1	β_0	0.1567	0.1026	0.0891	0.0839	0.1259	*	*	*	*	*
2	β_0	0.1298	0.1197	0.1192	0.1118	0.1462	*	*	*	*	*
	β_1	.2696	.2258	.2151	.2422	.2527					
3	β_0	-0.0114	0.0098	-0.0332	0.0004	0.0156	*	*	*	*	*
	β_1	.2564	.2354	.2396	.2542	.2607					
	β_2	.2882	.2282	.3290	.2329	.2463					
4	β_0	-0.0110	0.0098	-0.0317	-0.0008	0.0161	*	*	*	*	*
	β_1	.2421	.2427	.3033	.2117	.2864					
	β_2	.2877	.2282	.3268	.2346	.2453					
	β_3	.0189	-.0099	-.0846	.0562	-.0341					
5	β_0	-0.0056	0.0200	-0.0082	-0.0175	0.0314	*	*	*	*	*
	β_1	.2458	.2403	.2870	.2232	.2855					
	β_2	.2562	.1692	.1788	.3427	.1641					
	β_3	.0148	-.0080	-.0676	.0432	.0336					
	β_4	.0271	.0512	.1288	-.0966	.0694					
6	β_0	-0.0061	0.0209	-0.0130	-0.0166	0.0300	*	*	*	*	*
	β_1	.2429	.1945	.3250	.2050	.2278					
	β_2	.2583	.1767	.1921	.3418	.1767					
	β_3	.0265	.1913	-.2368	.1214	.1958					
	β_4	.0254	.0432	.1203	-.0968	.0584					
	β_5	-.0089	-.1555	.1320	-.0609	-.1739					
	p_0	0.13	0.15	0.18	0.17	0.17	0.18	0.10	0.11	0.17	0.13
	p_1	.02	.08	.03	.06	.10	.01	.04	.11	.12	.02
	p_2	.17	.24	.32	.27	.22	.32	.12	.12	.20	.17
	p_3	.08	.03	0	0	.02	0	.05	.02	.02	.08
	p_4	.06	.03	.02	.07	.03	.01	.07	.14	.07	.06
	p_5	.03	.03	.07	.03	.05	0	.11	.02	.04	.03
	p_6	0	.12	.14	.14	.02	.06	.02	.06	.07	0
	p_7	.29	.17	.11	.09	.21	.26	.16	.15	.15	.29
	p_8	.05	.02	.01	.05	.04	0	.16	.12	.04	.05
	p_9	.17	.13	.12	.12	.14	.16	.17	.15	.12	.17

*Not recorded.

The values of the posterior probabilities and parameter means after 10 simulations, of 100 observations each, of the sequential selection procedure. The last 5 columns are data from the first 100 observations of the 500 observation simulations tabulated in table 7. The posterior means were not recorded for these 5 cases. Also listed are the proportions of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 7. - L = 6, $i^* = 3$

Model	Param	After 500 observations				
1	θ_1	0	0	0	0	0
2	θ_2	0	0	0	0	0
3	θ_3	.6046	.9812	.9722	.9746	.8526
4	θ_4	.3388	.0175	.0257	.0230	.1316
5	θ_5	.0425	.0009	.0018	.0021	.0125
6	θ_6	.0141	.0003	.0003	.0003	.0032
1	β_0	0.1321	0.1378	0.1325	0.1349	0.1413
2	β_0	0.1356	0.1278	0.1247	0.1168	0.1383
	β_1	.2434	.2616	.2541	.2549	.2581
3	β_0	-0.0026	0.0067	-0.0066	-0.0021	-0.0029
	β_1	.2446	.2556	.2486	.2466	.2571
	β_2	.2542	.2336	.2605	.2615	.2622
4	β_0	-0.0027	0.0067	-0.0067	-0.0017	-0.0030
	β_1	.2076	.2515	.2340	.2341	.2872
	β_2	.2544	.2335	.2608	.2609	.2623
	β_3	.0491	.0055	.0194	.0167	-.0399
5	β_0	-0.0113	0.0077	0.0111	0.0063	-0.0062
	β_1	.2081	.2518	.2332	.2394	.2873
	β_2	.2926	.2279	.2845	.2061	.2774
	β_3	.0486	.0052	.0203	.0101	-.0401
	β_4	-.0300	.0049	-.0200	.0496	-.0121
6	β_0	-0.0118	0.0047	-0.0100	0.0061	-0.0073
	β_1	.2164	.3128	.2432	.2386	.2682
	β_2	.2949	.2455	.2801	.2070	.2819
	β_3	.0130	-.2307	-.0225	.0134	.0405
	β_4	-.0317	-.0098	-.0168	.0490	-.0155
	β_5	.0270	.1773	.0328	-.0026	-.0618
	p_0	0.178	0.136	0.148	0.116	0.168
	p_1	.004	.064	.028	.070	.012
	p_2	.314	.206	.198	.086	.282
	p_3	.004	.036	.092	.112	.018
	p_4	.010	.014	.038	.110	.014
	p_5	0	.092	.014	.024	.016
	p_6	.022	.004	.012	.022	.002
	p_7	.296	.190	.280	.288	.304
	p_8	0	.104	.026	.014	.010
	p_9	.172	.154	.164	.158	.174

