INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Simulation-based parameter estimation for multivariate distributions

Maa, Jen-Fue, Ph.D.
The Ohio State University, 1993
Simulation-based Parameter Estimation for Multivariate Distributions

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jen-Fue Maa, B. S., M. S.

The Ohio State University
1993

Dissertation Committee:

Professor Dennis K. Pearl
Professor Robert Bartoszyński
Professor David J. Horn

Approved By

Adviser
Department of Statistics
Acknowledgments

I sincerely appreciate Professor Dennis K. Pearl and Professor Robert Bartoszyński for their guidance, insight and encouragement throughout this research. I also thank Professor David Horn for his suggestions and comments. I am deeply grateful to my wife, Wei May Yen for her patience and continual support during my graduate study. I want to thank Doryce K. Lewis and Earl A. Lewis for helping me in reading galleys.

This material is based upon work supported by the Cooperative State Research Service, U.S. Department of Agriculture, under Agreement No. 91-37302-6745 and by Professor David J. Horn, Professor Dennis K. Pearl, and Professor Robert Bartoszyński at The Ohio State University.
VITA

March 4, 1962 .............................................. Born in Penhue, Taiwan, R.O.C.

1985 ............................................................... B. S. National Taiwan University
Taiwan, R.O.C.

1991 ............................................................... M. S. The Ohio State University
Columbus, Ohio

1991 - Present ............................................. The Ohio State University
Columbus, Ohio

Field of Study

Major Field: Statistics

Study in simulation for stochastic predator-prey model
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................. ii

VITA .................................................................................................... iii

TABLE OF CONTENTS ..................................................................... iv

LIST OF TABLES ............................................................................... vi

LIST OF FIGURES ............................................................................. vii

CHAPTER I. Introduction ............................................................... 1

CHAPTER II. Literature Review ...................................................... 4
  2.1 Literature of Optimization Method ........................................... 4
  2.2 Literature of Multivariate Goodness-of-Fit Tests ......................... 10
  2.3 Literature of Simulation-based Estimation ................................. 13

CHAPTER III. Theory About Two Distributions Being Equal .......... 19
  3.1 Theoretical Motivation of Proposing a New Random
      Criterion ........................................................................... 19
  3.2 The Main Theorem .................................................................. 20
  3.3 Reducing the Three Distances Equalities in the Theorem .... 32
CHAPTER VII. Conclusion ........................................... 122

REFERENCES ........................................................... 126
LIST OF TABLES

Table 5.1. Medians in the boxes of Figure 5.1 and Figure 5.2 ........ 68

Table 5.2. S.B.E. and classical estimators for discrete distributions,
  m=64 ........................................................................ 69

Table 5.3. S.B.E. and classical estimators for continuous distributions,
  m=64 ........................................................................ 69

Table 5.4. Estimated variation of simulation-based estimator, m=10 .. 74

Table 5.5. Estimated variation of simulation-based estimator, m=40 .. 75

Table 6.1. Results of estimated values from different methods ........ 83

Table 6.2. S.B.E. results for egg-laying, egg maturation, and
  nymph maturation of 2SSM and PPE populations ........... 87

Table 6.3  Cucumber leaf growth rate and half-time leaf surface ...... 113
LIST OF FIGURES

Fig. 2.1. Flow chart of simulated annealing optimization method .... 9

Fig. 4.1. Example of two sets of data for comparison ............... 37

Fig. 4.2. Illustration of mapping a time sequence observed data
to points in multi-dimensions ....................................... 38

Fig. 4.3. Rings defined by the observed data ......................... 40

Fig. 4.6. Confidence level versus empirical cdf of criterion for
local search .................................................................... 50

Fig. 4.4. Flow chart of optimization method for the simulation-
based estimator ................................................................. 53

Fig. 5.1. Values of the proposed criterion when the location
parameter deviates from N(0,1) ...................................... 62

Fig. 5.2. UMVU and S.B.E. in N(2,25) and N(9,4) study (m=16) . 66

Fig. 5.3. UMVU and S.B.E. in N(2,25) and N(9,4) study (m=64) . 67

Fig. 5.4. UMVU and S.B.E. in N(\mu,\sigma^2) study with m=32 ........... 70
Fig. 5.5. Values of the proposed criterion for parameters
\(\mu\) and \(\sigma^2\) deviate from \(N(0,1)\), \(m=32\), \(n=96\) ................. 71

Fig. 6.1. Histograms of Gibbs sampling posteriors ....................... 82

Fig. 6.2. PPE female adults survival under starvation ..................... 86

Fig. 6.3. 2SSM populations at 20\(^0\)C before 2 PPE females released . 89

Fig. 6.4. 2SSM populations at 22.5\(^0\)C before 2 PPE females released 90

Fig. 6.5. 2SSM populations at 25\(^0\)C before 2 PPE females released . 91

Fig. 6.6. 2SSM populations at 27.5\(^0\)C before 2 PPE females released 92

Fig. 6.7. 2SSM populations at 30\(^0\)C before 2 PPE females released . 93

Fig. 6.8. 2SSM populations at 20\(^0\)C after 2 PPE females released ... 94

Fig. 6.9. 2SSM populations at 22.5\(^0\)C after 2 PPE females released . 95

Fig. 6.10. 2SSM populations at 25\(^0\)C after 2 PPE females released ... 96

Fig. 6.11. 2SSM populations at 27.5\(^0\)C after 2 PPE females released . 97

Fig. 6.12. 2SSM populations at 30\(^0\)C after 2 PPE females released ... 98

Fig. 6.13. PPE populations at 20\(^0\)C ............................................. 99

Fig. 6.14. PPE populations at 22.5\(^0\)C ......................................... 100
Fig. 6.15. PPE populations at 25°C ........................................ 101

Fig. 6.16. PPE populations at 27.5°C .................................... 102

Fig. 6.17. PPE populations at 30°C ....................................... 103

Fig. 6.18. 2SSM egg populations on 4 lima bean plants .......... 106

Fig. 6.19. 2SSM nymph populations on 4 lima bean plants ........ 107

Fig. 6.20. 2SSM male populations on 4 lima bean plants .......... 108

Fig. 6.21. 2SSM female populations on 4 lima bean plants ........ 109

Fig. 6.23. Hockey Stick model for cucumber maximum leaf surface .. 112

Fig. 6.24. Cucumber leaf growth curve .................................. 116

Fig. 6.25. Average 2SSMs at day 13, N=4 (experiment 1 vs. simulation) ........................................ 117

Fig. 6.26. Average 2SSMs at day 20, N=4 (experiment 1 vs. simulation) ........................................ 118

Fig. 6.27. Average 2SSMs at day 15, N=4 (experiment 2 vs. simulation) ........................................ 119

Fig. 6.28. Average 2SSMs at day 23, N=4 (experiment 2 vs. simulation) ........................................ 120

Fig. 6.29. Average 2SSMs at day 17, N=5 (experiment 3 vs. simulation) ........................................ 121
The technique of using simulation-based estimation originally was discussed in the 1980's. Although several articles talk about simulation-based estimation, it is still a new research area and requires more systematic analysis. Simulation-based estimation method heavily depends on computer simulation and its basic idea lies in repeating simulations from different given parameter values to find the best fitted simulated distribution for the observed data. A high-speed computer is necessary if the simulation model is very complicated.

