RESPONSE SURFACE METHODS OF FITTING

STOCHASTIC BIOLOGICAL MODELS

A thesis presented in partial
fulfilment of the requirements
for the degree of M.Sc.
in Statistics at
Massey University

CATHERINE ANN MACKEN

1972

ABSTRACT

Response surface methods are discussed, with emphasis on the particular experimentation problems encountered in their use. A brief outline of simulation and modelling is given. This includes an indication of the role of randomness.

Two specific uses of computer simulation of biological phenomena are considered. The first is fitting growth curves to some cell growth data. This was done largely to develop techniques. The second and more significant use is in fitting stochastic selection values to some genotypic frequency data. To date, only deterministic estimates have been found from this data.

Attention is given to the careful design of simulation experiments, in order to reduce the number of simulation runs needed. Response surface methods were used and proved to be efficient experimentation techniques.

Acknowledgements

I would particularly like to express my appreciation of the very valuable guidance in the preparation of this thesis from my supervisor, Dr. B.S. Weir, of Massey University.

Thanks go to Dr. C.R. Boswell and Mr. D.J. Wilson of the Massey University Computer Unit for their help and suggestions on the computer programs; for the same reason, thanks also go to Mr. C.A. Freyberg, a Masterate student at Massey University.

Finally, I would like to thank Mrs. B. Robertson and Mrs. I. Cornish, for the typing of this thesis; Mr. A.E. Weber, for the printing of the programs; and Mr. W.D. Halford, for attending to the final copying.

TABLE OF CONTENTS

INTRODUCTION

A lack of exact analytical solutions to a mathematical system implies that numerical methods are needed to be able to study the system.

Simulation is the technique of imitating as best as possible the behaviour of a system. Using a mathematical model of the system, the experimenter can observe the effect that a different set of parameter values has on the outcome of the model by running a simulation trial using those parameter values. This technique has become practical since the advent of high speed computers. Stochastic models which previously defied solution by the mathematical analysts can now be studied by simulation.

The experimenter generally aims to estimate those values of the parameters which make the model as close to the real life situation as possible. Some criterion is needed for stating just how close the model is. If the simulated data is compared with observed data from the real life system, then the distance between them would be a measure of the goodness of fit of the model. The experimenter thus wishes to estimate those values of the parameters which make the distance as short as possible. An average distance must be taken to account for the variation in a stochastic model.

Thus experimentation, particularly on a stochastic model, involves many simulation runs. It is desirable to keep the number of runs down as much as possible. This means that it is important to plan experimentation so that the least distance is found with maximum efficiency. There are several alternative plans of experimentation available. Response surface methods were chosen for the following study since they involve experimental designs which are economical of simulation runs.

Literature Survey

Hoel and Mitchell

Hoel and Mitchell's (1971) paper first brought to notice the problem of fitting stochastic models using response surface methods. They proposed three competing stochastic models for the growth of a cell population and studied the goodness-of-fit of each model to the experimental data by measuring the sum of the squared differences between the simulated trials and the experimental data. They viewed the expectation of this distance as a response surface over the parameter space of the model, then using response surface methods optimized the fit of the model. The competing models were fitted to some data of Kubitschek (1962) on the growth of colonies of E.coli cells.

Response Surface Methods

A variable classified as a response can be explained or predicted by means of a functional relationship with a prespecified number of independent variables called factors. The functional relationship defines a response surface and measures of the response taken at different factor levels are points on this surface. Response surface methods provide a means of studying the functional relationship.

Initial interest in the use of response surface methodology was generated by Box and Wilson (1951). They first set forth the fundamentals and underlying philosophy of the use of this package of techniques and Box (1952) later extended this work for linear models. Davies (1956) edited an important textbook with a chapter dealing with the exploration of response surfaces.

There has been extensive development of second order designs. Box and Hunter (1957) studied rotatable second-order designs in general and central composite designs in particular. Hunter (1954) discussed the problem of finding a stationary point on a fitted second-order response surface and pointed out that a general second-order response surface could be transformed into a canonical form. Box and Hunter (1954) developed a method of setting a confidence region on this stationary point.

Box and Draper (1959) considered the problem of choosing a design such that a polynomial of degree d_1 might be most closely fitted to a response surface whose true representation is a polynomial of degree $d_2 > d_1$. Subject to this condition they chose their designs such that inadequate fit of the closest possible polynomial representation had a high chance of detection.

Since Box and Draper's, many other papers have been published on this subject. Hill and Hunter (1966) gave a review of the literature with particular emphasis on applications of the methodology. More recent publications were by V.J. Thomas (1971) who, in his M.Sc. thesis, concentrated on second-order designs including conditions for orthogonality of estimates; and by Myers (1971) whose textbook gave a comprehensive study of response surface methodology.

Response Surface Methods and Simulation

Modern use of the word 'simulation' traces its origin to the work of von Neumann and Ulam in the late 1940's when they coined the term "Monte Carlo analysis" to apply to a mathematical technique they used to solve certain involved nuclear-shielding problems. An interesting history of the technique is given in Hammersley and Handscomb (1964). In the early 1950's, the advent of high speed computers made simulation much more feasible. It is now a standard technique dealt with in many texts, including that of Naylor, Balintfy, Burdick and Chu (1966).

Computer simulation techniques have made it possible to perform a type of pseudo-experiment in areas where real-world experiments were otherwise impossible or impractical.

Simulation has also enabled study of models for which the nature of the model as much as the nature of the equations prohibits analytical solution of the equations. Such a situation may arise, for example, upon introduction of stochastic variation to parameters of a model, thus making closed forms for maximum likelihood parameter estimates not only difficult but no longer possible to obtain.

Hence an increasing concern with experimental design, response surface methods in particular.

Hufschmidt (1962) analysed, using response surface methods, the response surface obtained from simulation of a simplified river-basin system. He gave in detail an account of the complete experimental plan undertaken. Burdick and Naylor (1969) gave a general discussion of response surface methods applied to problems in Economics. They used simulation to study a model in a situation where real-world experiments would not have been feasible. Hoel and Mitchell (1971) used simulation and response surface methods to fit a model to some experimental data. Hunter and Naylor (1970) referred to a specific example in order to discuss in detail the experimental design problems encountered when using simulation to explore response surfaces.

Selection

Allard, Kahler and Weir (in press) used genotypic frequency data from barley populations to obtain maximum likelihood estimates of selective values. They made selective value estimates from a pair of consecutive generations, then averaged these estimates over several pairs of generations.

The next step might be to study the effect of allowing stochastic variation of selective values. Jain and Marshall (1968) reviewed the literature and found support for the idea of varying selection values. They examined by means of computer simulation the effect on genotypic equilibria of random fluctuations from generation to generation in selective values. They concentrated on values distributed according to a normal distribution. Barker and Butcher (1966) also studied the effect of generation - to - generation fluctuations in selective values. They chose selective values from a uniform distribution and, using simulation, observed quasi-fixation of genes in a population.