The values of the posterior probabilities and parameter means after five simulations, of 500 observations each, of the sequential selection procedure. Also listed are the proportions of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 8. - L = 6, $i^* = 5$

Model	Param	After 100 observations					After first 100 of 500 observations				
1	θ_1	0	0	0	0	0	0	0	0	0	0
2	θ_2	0	0	0	.021	.042	0	0	.011	.003	.026
3	θ_3	.945	.942	.956	.895	.848	.943	.877	.852	.904	.857
4	θ_4	.043	.043	.038	.063	.042	.048	.106	.089	.070	.076
5	θ_5	.007	.012	.005	.015	.033	.007	.013	.035	.018	.028
6	θ_6	.004	.003	.001	.006	.035	.002	.004	.013	.006	.013
1	β_0	0.1502	-0.0299	0.0231	-0.0079	0.0034	*	*	*	*	*
2	β_0	0.0356	0.0189	0.0288	0.0316	0.0431	*	*	*	*	*
	β_1	.5159	.5101	.5123	.5079	.5106					
3	β_0	-0.0348	-0.0413	-0.0412	-0.0098	0.0026	*	*	*	*	*
	β_1	.5040	.5077	.5096	.5084	.5133					
	β_2	.1467	.1265	.1478	.0837	.0850					
4	β_0	-0.0333	-0.0396	-0.0414	-0.0070	0.0026	*	*	*	*	*
	β_1	.4874	.5288	.5019	.5513	.5146					
	β_2	.1450	.1252	.1478	.0810	.0849					
	β_3	.0217	-.0261	.0099	-.0456	-.0016					
5	β_0	-0.0232	-0.0242	-0.0431	-0.0041	0.0161	*	*	*	*	*
	β_1	.4985	.5201	.5039	.5516	.4861					
	β_2	.0642	-.0063	.1616	.0184	-.1260					
	β_3	.0101	-.0185	.0074	-.0463	.0235					
	β_4	.0774	.1223	-.0131	.0605	.2068					
6	β_0	-0.0298	-0.0234	-0.0435	-0.0067	0.0135	*	*	*	*	*
	β_1	.6005	.5398	.5233	.5143	.3504					
	β_2	.1187	-.0120	.1655	.0517	-.0226					
	β_3	-.3886	-.1048	-.0708	.1797	.6034					
	β_4	.0292	.1270	-.0167	.0301	.1071					
	β_5	.3037	.0673	.0599	-.1893	-.4482					
	p_0	0.10	0.16	0.14	0.24	0.18	0.18	0.14	0.17	0.18	0.22
	p_1	.05	.06	.07	.02	.13	.25	.04	.02	.01	.01
	p_2	.08	.20	.15	.03	.03	.06	.17	.04	.25	.02
	p_3	.04	.05	.07	.07	.02	.01	.06	.01	.03	0
	p_4	.08	.13	.08	.26	.18	.16	.06	.10	.10	.05
	p_5	.08	.08	.10	.12	.19	.07	.17	.25	.06	.41
	p_6	.08	.04	.06	.02	.03	.03	0	0	.14	.01
	p_7	.11	.10	.09	.03	.05	.09	.19	.19	.05	.02
	p_8	.22	.02	.10	0	.01	.06	.03	.01	.01	0
	p_9	.16	.16	.14	.21	.18	.09	.14	.21	.17	.26

*Not recorded.