Our motivation for developing a system of simulation-based estimation is from the USDA project (Horn et al, 1991,b) which involved stochastic modeling of a greenhouse prey-predator system, and is currently investigated by Professor David J. Horn at the Entomology Department, Professor Dennis K. Pearl and Professor Robert Bartoszyński at the Statistics Department of The Ohio State University. Here we observed greenhouse mite populations at different days on several plants from a prey-predator system, so that realistic modeling requires the estimation of many parameters. In such a complex system, the likelihood function is unavailable but the idea of simulation-based estimation technique can still be
applied by adjusting the value of parameter until the simulation fits the observed data. The primary advantage of the simulation-based method over analytical methods is that it can be used in cases where closed form solutions are unknown or difficult to get, for instance, when the form of the likelihood is very complicated or unavailable.

Once the simulation model has been built, there are two sets of data available, the simulated data and the observed data. Fitting the simulated data to the observed data can be characterized as follows:

(i) by the replicated simulated and observed data,
(ii) by a random criterion for the goodness-of-fit,
(iii) by a minimization algorithm.

For (i), a multiple-replicated observed data will yield information about the variation among experiments. In addition one can have multiple replicated simulation data which yield information about the variation from simulation. Therefore, the components of variation for these two sets of data together are variation among experiment and variation among simulations. For (ii), one will need a random criterion to measure the closeness between simulated data and observed data. For (iii), a minimization algorithm is required for the random criterion since minimizing the criterion usually does not have a closed form solution.

Chapter II reviews the literature of optimization algorithms and simulation-based estimation. It also includes the literature of multivariate goodness-of-fit tests for two sets of data which can be considered as the random criterion for the measure of closeness. Chapter III includes the motivation of the proposed fitting
criterion that is used in Chapter IV. It also introduces several theorems relating equivalence between two distributions and equivalence of the distributions derived from the interpoint distances. Chapter IV proposes a systematic method of performing the simulation-based estimation and includes a proposed random criterion for the goodness-of-fit and a proposed optimization algorithm. In addition, Chapter IV also compares the proposed method to tests and minimization algorithms in the literature. In Chapter V, a precision interval is proposed in order to obtain an accurate estimate of the simulation-based interval. Also it discusses a consistency property of the simulation-based estimation. Furthermore, Chapter V includes the empirical analysis of the simulation-based estimation method by studying the parameter estimates for common distributions. Chapter VI discusses a very complicated simulation model, a greenhouse prey-predator model, and applies the estimation technique to different greenhouse experiments. Some of them are small experiments with an available likelihood function, but usually the simulation-based estimation is the only method for the parameter estimation in large scale experiments. Finally, Chapter VII gives some concluding remarks about the new proposed simulation-based estimation and optimization methodology.
Chapter II

Literature Review

Introduction

Since high-speed digital computers became available, there have been several articles talking about simulation-based optimization and estimation. However, few of them talk about simulation-based estimation for multi-dimensional multiple-replicated data. In this chapter the review will focus on the literature of optimization algorithms and multivariate goodness-of-fit tests. Both of the issues will play an important role in the simulation-based estimation to provide the solutions for the minimization algorithm and the problem of measuring the closeness-of-fit.

2.1 Literature of Optimization Method

The literature for optimization techniques is mainly focused on optimizing deterministic objective function. The most popular classical approach for deterministic function minimization is the Newton-Raphson method. However,
deterministic search procedures, while generating monotonically improving sequences of solutions, encounter the following problems (see Anily, 1987):

(i) the final solution heavily depends on the starting point, and
(ii) deterministic methods tend to get trapped in local optimum.

In addition, the deterministic minimization type of algorithm cannot be applied to a random objective function since it will produce different function values for each simulation though the parameter values are given.

Other newly developed methods such as the method proposed by Mockus (1991) use a Bayesian approach toward global optimization to compare with the usual minimax approach. For other methods see Evtushenko (1985) and Horst (1990). The simulated annealing method was introduced by Kirkpatrick, Gelatt and Vecchi (1983). In the 1980's, using simulated Markov chains to yield observations from the corresponding stationary distribution has been adapted to the development of stochastic search for global optimization. Simulated annealing methods tend to avoid problems (i) and (ii) above by randomizing the procedure so as to allow for occasional changes that may temporarily worsen the solution.

Following are the general steps of simulated annealing algorithm for minimization (assume continuous objective function f and its support),

Step 1 - from the initial start x, select a jump δ, and let the new \( x_j = x_i + \delta \),

Step 2 - if \( x_j \) is not in the bounded support of \( f \) then generate a new \( x_j \),

\[ \Delta f = f(\text{new } x_j) - f(\text{old } x_i) \]

Step 3 - accept new \( x_j \) if \( \Delta f < 0 \),

Step 4 - accept \( x_j \) with probability \( p_{ij}(c) = \exp\{c h(f, \Delta f)\} \) if \( \Delta f > 0 \) where \( c \) is the present control parameter and \( c \) can be time dependent which will affect the convergence rate,

Step 5 - repeat the process until the user specified criterion is satisfied.

For step 1, Goldstein (1988) emphasizes that the neighborhood structure is not the only aspect of the simulated annealing algorithm that is free to be chosen to improve the performance of the algorithm. The form of the energy function may also affect the behavior of the algorithm. For example, the objective function can be \( f \) or \( \log f \). If the neighborhood size is too small relative to the size of the possible range, the Markov chain cannot move around the state space fast enough to find the minimum in a reasonable time. On the other hand, a neighborhood too large has the algorithm merely sampling randomly from a large portion of the state space. In Goldstein's paper an optimum value of jump size is illustrated by the traveling salesman example. For step 4, in order to speed up the convergence rate, the control parameter \( c \) must be investigated. If \( c \) varies with time, like \( c/\log(t) \), then the annealing process tends to reach quasi-equilibrium or escape local minimums at a much faster rate, see Hajek (1985). Anily (1987) discussed
the properties of the transition probabilities \( p_{ij}(c) \) in the parameter space and suggested a general form:

\[
p_{ij}(c) = \exp \left\{ g \frac{f_i - f_j}{c_i} \right\}
\]

where \( g \) is constant (Bohachevsky (1986) suggested \( g = -1 \)). Lundy (1985) used \( g = 0 \) and \( c_{i+1} = \frac{c_0}{1 + i \alpha c_0 / U} \) where \( 0 < \alpha \ll 1 \) and \( U \) is an upper bound for the increase in the objective function. Hajek (1988) shows that for \( c_i = \frac{c_0}{\log(1 + t)} \), the condition for convergence is that \( c_0 \) must be greater than or equal to the depth of the deepest local minimum which is not a global minimum state.