The values of the posterior probabilities and parameter means after 10 simulations, of 100 observations each, of the sequential selection procedure. The last 5 columns are data from the first 100 observations of the 500 observation simulations tabulated in table 9. Also listed are the proportions of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 9. - L = 6, $i^* = 5$

Model	Param	After 500 observations				
1	θ_1	0	0	0	0	0
2	θ_2	0	0	0	0	0
3	θ_3	.974	.882	.899	.976	.976
4	θ_4	.020	.024	.029	.021	.022
5	θ_5	.003	.075	.062	.002	.002
6	θ_6	.002	.020	.009	0	0
1	β_0	-0.1051	0.0255	-0.0926	0.0130	0.0390
2	β_0	0.0290	0.0373	0.0436	0.0351	0.0458
	β_1	.4860	.5032	.4903	.5038	.5053
3	β_0	-0.0181	-0.0137	0.0018	-0.0258	0.0065
	β_1	.5016	.5008	.5019	.5056	.5061
	β_2	.1075	.1046	.0803	.1189	.0791
4	β_0	-0.0179	-0.0142	0.0027[c]	-0.0257	0.0066
	β_1	.5035	.4859	.5201	.5177	.4956
	β_2	.1071	.1050	.0782	.1187	.0790
	β_3	-.0025	.0198	-.0237	-.0160	.0138
5	β_0	-0.0100	0.0012	0.0159	-0.0202	0.0136
	β_1	.4959	.4940	.5016	.5169	.4962
	β_2	.0413	-.0152	-.0328	.0897	.0306
	β_3	.0061	.0081	0	-.0151	.0129
	β_4	.0653	.1185	.1116	.0241	.0455
6	β_0	-0.0166	-0.0044	0.0157	-0.0186	0.0135
	β_1	.4232	.4352	.4617	.5039	.4949
	β_2	.0891	.0094	-.0250	.0835	.0309
	β_3	.2801	.2087	.1232	.0400	.0179
	β_4	.0224	.0996	.1030	.0288	.0452
	β_5	-.2070	-.1452	-.0854	-.0430	-.0039
	p_0	0.158	0.106	0.208	0.174	0.132
	p_1	.252	.190	.314	.010	.128
	p_2	.108	.054	.016	.306	.138
	p_3	.040	.052	.014	.006	.008
	p_4	.170	.184	.100	.020	.104
	p_5	.022	.034	.114	.030	.088
	p_6	.056	0	0	.072	.002
	p_7	.074	.134	.052	.208	.172
	p_8	.036	.118	.110	.022	.092
	p_9	.084	.128	.072	.152	.132

The values of the posterior probabilities and parameter means after five simulations, of 500 observations each, of the sequential selection procedure. Also listed are the proportions of the times each $a^{(i)}$ was chosen as the optimal experiment.