At step 5, a stopping rule must be specified. The straightforward method is to terminate the algorithm after a specified number of iterations without a move, see Brooks (1988). Another criterion by Raatikainen (1987) terminates the simulation by the relative half-length of the confidence interval, which is commonly used to control the accuracy of one-dimensional estimates. For example, the \( P^2 \)-algorithm given in Raatikainen (1986) and Heidelberger and Lewis (1984) uses \( P \left( \frac{|x - \theta|}{|x|} < \xi \right) > 1 - \alpha \), where \( \theta = \text{true(unknown)} \) value and \( \xi = \) maximum relative half-length of the confidence interval.

For the rate of convergence, let us define the transition probabilities as follows:

\[
\forall i, j \in S : P_{ij}(c_k) = \begin{cases} G_{ij}(c_k)A_{ij}(c_k) & \text{if } i \neq j \\ 1 - \sum_{l \neq i} P_{ll}(c_k) & \text{if } i = j \end{cases}
\]
where \( G_{ij}(c_k) \) denotes the probability of generating a solution at the state \( j \) to solution at the state \( i \) and \( A_{ij}(c_k) \) is the acceptance probability in step 4. It can be proved that under certain conditions (such as irreducible and aperiodic stationary Markov chain), the rate of convergence is geometric; namely, there exists \( \alpha > 0 \) and \( 0 \leq \beta < 1 \) such that:

\[
| P^n(i,j) - \pi_j | \leq \alpha \beta^n,
\]

where \( \beta \) = the second largest eigenvalue of \( P' \).

More details can be found in Lundy (1986), Aarts (1989) and Collins (1988). Although the above shows the convergence of the algorithm, it is practically difficult to find the convergence rate \( \beta \). One may be still not sure about whether it is the right time to stop the minimization process. Figure 2.1 is the flowchart of simulated annealing methods.

The trade off between the deterministic approach and simulated annealing is that annealing method can avoid being trapped in the local minimum and the objective function doesn't have to be continuous or differentiable, but it will be more time consuming for the searching process. The classical numerical approach is the choice if one has a simple objective function like monotonicity of first derivative. Simulated annealing will be favored if the simulation itself is quick or the function is very complicated, especially with many sharp local minimal in high dimensions. In addition to the above, when the objective criterion is a random function, one cannot use the deterministic minimization algorithm since it will produce different function values at each time of simulation for fixed given parameter value. Other global optimization schemes for deterministic
functions in the literature are given by Olssen and Nelson (1975), Bohachevsky (1986), and Levy and Montalvo (1985).

Figure 2.1  Flow chart of simulated annealing optimization method
2.2 Literature of Multivariate Goodness-of-fit Tests

Multivariate tests are a fundamental problem of data analysis. In many areas, there has been a long-standing need for robust multidimensional goodness-of-fit tests (see Bickel, 1983).

For the multivariate one-sample problem, the classical procedure is Hotelling's $T^2$ which assumes that the underlying population is a $p$-variate normal distribution. Many nonparametric tests had been proposed such as the Component Sign Test (see Bickel, 1965), which uses a sign statistic for each component of the vectors and combines them in a quadratic form. Bickel demonstrated that this statistic may not perform well when there are substantial correlations among variates in the vectors.

Another test statistic is the Bivariate Sign Test proposed by Hodges (1955) and Blumen (1958) who show that its local Bahadur efficiency is .636 when the underlying distribution is bivariate normal. Oja and Nyblm (1989) generalized Hodges's and Blumen's test and proposed a class of invariant sign tests. They showed that among them, Blumen's procedure is optimal against an elliptic alternative. Brown and Hettmansperger (1987) use Oja's generalized multivariate median and defined a multivariate quantile to develop a bivariate test for the location problem in the one- and two-sample setting. This sign test appears to behave in a large sample setting much as Blumen's test and has a bounded influence function, hence it is highly robust. Efficiencies for Blumen's bivariate rank sum and sign-rank procedures have not yet been exhibited, see Randles (1990,a). Another distribution-free multivariate sign test based on interdirections was proposed by Randles (1989). This test statistic is based on the number of
pairs of points that fall on opposite sides of a series of data-based hyperplanes. If
the observations are from distributions with elliptical directions, then the statistic
has a limiting chi-square null distribution. It also has a small-sample distribution-
free and invariant properties.

Methods for detecting clustering can be grouped into three categories (see
Cuzick, 1990):

(a) methods based on cell counts,
(b) methods based on adjacencies of cells with high counts, and
(c) the distance method.

Cell count methods are based on a $\chi^2$ test for heterogeneity of cell counts on the
empirical distribution function. The cells are usually defined by pre-existing
administrative boundaries. In addition, the computation of the expected counts
requires detailed knowledge of the population. An article by the Committee on
Medical Aspects of Radiation in the Environment (1988) has developed an
algorithm for assembling small areas into regions of roughly constant expected
numbers. Openshaw et al. (1988) considered a method of multiple overlapping
circles in which the observed and expected numbers are computed for circles with
centers at every point of a fine grid and for a variety of radii, (see Cuzick, 1990).

The problems with cell-count methods are in the difficulty of obtaining
accurate expected cell counts and in the definition of the cell boundaries. When
expected numbers are much less than unity, a single case will generate a
significant result for that cell. Also, when cases are rare, failure to account for
nearby cases in neighboring cells represents a significant loss of power.
Khmaladze (1988) used the empirical process to approach a goodness-of-fit test in
k-dimension. Schilling (1983) proposed a practicable multidimensional goodness-of-fit test based on nearest neighbor and obtained its asymptotic behavior, but the test requires very large samples.

For the multivariate two-sample testing problem, Friedman and Rafsky (1979) introduced a nonparametric two-sample test based on the minimal spanning tree of the pooled sample. It is a multivariate generalization of the Wald-Wolfowitz and Smirnov univariate two-sample test. The tests proposed by Friedman and Steppel (1974), and Rogers (1976) for the multivariate two-sample problem is based on the proportion of nearest neighbors from the same sample for each index point. Schilling (1986) demonstrated the limiting normality of a generalization of this statistics which examines the k-nearest neighbor. In particular, the test that rejects the null if $T_{kn}$ is large, where

$$T_{kn} = \frac{1}{kn} \sum_{i=1}^{n} \sum_{r=1}^{k} I_i(r),$$

and $I_i(r) = \begin{cases} 1 & \text{if } \text{NN}_i(r) \text{ and } Z_i \text{ from the same sample} \\ 0 & \text{otherwise} \end{cases}$

Here, $\text{NN}_i(r) = r^{th}$ nearest neighbor to the sample, and $Z_i, Z_1, Z_2, \ldots, Z_{k+n}$ is the pooled sample. Schilling also proposed two weighted versions of above test. The first is weighted by a function of rank of the neighbor. The optimal system of weights does not depend on the alternative and thus cannot be tuned for particular alternatives to $H_0$. The second version is weighted by the reference point for continuous weight functions $w$, but it is not distribution free since $w$ can be designed to perform well against specific types of alternatives. Henze (1988) gives more detail about the testing statistic and concludes that practical
implementation of this approximate permutation test required the determination of all k-nearest neighbors. Cuzick (1990) analytically evaluated several tests including the nearest-neighbors test, the run-length test, the inverse sampling k-nearest neighbors test and the use of general rank-based procedures, then applied them all to a data set on the locations of cases of childhood leukemia and lymphoma.

2.3 Literature of Simulation-based Estimation

Thompson (1987) proposed an algorithm for simulation-based estimation of parameters characterizing a stochastic process which is summarized as follows:

Divide the time axis into k bins and get the multinomial distribution $\pi_1, \ldots, \pi_k$.

Let

$\hat{P}_j(\theta) =$ proportion of observed data in j-th bin, and let

$\hat{P}_{nj}(\theta) =$ proportion of simulated data in j-th bin with simulation size n

for $j=1,2,\ldots,k$.

The simulation-based estimation is to pick $\theta$ which minimizes the criterion:

$$S(\theta) = \sum_{j=1}^{k} \frac{(\hat{P}_{nj} - \hat{P}_j)^2}{\hat{P}_j}.$$ 

Once the algorithm has converged to a value, say $\hat{\theta}$, one can use this value to generate $M$ sets of simulated data each with size $n$ and get $S_1(\hat{\theta})$, $S_2(\hat{\theta})$, $\ldots$, $S_M(\hat{\theta})$. With roughly 95% certainty
\[ S(\theta) = \overline{S} \pm \frac{2}{\sqrt{M}} \sigma_s, \]

where \( \overline{S} = \frac{1}{M} \sum_{j=1}^{M} S_j(\hat{\theta}) \) and \( \sigma_s = \frac{1}{M} \sum_{j=1}^{M} (S_j(\hat{\theta}) - \overline{S})^2. \)

Next, using \( \hat{\theta} \) as the center of rotatable design, one can fit the quadratic curve \( S(\theta) = A + B\theta + C\theta^2. \) The 95% confidence set for \( \theta \) can be approximated by

\[ S_0(\hat{\theta}) - \frac{2}{\sqrt{M}} \sigma_s \leq A + B\theta + C\theta^2 \leq S_0(\hat{\theta}) + \frac{2}{\sqrt{M}} \sigma_s, \]

where \( S_0(\hat{\theta}) \) is obtained from observed data. The determination of the relative sizes of the \( k \) bins used to discretize the data was not specified. One might try to set the \( p \)’s such that the binning scheme has equal numbers of observations in each bin, although it may not be easy. Thompson used an optimization method for looking for the minimization solution from Chandler (1969) called STEPIT which is robust although computationally slow. A brief outline of STEPIT is as follows:

STEPIT varies each parameter value individually up and down. If either variation yields an improvement, the step size of each jump is doubled and another step is taken. The number of such steps allowed is limited. When a local minimum has been bracketed by this process, quadratic interpolation is used to attempt to refine the position of the minimum. If both up and down parameter values lead to no improvement, the step sizes is decreased. The algorithm terminates when the step size becomes smaller than a user-specified minimum. However, a generalization of the step size stopping criterion for multi-dimensional parameters version was not specified in the algorithm.
One shortcoming of Thompson's method is that the criterion for comparing the simulation proportions and the observed proportion is not well defined. For example, how to choose the intervals to count the proportions? What is the generalized criterion for multidimensional data? Also, the confidence set is approached by approximation. Inverting the quadratic curve to find the confidence set may not perform well, even with large M (the number of data sets from simulation). Also, there is no guarantee of convergence to $\theta_0$ in the STEPIT procedure.

The next example is an application of the simulation-based estimator proposed by Pakes (1986). Pakes fits an optimal stopping model to patent renewal data. Each year patent holders have to decide whether to pay a renewal fee in order to keep their patents in force. Let

$$\pi(\theta) = \text{the expected proportion dropping out at each age, and}$$

$$p_n = \text{observed drop out proportions.}$$

If the $\pi(\cdot)$ had been an easily calculable function of $\theta$, any of the usual estimation procedures for the multinomial distribution could have been used to estimate $\theta$. Pakes substitutes a simulation estimator, $\pi_S(\theta)$ based on an average of $s$ simulations, for $\pi(\theta)$. For fixed $\theta$, the simulation estimate was obtained by taking $ns$ random draws from simulation, and simply counting up the proportions that dropped out at each age. Then the estimator is the one which minimizes

$$\| G_n(\theta) \| = \| p_n - \pi_S(\theta) \|.$$

This is the simulated analogue of the estimator which minimize
an estimator which satisfies the conditions of theorems in Chapter V for the consistency and the asymptotic normality. The multivariate central limit theorem by Breiman (1968) guarantees that

\[ \sqrt{n} \left[ p_n - \pi(\theta_0) \right] \rightarrow N(0, V) \quad \text{where } V = \text{diag} [\pi(\theta_0)] - \pi(\theta_0)\pi(\theta_0)'. \]

Since, at the true \( \theta_0 \), the simulation mimics the data for a sample size of \( sn \),

\[ \sqrt{n} \left[ \pi_s(\theta_0) - \pi(\theta_0) \right] \text{ has the limit distribution } N(0, s^{-1}V). \]

Moreover, since the data generating process and the simulation process are independent,

\[ \sqrt{n} \, G_n(\theta_0) = \sqrt{n} \left[ p_n - \pi(\theta_0) \right] - \sqrt{n} \left[ \pi_s(\theta_0) - \pi(\theta_0) \right] \rightarrow N(0, (1 + s^{-1})V). \]

The limit distribution of \( \sqrt{n} \, G_n(\theta_0) \) differs from that of \( \sqrt{n} \, g_n(\theta_0) \) (which is the true distribution of the observed data = \( \sqrt{n} \left[ p_n - \pi(\theta_0) \right] \)) only through the presence of the scalar \( (1 + s^{-1}) \), which reflects the extra independent source of randomness generated by the simulation process. The estimator formed by minimizing a weighted version, \( \| A_n(\theta) G_n(\theta) \| \), has the property that

\[ \sqrt{n} \left( \hat{\theta}_0 - \theta_0 \right) \rightarrow N(0, (1 + s^{-1})M(A)), \]

where \( M(A) = (\Gamma^A\Lambda\Gamma')^{-1}\Gamma^A A \Lambda^2 A \Gamma (\Gamma^A A \Gamma')^{-1} \),