TABLE 10. - LARGE SAMPLE STUDY TWO

Model	Param	$\vec{\theta}_0$ = (0.1,0.2,0.3,0.4)						$\vec{\theta}_0$ = (0.25,0.25,0.25,0.25)						$\vec{\theta}_0$ = (0.4,0.3,0.2,0.1)					
		5 simulations					Average	5 simulations					Average	5 simulations					Average
1	θ_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	θ_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	θ_3	.957	.947	.828	.926	.960	.924	.969	.969	.966	.967	.930	.960	.984	.979	.976	.957	.984	.976
4	θ_4	.043	.053	.172	.074	.040	.076	.031	.031	.034	.033	.070	.040	.016	.021	.024	.043	.016	.024
1	β_0	-0.0338	-0.0460	.0015	-0.0282	-0.0125	-0.0238	-0.0413	-0.0230	-0.0204	-0.0501	-0.0172	-0.0304	0.0392	0.0028	0.0045	-0.0195	0.0180	0.0090
2	β_0	-0.0338	-0.0460	0.0015	-0.0282	-0.0125	-0.0238	-0.0413	-0.0230	-0.0204	-0.0501	-0.0172	-0.0304	0.0392	0.0028	0.0045	-0.0195	0.0180	0.0090
	β_1	1.0044	.9810	.9938	.9989	.9644	.9885	1.0142	.9779	1.0510	1.0487	.9697	1.012	.9484	1.0213	1.0603	.9906	.9614	.9964
3	β_0	-0.0338	-0.0460	0.0015	-0.0282	-0.0125	-0.0238	-0.0413	-0.0230	-0.0204	-0.0501	-0.0172	-0.0304	0.0392	0.0028	0.0045	-0.0195	0.0180	0.0090
	β_1	1.0044	.9810	.9938	.9989	.9644	.9885	1.0142	.9779	1.0510	1.0487	.9697	1.012	.9484	1.0213	1.0603	.9906	.9614	.9964
	β_2	-1.0555	-1.0012	-.9594	-1.0026	-.9948	-1.003	-.9753	-1.0205	-1.0208	-1.0139	-.9762	-1.001	-.9451	-.9539	-1.0253	-.9339	-1.0012	-.9719
4	β_0	-0.0338	-0.0460	0.0015	-0.0282	-0.0125	-0.0238	-0.0413	-0.0230	-0.0204	-0.0501	-0.0172	-0.0304	0.0392	0.0028	0.0045	-0.0195	0.0180	0.0090
	β_1	1.0044	.9810	.9938	.9989	.9644	.9885	1.0142	.9779	1.0510	1.0487	.9697	1.012	.9484	1.0213	1.0603	.9906	.9614	.9964
	β_2	-1.0555	-1.0012	-.9594	-1.0026	-.9948	-1.003	-.9753	-1.0205	-1.0208	-1.0139	-.9762	-1.001	-.9451	-.9539	-1.0253	-.9339	-1.0012	-.9719
	β_3	.0110	-.0236	.0564	.0355	-.0016	.0155	-.0012	-.0052	-.0152	-.0121	.0418	.0016	.0019	-.0239	.0290	-.0456	0.0087	-.0060

The values of the posterior probabilities and parameter means after five simulations, of 500 observations each, of the sequential selection procedure with three different prior distributions on the models.

TABLE 11. - LARGE SAMPLE STUDY THREE

Model	Parameter	τ = 100.0						τ = 1.0						τ = 0.01					
		5 simulations					Avg	5 simulations					Avg	5 simulations					Avg
1	θ_1	1	1	1	1	1	1	1	1	1	1	1	1	0.9979	0.8155	0.9999	0.8951	0.9990	0.9415
2	θ_2	0	0	0	0	0	0	0	0	0	0	0	0	.0006	.1843	0	.0012	.0004	.0373
3	θ_3	0	0	0	0	0	0	0	0	0	0	0	0	.0015	.0002	.0001	.1037	.0006	.0212
1	β_0	0.953	1.027	1.094	0.979	1.106	1.032	1.111	1.091	1.046	1.298	0.932	1.096	1.415	0.454	1.221	0.916	0.906	0.982
	β_1	1.069	1.046	.972	1.067	.979	1.027	1.009	.890	.957	.800	.874	.906	1.093	1.274	1.318	.632	1.166	1.097
	β_2	.984	1.039	.997	.907	1.033	.992	.862	.910	1.210	.822	1.160	.993	1.163	.542	1.607	1.119	1.193	1.125
2	β_0	1.230	1.254	1.298	1.262	1.295	1.268	1.135	1.075	1.567	1.281	0.527	1.117	1.413	0.454	0.754	0.914	0.904	0.888
	β_2	1.559	1.564	1.469	1.520	1.494	1.521	1.030	.911	1.009	.816	.907	.935	1.122	1.275	1.319	.673	1.172	1.112
	β_3	.770	.746	.702	.738	.705	.732	.039	.019	.039	-.020	.219	.059	.074	.154	.031	.201	.018	.096
3	β_0	1.742	1.774	1.788	1.754	1.796	1.771	1.122	1.108	1.002	1.306	0.940	1.096	1.415	0.418	1.221	0.916	0.906	0.976
	β_2	1.515	1.559	1.480	1.438	1.520	1.502	.882	.928	1.246	.842	1.159	1.011	1.184	.613	1.607	1.121	1.198	1.145
	β_3	.258	.226	.212	.246	.204	.229	-.048	.013	-.041	-.048	.057	-.013	.056	.301	.005	.086	.012	.092
	Number of trials until $\theta_{1,3}$ = 1.0	3	3	3	3	3	3	49	47	39	59	52	49.2						

The posterior probabilities and posterior parameter means after five simulations of the sequential selection procedure for three different values of τ. For τ = 100.0 and τ = 1.0 the number of trials until $\theta_{1,3}$ = 1.0 (within the accuracy of the computer) is also tabulated. For τ = 0.01 the values are based upon 1000 observations.

TABLE 12. - SMALL SAMPLE STUDY ONE (H_3 TRUE)

[$J_{MAX} = 8$]

θ_m	τ	$\vec{\mu}_{3,0}$	PCS	ASN	Starting value for random seq.
0.70	0.5	(0, 0)	0.133	6.36	*
.70	.5	(0.5, 0.5)	.458	7.15	041 574 501 221
.70	.5	(1.0, 1.0)	.544	6.82	261 404 147 531
.70	.5	(1.5, 1.5)	.446	5.89	251 233 175 021
.80	.5	(0, 0)	.173	7.50	265 603 111 061
.80	.5	(0.5, 0.5)	.468	7.78	237 616 233 015
.80	.5	(1.0, 1.0)	.531	7.52	066 231 644 355
.80	.5	(1.5, 1.5)	.460	7.24	124 715 646 251
.90	.5	(0, 0)	.229	7.98	202 255 025 241
.90	.5	(0.5, 0.5)	.479	7.92	020 625 757 465
.90	.5	(1.0, 1.0)	.513	7.81	154 510 176 555
.90	.5	(1.5, 1.5)	.439	7.76	043 355 261 141
.70	1.0	(0, 0)	.397	5.49	031 264 722 101
.70	1.0	(0.5, 0.5)	.673	5.88	142 153 215 611
.70	1.0	(1.0, 1.0)	.737	5.29	025 206 250 121
.70	1.0	(1.5, 1.5)	.621	4.84	244 233 735 061
.80	1.0	(0, 0)	.558	6.90	337 020 177 205
.80	1.0	(0.5, 0.5)	.755	6.94	361 341 044 651
.80	1.0	(1.0, 1.0)	.771	6.50	231 737 436 405
.80	1.0	(1.5, 1.5)	.700	6.22	107 152 460 271
.90	1.0	(0, 0)	.605	7.80	316 753 345 645
.90	1.0	(0.5, 0.5)	.765	7.45	042 264 053 551
.90	1.0	(1.0, 1.0)	.777	7.15	304 456 707 705
.90	1.0	(1.5, 1.5)	.689	7.12	324 670 521 455
.70	2.0	(0, 0)	.699	4.24	034 773 264 025
.70	2.0	(0.5, 0.5)	.871	4.03	361 656 711 721
.70	2.0	(1.0, 1.0)	.877	3.62	110 151 661 121
.70	2.0	(1.5, 1.5)	.723	3.48	000 766 306 641
.80	2.0	(0, 0)	.868	5.45	020 542 277 271
.80	2.0	(0.5, 0.5)	.962	4.99	073 755 766 635
.80	2.0	(1.0, 1.0)	.970	4.63	013 527 071 701
.80	2.0	(1.5, 1.5)	.872	4.61	041 554 522 311
.90	2.0	(0, 0)	.944	6.46	001 231 353 331
.90	2.0	(0.5, 0.5)	.967	5.66	361 503 245 415
.90	2.0	(1.0, 1.0)	.969	5.48	151 650 040 041
.90	2.0	(1.5, 1.5)	.939	5.80	233 434 565 701

*Not recorded.

Resulting PCS and ASN values for $J_{MAX} = 8$ and the combinations of θ_m, τ, and $\vec{\mu}_{3,0}$. Results are based upon 1500 simulations of the procedure for each combination.