\( A \) is the probability limit of \( A_n(\theta_0) \), and \( \Gamma \) is the derivative matrix of \( G(\cdot) \).
For \( A_n(\theta) = \text{diag}[\pi_\theta(\theta)^{-1/2}] \) or \( A_n(\theta) = \text{diag}[p_n^{-1/2}] \), these are the weighting metrics used by the method of minimum chi-square and modified minimum chi-square, respectively. In both cases \( A = \text{diag}(\pi(\theta_0)^{-1/2}) \). Theil (1984), shows that \( M(A) - M(\text{diag}[\pi(\theta_0)^{-1/2}]) \) is positive semi-definite for every nonsingular \( A \), so the use of \( A = \text{diag}(\pi(\theta_0)^{-1/2}) \) leads to an asymptotically efficient estimator for any fixed value of \( s \). A second example in the Pakes (1989) paper is from McFadden (1989) which estimates the parameters of a multinomial probit model using the simulation method.

The above simulation method is a generalization of Thompson's method and does not required to calculate the theoretical expectation, \( \pi(\theta) \), which may be sometimes difficult to get or even not available. Instead, we can use the simulated \( \hat{\pi}_\theta(\theta) \) to obtain \( \hat{\theta}_0 \) by minimizing \( \|G_n(\theta)\| \) if the simulation can produce a good estimate, \( G_n(\theta) \), of \( g_n(\theta) \). Second, Pakes shows that under suitable conditions, the simulation estimator \( \hat{\theta}_0 \) will be a consistent estimator of \( \theta_0 \) when the observed sample size goes to infinity, (see Chapter V). The limiting normality of \( G_n(\hat{\theta}_0) \) and \( \hat{\theta}_0 \) gives us knowledge of the estimator and the criterion function.

Pakes did not specify a general form of the criterion \( \|G_n(\theta)\| \), so the user has to define it from case to case. It would be nice if there is a distribution-free criterion which would work for all cases and all dimensional data. For instance, in Pakes (1986) example, we may encounter that the observed sample size \( n \) is too small to count the observed sample proportion in each age, especially if we have data from higher dimensions. In order to get consistency and limiting normality for the simulation estimator, we must have a very large sample of observations. There is no formula for small sample size. In addition, one may not be able to get
the exact form of the limiting distribution for $\hat{\theta}_0$ and $G_n(\cdot)$, if the exact form of $G(\cdot)$ is not available in terms of $\theta$ for differentiating $G(\cdot)$ in Theorem 6.2. Next, it is crucial to get a good algorithm so that we can approach the minimizing solution of $\|G_n(\cdot)\|$. There will be no closed form for $G_n(\cdot)$ in the case of simulation. After all, it is nice to have the asymptotic properties for a simulation-based estimator from such a general criterion under some reasonable regulatory conditions.

Lee (1991) proposed a simulation-based estimation method for time-series models which applies Pakes method with some modification. As in Chapter V, the asymptotic covariance matrix for the estimator depends on the choice of the weights in the fitting criterion. Lee discussed the optimal weights choice to yield the smallest asymptotic covariance matrix for the estimator. By increasing the sample size of the simulation, one can reduce the randomness in the simulation and allow the reduction of the variance of the estimator. Again, the minimization algorithm is not discussed. Furthermore, the author emphasized that the algorithm that they used would have difficulty sorting out a change in the criterion function due to a variation in the random draw from a change due to a modification in the parameters. Akinson (1989) avoids the difficulty of stochastic optimization by the "fixed-seed" method, where all the random numbers needed to simulate the data set are generated at the beginning of the estimation procedure. Thus variation in the objective function are due solely to changes in the parameter values which allows the application of standard optimization techniques.
Chapter III

Theory About Two Distributions Being Equal

Introduction

From previous chapters, we have seen the characterization of simulation-based estimation. In this chapter, a theoretical motivation for the development of a new multivariate goodness-of-fit random criterion is shown. A formulation for the characterization of two distributions being equal is described in section 3.2. The rest of this chapter discusses the conditions for the theorem to be true.

3.1 Theoretical Motivation of Proposing a New Random Criterion

A functional mapping (from $\mathbb{R}^k \times \mathbb{R}^k$ to $\mathbb{R}$) of the multi-dimensional data into a one-dimensional value plays an important role in the multivariate two sample testing problem. The mapping function which is commonly used in this problem is the distance. For example, Friedman and Steppel (1974) and Rogers (1976) proposed two-sample tests which are based on nearest neighbors. Friedman and
Rafsky (1979) proposed a two-sample test based on the minimal spanning tree. Atkinson (1989) measured the goodness-of-fit by the squared difference between the reciprocal of the distance within the observed data and the reciprocal of the distance within the simulated data. Other distance related tests are proposed by Schilling (1986) and Henze (1988). Therefore, for testing the equality of two distributions, one often compares the interpoint distance of two points from either the two distributions.

This technique for reducing the dimensionality of the multivariate two-sample problem can be justified as follows. Suppose $X_1, X_2, \ldots$ are iid from $F$ and $Y_1, Y_2, \ldots$ are iid from $G$, and $X$'s and $Y$'s are independent. We want to see whether

$$h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \overset{d}{=} h(X_3, Y_3) \text{ iff } F = G \quad (3.1)$$

where $h$ is a function that maps two points in $\mathbb{R}^k$ to a one-dimensional value in $\mathbb{R}$. In section 3.2, conditions on the function $h$ such that (3.1) is true for some distributions are discussed. In section 3.3, additional conditions imposed to reduce the three distances equalities in (3.1) are considered.

3.2 The Main Theorem

In the sequel, we let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be iid random variables from $k$-dimensional distributions $F$ and $G$, respectively. We assume that $X_i$ and $Y_j$ are independent for all $i, j$. Also, we use the symbol $\overset{d}{=}$ to mean the equality of distributions.
Theorem 3.1. Assume that F and G are discrete distributions on \( \mathbb{R}^k \), and let \( h(x,y) \) be any real-valued non-negative function on \( \mathbb{R}^k \times \mathbb{R}^k \) such that \( h(x,y) = 0 \) iff \( x = y \). Then

\[
h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \overset{d}{=} h(X_3, Y_3) \quad \text{iff} \quad F = G.
\]

PROOF. It is trivial that \( F = G \) implies \( h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \overset{d}{=} h(X_3, Y_3) \), so one needs only to prove the converse. From \( h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \overset{d}{=} h(X_3, Y_3) \), we have

\[
P(h(X_1, X_2) \leq t) = P(h(Y_1, Y_2) \leq t) = P(h(X_3, Y_3) \leq t) \quad \text{for all} \ t \geq 0.
\]

In particular, when \( t=0 \),

\[
P(X_1 = X_2) = P(Y_1 = Y_2) = P(X_3 = Y_3), \tag{3.2}
\]

since \( h(x,y) = 0 \) implies \( x = y \).