TABLE 13. - SMALL SAMPLE STUDY ONE (H_3 TRUE)

[J_{MAX} = 16]

θ_m	τ	$\vec{\mu}_{3,0}$	PCS	ASN	Starting value for random seq.
0.70	0.5	(0, 0)	0.354	9.48	272 036 225 461
.70	.5	(0.5, 0.5)	.665	10.7	057 343 345 741
.70	.5	(1.0, 1.0)	.723	9.63	073 144 502 151
.70	.5	(1.5, 1.5)	.555	7.38	231 500 657 525
.80	.5	(0, 0)	.508	13.6	033 254 034 051
.80	.5	(0.5, 0.5)	.761	13.3	225 553 740 341
.80	.5	(1.0, 1.0)	.806	12.3	134 537 257 651
.80	.5	(1.5, 1.5)	.661	11.8	251 356 646 745
.90	.5	(0, 0)	.574	15.5	056 537 424 615
.90	.5	(0.5, 0.5)	.752	14.6	246 632 674 651
.90	.5	(1.0, 1.0)	.800	13.9	140 077 157 311
.90	.5	(1.5, 1.5)	.710	13.8	044 035 362 005
.70	1.0	(0, 0)	.548	6.53	173 052 463 251
.70	1.0	(0.5, 0.5)	.821	6.82	063 364 104 441
.70	1.0	(1.0, 1.0)	.825	6.09	233 034 770 255
.70	1.0	(1.5, 1.5)	.637	5.36	017 237 125 325
.80	1.0	(0, 0)	.808	9.48	275 264 535 015
.80	1.0	(0.5, 0.5)	.971	9.30	015 352 360 531
.80	1.0	(1.0, 1.0)	.961	8.16	017 142 770 505
.80	1.0	(1.5, 1.5)	.865	7.86	004 724 275 765
.90	1.0	(0, 0)	.927	12.1	161 027 043 101
.90	1.0	(0.5, 0.5)	.973	10.8	101 732 737 651
.90	1.0	(1.0, 1.0)	.964	10.1	016 351 614 135
.90	1.0	(1.5, 1.5)	.958	10.6	171 716 572 235
.70	2.0	(0, 0)	.700	4.25	073 021 660 321
.70	2.0	(0.5, 0.5)	.878	4.17	003 466 340 375
.70	2.0	(1.0, 1.0)	.855	3.59	337 170 131 645
.70	2.0	(1.5, 1.5)	.714	3.51	055 666 256 215
.80	2.0	(0, 0)	.911	5.67	037 537 412 725
.80	2.0	(0.5, 0.5)	.990	5.12	111 525 350 761
.80	2.0	(1.0, 1.0)	.988	4.84	003 413 673 201
.80	2.0	(1.5, 1.5)	.894	4.71	055 643 644 455
.90	2.0	(0, 0)	.996	7.13	374 543 153 375
.90	2.0	(0.5, 0.5)	1.00	6.25	133 225 727 441
.90	2.0	(1.0, 1.0)	1.00	5.94	332 405 117 171
.90	2.0	(1.5, 1.5)	.995	6.20	010 312 536 461

Resulting PCS and ASN values for J_{MAX} = 16 and the combinations of θ_m, τ, and $\vec{\mu}_{3,0}$. Results based upon 1000 simulations.

TABLE 14. - SMALL SAMPLE STUDY TWO (H_2 TRUE)

[J_{MAX} = 8]

θ_m	τ	$\vec{\mu}_{2,0}$	PCS	ASN	Starting value for random seq.			
0.70	0.5	(1.0)	0.760	7.86	052	516	237	355
.80	.5	(1.0)	.734	7.98	016	160	602	721
.90	.5	(1.0)	.740	7.98	245	577	171	655
.70	1.0	(.5)	.828	7.63	321	722	414	631
.70	1.0	(1.0)	.882	7.20	340	321	470	071
.70	1.0	(1.5)	.800	6.86	360	415	546	645
.80	1.0	(1.0)	.872	7.98	273	760	237	431
.90	1.0	(.5)	.880	7.97	006	761	404	325
.90	1.0	(1.0)	.898	7.98	331	151	347	271
.90	1.0	(1.5)	.832	7.99	372	024	174	011
.70	2.0	(1.0)	.900	5.13	004	415	604	245
.80	2.0	(1.0)	.936	7.89	063	456	575	211
.90	2.0	(1.0)	.934	7.98	065	654	616	225

The PCS and ASN values resulting from 500 simulations of the sequential procedure for each of the tabulated combinations of θ_m, τ, and $\vec{\mu}_{2,0}$.

TABLE 15. - SMALL SAMPLE STUDY THREE (FOUR MODEL PROBLEM)

[J_{MAX} = 8]

θ_m	τ	$\vec{\mu}_{3,0}$	PCS	ASN	Starting value for random seq.			
0.70	1.0	(0, 1, -1)	0.767	7.55	*			
.70	1.0	(0, 0.5, 0)	.442	7.10	006	171	767	411
.70	1.0	(0, 0, 0.5)	.027	5.74	113	071	707	045
.70	1.0	(1, 0.5, 0)	.154	6.40	032	457	065	345
.80	1.0	(0, 1, -1)	.792	7.91	315	037	701	221
.80	1.0	(0, 0.5, 0)	.524	7.67	070	131	010	071
.80	1.0	(0, 0, 0.5)	.030	6.73	044	541	754	365
.80	1.0	(1, 0.5, 0)	.200	7.31	034	264	602	535
.90	1.0	(0, 1, -1)	.790	7.99	175	260	740	521
.90	1.0	(0, 0.5, 0)	.506	7.91	000	247	732	655
.90	1.0	(0, 0, 0.5)	.025	7.51	276	504	634	101
.90	1.0	(1, 0.5, 0)	.210	7.81	243	240	621	255

*Not recorded.

PCS and ASN values resulting from 1000 simulations performed for the indicated combinations of θ_m, τ, and $\vec{\mu}_{3,0}$.

TABLE 16. - CODED DATA FOR SAMPLE PROBLEM

(DATA TAKEN FROM DRAPER AND SMITH)

z_1	z_2	z_3	z_4	y
-75	0	0	-65	1.4
175	0	0	150	26.3
0	0	-65	150	29.4
0	0	165	-65	9.7
0	0	0	150	32.9
-75	-75	0	150	26.4
175	175	0	-65	8.4
-75	-75	-65	150	28.4
175	175	165	-65	11.5
0	0	-65	-65	1.3
0	0	165	150	21.4
0	-75	-65	-65	.4
0	175	165	150	22.9
0	0	0	-65	3.7
0	-75	0	150	26.5
0	-75	0	150	23.4
0	-75	0	150	26.5
0	175	0	-65	5.8
0	175	0	-65	7.4
0	175	0	-65	5.8
0	-75	-65	150	28.8
0	-75	-65	150	26.4
0	175	165	-65	11.8
0	175	165	-65	11.4

TABLE 17. - SUMMARY OF ANALYSIS OF EQUATION (7-1)

USING DATA OF TABLE 16

Term of model	Estimated coefficient	t-statistic	Descriptive significance level
z_4	0.112	28.4	0.999+
z_3z_4	-.354E-3	6.8	.999+
z_2	.323E-1	3.4	.986
z_3	.235E-1	2.5	.955
z_1	.319E-1	2.1	.920
z_1z_4	-.416E-3	1.9	.890
z_2^2	.705E-4	1.9	.886
z_3^2	-.128E-3	1.6	.836
z_1z_2	-.339E-3	1.2	.717
z_1z_3	-.669E-4	.9	.576
z_1^2	.705E-4	.5	.367
z_2z_3	.332E-4	.5	.347
z_2z_4	.178E-4	.3	.217

R^2 = 0.988
Residual mean square = 3.25
Replication mean square = 1.85

$$F = \frac{\text{Lack-of-fit mean square}}{\text{Replication mean square}} = 2.90$$

TABLE 18a. - THE MODELS FOR EXAMPLE 1

$$H_1: \quad y = \beta_0 + \beta_1 z_4 + \beta_2 z_3 z_4 + \varepsilon$$

$$H_2: \quad y = \beta_0 + \beta_1 z_4 + \beta_2 z_3 z_4 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + \varepsilon$$

$$H_3: \quad y = \beta_0 + \beta_1 z_4 + \beta_2 z_3 z_4 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + \beta_6 z_2^2 + \beta_7 z_3^2 + \beta_8 z_1 z_4 + \varepsilon$$

$$H_4: \quad y = \beta_0 + \beta_1 z_4 + \beta_2 z_3 z_4 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + \beta_6 z_2^2 + \beta_7 z_3^2 + \beta_8 z_1 z_4 + \beta_9 z_1^2 + \beta_{10} z_1 z_2 + \beta_{11} z_1 z_3 + \beta_{12} z_2 z_3 + \beta_{13} z_2 z_4 + \beta_{14} z_4^2 + \varepsilon$$