Let the probability mass function of \( X_1 \) and \( Y_1 \) be

\[
P( X = x_i ) = \pi_i, \ i = 1, 2, \ldots \quad \text{and} \quad P( Y = y_i ) = \gamma_i, \ i = 1, 2, \ldots
\]

with \( \sum \pi_i = \sum \gamma_i = 1 \), where \( x_i \) and \( y_i \) are \( k \)-dimensional vectors.

Let \( S_1 = \{ x_i, \ i=1, 2, \ldots \} \) and \( S_2 = \{ y_i, \ i=1, 2, \ldots \} \). Suppose that \( S_1 \cap S_2 \) consists of the matched pairs \( x_{ij} = y_{ij} \). Thus, from (3.2)

\[
P(X_1 = X_2) = \sum_{x_{ij} \in S_1} \pi_{ij}^2
= P(Y_1 = Y_2) = \sum_{y_{ij} \in S_2} \gamma_{ij}^2
= P(X_3 = Y_3) = \sum_{x_{ij} \in S_1 \cap S_2} \pi_{ij} \gamma_{ij}. \tag{3.3}
\]
and it follows that \( S_1 \cap S_2 \neq \emptyset \) since \( P(X_1 = X_2) > 0 \) from the assumption of discreteness. Omitting the terms from outside \( S_1 \cap S_2 \) and using the Cauchy inequality, one has

\[
\sum_{x_i \in S_1} \pi_j^2 \sum_{y_j \in S_2} \gamma_j^2 \geq \sum_{x_{ij} \in S_1 \cap S_2} \pi_{ij}^2 \sum_{y_{ij} \in S_1 \cap S_2} \gamma_{ij}^2 \geq \left( \sum_{x_{ij} \in S_1 \cap S_2} \pi_{ij} \gamma_{ij} \right)^2.
\]

Since (3.3) implies that the first and last terms in the above are equal, we obtain

\[
\sum_{x_i \in S_1} \pi_j^2 \sum_{y_j \in S_2} \gamma_j^2 = \sum_{x_{ij} \in S_1 \cap S_2} \pi_{ij}^2 \sum_{y_{ij} \in S_1 \cap S_2} \gamma_{ij}^2 = \left( \sum_{x_{ij} \in S_1 \cap S_2} \pi_{ij} \gamma_{ij} \right)^2.
\]

Since the \( \pi \)'s and \( \gamma \)'s are nonnegative, only \( \pi_{ij} \), \( \gamma_{ij} \) for \( x_{ij} \in S_1 \cap S_2 \) can be positive. In addition, equality in the Cauchy formula means \( \pi_{ij} = c \gamma_{ij} \) for \( x_{ij} \in S_1 \cap S_2 \), and \( c \) must equal 1 since

\[
\sum_{x_{ij} \in S_1 \cap S_2} \pi_{ij} = 1 \quad \text{and} \quad \sum_{y_{ij} \in S_1 \cap S_2} \gamma_{ij} = 1. \quad \blacksquare
\]

**Remark 3.1.** Observe that the proof does not require that the random variables \( X_i \) and \( Y_j \) be defined on \( \mathbb{R}^k \): they can be, for example, multivariate nominal random variables.

The proof above relies on the fact that the probability of two iid discrete random variables assuming the same value is positive. Therefore, it cannot be applied to the continuous case. However, the theorem is also proved in the
continuous case with some restriction on the density function. We start from the following lemma.

**Lemma 3.1.** Let the cdf $F$ and the cdf $G$ have bounded density functions $f$ and $g$, respectively which have continuous first partials almost everywhere. If $a_i < b_i$, for $i = 1$ to $k$, then

$$
\lim_{t \to 0} \frac{\sum_{i=1}^{k} (b_i - a_i) \int f(x) g(x) dx}{t^k} = \prod_{i=1}^{k} (b_i - a_i)
$$

**PROOF.** We will provide the proof for $k=2$. For general $k$ the proof follows by analogy. Conditioning on the value of $Y=(Y_1,Y_2)=(y_1,y_2)$ and using the fact that the random variables are continuous, we may write

$$
P(a_1 t < X_1 - Y_1 < b_1 t, a_2 t < X_2 - Y_2 < b_2 t) = \int_{y_1}^{y_2} \int_{y_1}^{y_2} \prod_{i=1}^{k} (b_i - a_i) \int f(x) g(x) dx.
$$

Letting $\Delta_i = \frac{b_i - a_i}{2}$, $y_i^* = y_i + \frac{b_i + a_i}{2}$, $i = 1, 2$, the integrand can be rewritten as follows:

$$
\frac{1}{t^2} \left\{ \left[ F(y_1^* + \Delta_1 t, y_2^* + \Delta_2 t) - F(y_1^* - \Delta_1 t, y_2^* - \Delta_2 t) \right] - \left[ F(y_2^* - \Delta_1 t, y_2^* + \Delta_2 t) + F(y_1^* + \Delta_1 t, y_2^* - \Delta_2 t) \right] \right\} g(y_1, y_2)
$$

Using a Taylor expansion, and the assumed a.e. continuity of the partial derivatives, it is easy to see that the limit of (3.5) as $t \to 0$ is $4 \Delta_1 \Delta_2 f(y_1, y_2) g(y_1, y_2)$. 

The boundedness of the density allows us to use the Dominated Convergence Theorem to pass to the limit under the integral sign. Thus

$$\lim_{t \to 0} \frac{P(a_it < X_i - Y_i < b_it, i=1,2)}{t^2} = (b_1-a_1)(b_2-a_2) \int \int f(y_1,y_2)g(y_1,y_2)dy_1dy_2. \quad \blacksquare$$

**Remark 3.2.** In Lemma 3.1, the condition of the boundedness of the density function is applied when we want to move the limit into the integral. Clearly any conditions on the density function which allows us to pass the limit under the integral in (3.4) would be appropriate.

**Lemma 3.2.** Let $X$ and $Y$ have densities $f$ and $g$ respectively which satisfy the conditions of Lemma 3.1. Let $h : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ be a nonnegative continuous function such that $h(x,y) = 0$ iff $x = y$, and $h(ax+b,ay+b) = l|a|h(x,y)$, $\forall a \in \mathbb{R}$, $\forall b \in \mathbb{R}^k$. Then

$$\lim_{t \to 0} \frac{P(h(X,Y)<t)}{t^k} = \alpha \int f(y)g(y)dy$$

where $\alpha =$volume of region defined by $\{x: h(x,0)\leq 1\}$, that is, $\alpha = \int_{\{h(x,0)<1\}} dx$.