TABLE 18b. - THE PRIOR MEANS FOR EXAMPLE 1

$$\vec{\mu}_{1,0} = \begin{pmatrix} 1.215\times10 \\ 9.791\times10^{-2} \\ -2.650\times10^{-4} \end{pmatrix}$$

$$\vec{\mu}_{2,0} = \begin{pmatrix} 1.064\ 10 \\ 1.113\ 10^{-1} \\ -3.258\ 10^{-4} \\ 2.211\ 10^{-3} \\ 1.761\ 10^{-2} \\ 1.066\ 10^{-2} \end{pmatrix}$$

$$\vec{\mu}_{3,0} = \begin{pmatrix} 11.76 \\ .1137 \\ -.3376\times10^{-3} \\ .3322\times10^{-2} \\ .3114\times10^{-1} \\ .1768\times10^{-1} \\ -.1158\times10^{-3} \\ -.6788\times10^{-4} \\ -.1076\times10^{-3} \end{pmatrix}$$

$$\vec{\mu}_{4,0} = \begin{pmatrix} 12.70 \\ 0.1119 \\ -.3542\times10^{-3} \\ .3194\times10^{-1} \\ .3226\times10^{-1} \\ .2354\times10^{-1} \\ -.1512\times10^{-3} \\ -.1277\times10^{-3} \\ -.4164\times10^{-3} \\ .7045\times10^{-4} \\ -.3393\times10^{-3} \\ -.6690\times10^{-4} \\ .3323\times10^{-4} \\ .1785\times10^{-4} \\ 0 \end{pmatrix}$$

TABLE 18c. - THE MATRIX FROM WHICH $W_{\ell,0}$ MAY BE TAKEN

Row	1	24.00000					
Row	2	1020.000	320700.0				
Row	3	-23950.00	0.381425E+07	0.210118E+10			
Row	4	300.0000	-14125.00	-0.114562E+07	108750.0		
Row	5	800.0000	-127250.0	577500.0	72500.00	290000.0	
Row	6	600.0000	-23950.00	0.307475E+07	33750.00	135000.0	188700.0
Row	7	290000.0	-0.380000E+07	-0.368156E+09	0.987500E+07	0.395000E+08	0.187500E+08
Row	8	188700.0	0.307475E+07	0.506112E+08	0.444750E+07	0.177900E+08	0.253050E+08
Row	9	-14125.00	0.172438E+07	0.231684E+09	0.193437E+07	-0.229375E+07	-0.114562E+07
Row	10	108750.0	0.193437E+07	-0.383297E+09	0.148125E+08	0.987500E+07	0.468750E+07
Row	11	72500.00	-0.229375E+07	-0.383297E+09	0.987500E+07	0.987500E+07	0.468750E+07
Row	12	33750.00	-0.114562E+07	-0.357216E+09	0.468750E+07	0.468750E+07	0.444750E+07
Row	13	13500 .0	577500.0	-0.336394E+09	0.468750E+07	0.187500E+08	0.177900E+08
Row	14	-127250.0	-0.301625E+07	0.136534E+10	-0.229375E+07	-0.380000E+07	577500.0
Row	15	0	0	0	0	0	0

Row	7	0.775625E+10					
Row	8	0.343012E+10	0.455431E+10				
Row	9	-0.823281E+09	-0.357216E+09	0.122473E+10			
Row	10	0.193906E+10	0.857531E+09	0.804688E+07	0.290859E+10		
Row	11	0.193906E+10	0.857531E+09	-0.823281E+09	0.193906E+10	0.193906E+10	
Row	12	0.911719E+09	0.806719E+09	-0.383297E+09	0.911719E+09	0.911719E+09	0.857531E+09
Row	13	0.364687E+10	0.322687E+10	-0.383297E+09	0.911719E+09	0.911719E+09	0.857531E+09
Row	14	-0.205016E+10	-0.336394E+09	0.511906E+09	-0.823281E+09	-0.823281E+09	-0.383297E+09
Row	15	0	0	0	0	0	0

Row	13	0.343012E+10		
Row	14	-0.368156E+09	0.250450E+10	
Row	15	0	0	1.0