**PROOF.** For any fixed $y$, let $A(y) = \{x: h(x,y)<t\}$ for $t>0$. Since $h$ is continuous and $h(y,y) = 0$, $A(y)$ is an open region in $\mathbb{R}^k$ containing $y$. Consequently, $A(y)$ is representable as $A(y) = \bigcup_{j=1} B_j \cup \mathcal{N}$, a union of a countable family of disjoint open rectangles $B_j$ and a set $\mathcal{N}$ of Lebesgue measure zero. Since the random variables under consideration are continuous, we have
\[ P(A(y)) = \sum_{j=1}^{\infty} P(B_j) = \sum_{j=1}^{\infty} P(a_{jm} < X_m < b_{jm}, m = 1, \ldots, k), \]

where \(a_{jm}, b_{jm}\) possibly depend linearly on \(y\) and \(t\). Therefore

\[
P(h(X,Y) < t) \quad t^k
\]

\[= \int P(h(X,y) < t) g(y) \, dy\]

\[= \int P(\frac{A(y)}{t^k} < t) g(y) \, dy\]

\[= \int \sum_{j=1}^{\infty} P(a_{jm} < X_m < b_{jm}, m = 1, \ldots, k) \quad t^k g(y) \, dy.\]

Interchanging the integral and summation signs and recentering the rectangles, the last integral becomes

\[
\sum_{j=1}^{\infty} \int P(\omega_{jm} t < X_m - y_m < \delta_{jm} t, \forall m) \quad t^k g(y) \, dy
\]

where \(\omega_{jm} = \frac{a_{jm} - y_m}{t}\) and \(\delta_{jm} = \frac{b_{jm} - y_m}{t}\). Now, taking limit as \(t \downarrow 0\) and using Lemma 3.1, we obtain

\[
\lim_{t \downarrow 0} \frac{P(h(X,Y) < t)}{t^k} = \left[ \sum_{j=1}^{\infty} \prod_{m=1}^{k} (\delta_{jm} - \omega_{jm}) \right] f(y) g(y) \, dy
\]

\[= \alpha \int f(y) g(y) \, dy. \quad \blacksquare\]

**Theorem 3.2.** Let \(X_1, X_2, X_3\) be iid \(k\)-dimensional random variables with density \(f\) and cdf \(F\) and let \(Y_1, Y_2, Y_3\) be iid \(k\)-dimensional random variables with density \(g\) and cdf \(G\), and suppose the \(X\)'s and \(Y\)'s are independent with both \(f\) and
g satisfying the conditions of Lemma 3.1. If h satisfies the conditions of Lemma 3.2, then

\[ h(X_1, X_2) = h(Y_1, Y_2) = h(X_3, Y_3) \text{ if and only if } F = G. \quad (3.6) \]

**PROOF.** It is trivial that \( F = G \) implies \( h(X_1, X_2) = h(Y_1, Y_2) = h(X_3, Y_3) \), so one needs only to prove the converse. From

\[ h(X_1, X_2) = h(Y_1, Y_2) = h(X_3, Y_3), \]

one has

\[ P(h(X_1, X_2) < t) = P(h(Y_1, Y_2) < t) = P(h(X_3, Y_3) < t) \text{ for any } t \geq 0. \]

Therefore

\[ \lim_{t \to 0} \frac{P(h(X_1, X_2) < t)}{t^k} = \lim_{t \to 0} \frac{P(h(Y_1, Y_2) < t)}{t^k} = \lim_{t \to 0} \frac{P(h(X_3, Y_3) < t)}{t^k}. \]

From Lemma 3.2 with \( f = g \) for the first two lines, we have

\[ \lim_{t \to 0} \frac{P(h(X_1, X_2) < t)}{(2t)^k} = \alpha \int f^2(x)dx, \]

\[ \lim_{t \to 0} \frac{P(h(Y_1, Y_2) < t)}{(2t)^k} = \alpha \int g^2(x)dx, \]

\[ \lim_{t \to 0} \frac{P(h(Y_1, Y_2) < t)}{(2t)^k} = \alpha \int f(x)g(x)dx. \]

Therefore, since \( 0 < \alpha < \infty \),

\[ \int f^2(x)dx = \int g^2(x)dx = \int f(x)g(x)dx. \quad (3.7) \]

From the Schwarz inequality, one has
\[ \int f^2(x)dx \int g^2(x)dx \geq (\int f(x)g(x)dx)^2. \] \hspace{1cm} (3.8)

But (3.7) shows that we have equality in (3.8). Thus, since \( f \) and \( g \) are both density functions, they must be identical a.e. \( \Box \)

**Remark 3.3.** The proof of Theorem 3.1 and Theorem 3.2 show that our main result is true for a wider class of situations. For example, we can allow for mixtures of continuous and discrete distributions, with any function \( h \) satisfying the conditions of Theorem 3.2. Also, in Theorem 3.2 we can widen the class of \( h \)'s by allowing continuous monotone functions of \( h \)'s satisfying the conditions of the theorem.

**Remark 3.4.** None of the equations in (3.6) can be dropped. For example, if \( X \equiv 0, Y = 0 \) or \( 1 \) with probability \( \frac{1}{2} \), then \( h(Y_1,Y_2) \equiv h(X_1,Y_3) \), but \( F \neq G. \)

**Remark 3.5.** The proof of Theorem 3.1 and Theorem 3.2 did not require the independence of all three interpoint distances. In particular, we have

\[ h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \overset{d}{=} h(X_1, Y_1) \text{ iff } F = G. \]

This may be useful in applications since it allows \( X \) to \( X \), \( Y \) to \( Y \), and \( X \) to \( Y \) distances to be computed from the same reference points.

Statement (3.6) is true under a variety of situations not covered by above. In particular, it is true when \( h(x,y) = \max_{1 \leq i \leq k} |x_i - y_i| \) under some additional restrictions.
imposed on the characteristic functions of X and Y which is described in Theorem 3.3. In fact, we believe (3.6) is true for all distribution F and G and every h which is a function of the Euclidean metric. We begin the proof of Theorem 3.3 from the following.

**Lemma 3.3.** If \( k(s) \) is an odd function, i.e., \( k(-s) = -k(s) \), then the inversion formula of the Fourier transform for

\[
G(t) = -2 \lim_{H \to \infty} \int_{-H}^{H} k(s) \sin(ts) ds
\]

is

\[
k(s) = \frac{1}{\pi} \lim_{H \to \infty} \int_{0}^{H} G(t) \sin(ts) dt.
\]

**Proof.** See Fourier integral, Papoulis(1962). ■

**Lemma 3.4.** Let \( A(s) \) and \( B(s) \) be a complex function, and \( k(s) = \frac{1}{s} (A(s) - B(s)) \). If \( k(s) \) is an odd function then

\[
\lim_{H \to \infty} \int_{-H}^{H} \frac{\sin(ts) A(s)}{s} ds = \lim_{H \to \infty} \int_{-H}^{H} \frac{\sin(ts) B(s)}{s} ds \Rightarrow A(s) = B(s).
\]

**Proof.** Let \( G(t) = \lim_{H \to \infty} \int_{-H}^{H} \frac{\sin(ts) A(s)}{s} ds \). Let \( G(t) = \lim_{H \to \infty} \int_{-H}^{H} \frac{\sin(ts) B(s)}{s} ds \). Then

\[
G(t) = \lim_{H \to \infty} \int_{-H}^{H} \frac{\sin(ts) A(s) - B(s)}{s} ds
\]

\[
= \lim_{H \to \infty} \int_{-H}^{H} \sin(ts) k(s) ds
\]

\[
= \lim_{H \to \infty} \left[ \int_{0}^{H} \sin(ts) k(s) ds + \int_{-H}^{0} \sin(ts) k(s) ds \right]
\]

(since \( k(-s) = -k(s) \), therefore, \( \sin(ts) k(s) \) is an even function)
\[
= 2 \lim_{H \to \infty} \int_0^H k(s) \sin(ts) ds.
\]

From Lemma 3.3, one has \( k(s) = \frac{1}{\pi} \lim_{H \to \infty} \int_0^H G(t) \sin(ts) dt \). Since \( G(t) = 0 \) for all \( t \), \( k(s) = 0 \). Therefore, \( A(s) = B(s) \). \( \blacksquare \)

In general, one can show that Lemma 3.4 is also true for the function \( A(s) \) and \( B(s) \) in more than one dimension situation.

**Theorem 3.3.** Let \( X_1, X_2, X_3 \) be iid \( k \) dimensional random variables with cdf \( F \) and characteristic function \( \phi \). Let \( Y_1, Y_2, Y_3 \) be iid \( k \) dimensional random variables with cdf \( G \) and characteristic function \( \varphi \). If \( h(X_i, Y_i) \) is the distance that equals the maximum of all the differences between the coordinates of the points, and both \( \phi, \varphi \) are odd functions or both are even functions, then

\[ h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \overset{d}{=} h(X_3, Y_3) \text{ if and only if } F = G \]

**PROOF.** Let \( \phi(s_1, s_2, ..., s_k) \) be the characteristic function of \( X_1 \), and \( \varphi(s_1, s_2, ..., s_k) \) be the characteristic function of \( Y_1 \). Let vector \( X_1 = (U_1, U_2, ..., U_k) \) and vector \( Y_1 = (V_1, V_2, ..., V_k) \). Then

\[
P(h(X_1, Y_1) \leq t) = P(\max_i \left| U_i - V_i \right| \leq t)
= \int P(-t+V_i \leq U_i \leq t+V_i, i = 1 \text{ to } k \mid V = v) \ dG
\]
\[
\lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \int_{-H}^{H} \int_{-H}^{H} \sin t \frac{\sin t s_1}{s_1} \sin t \frac{\sin t s_2}{s_2} \cdots \sin t \frac{\sin t s_k}{s_k} e^{-i \Sigma s_i v_i} \phi(s_1, s_2, \ldots, s_k) \, ds_1 \, ds_2 \cdots \, ds_k \, dG
\]

(from the inversion formula of the characteristic function)

Let \( S = (s_1, s_2, \ldots, s_k) \), therefore above is then

\[
\lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \int_{-H}^{H} \int_{-H}^{H} \sin t \frac{\sin t s_1}{s_1} \sin t \frac{\sin t s_2}{s_2} \cdots \sin t \frac{\sin t s_k}{s_k} \{ \int e^{-i \Sigma s_i v_i} \, dG \} \phi(S) \, dS
\]

\[
= \lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \int_{-H}^{H} \int_{-H}^{H} \sin t \frac{\sin t s_1}{s_1} \sin t \frac{\sin t s_2}{s_2} \cdots \sin t \frac{\sin t s_k}{s_k} \phi(-S) \phi(S) \, dS \quad (3.9)
\]

Similarly, one has

\[
P(\rho(X_1, X_2) \leq t) =
\lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \int_{-H}^{H} \int_{-H}^{H} \sin t \frac{\sin t s_1}{s_1} \sin t \frac{\sin t s_2}{s_2} \cdots \sin t \frac{\sin t s_k}{s_k} |\phi(S)|^2 \, dS \quad (3.10)
\]

and

\[
P(\rho(X_1, Y_2) \leq t) =
\lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \int_{-H}^{H} \int_{-H}^{H} \sin t \frac{\sin t s_1}{s_1} \sin t \frac{\sin t s_2}{s_2} \cdots \sin t \frac{\sin t s_k}{s_k} |\phi(S)|^2 \, dS \quad (3.11)
\]

Since (3.9), (3.10) and (3.11) are equal for all \( t \), from (3.9) = (3.11) one has
\[
\lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \cdots \int_{-H}^{H} \left( \prod_{i=1}^{k} \frac{\sin \frac{ts_i}{s_i}}{s_i} \right) [\varphi(-S)\varphi(S) - I\varphi(S)^2] \, dS = 0.
\]

Since both \( \varphi, \varphi \) are odd functions or both are even functions,

\[
\left( \prod_{i=1}^{k} \frac{\sin \frac{ts_i}{s_i}}{s_i} \right) [\varphi(-S)\varphi(S) - I\varphi(S)^2] \text{ is an even function. By the Lemma 3.4, one has}
\]

\[
\varphi(-S)\varphi(S) - I\varphi(S)^2 = 0, \quad \text{i.e.}
\]

\[
\varphi(S)\varphi(S) = I\varphi(S)^2 = \varphi(S)\varphi(-S).
\] (3.12)

Similarly, from (3.10) = (3.11), one has

\[
|\varphi(S)|^2 = I|\varphi(S)|^2.
\] (3.13)

(3.12) and (3.13) implies \( \varphi(S) = \varphi(S) \). ■

**Remark 3.6.** Formula (3.9) is equivalent to

\[
\lim_{H \to \infty} \frac{1}{\pi^k} \int_{-H}^{H} \cdots \int_{-H}^{H} \frac{\sin \frac{ts_1}{s_1}}{s_1} \frac{\sin \frac{ts_2}{s_2}}{s_2} \cdots \frac{\sin \frac{ts_k}{s_k}}{s_k} \varphi(S)\varphi(-S) \, dS.
\]

The odd or even characteristic function assumptions in Theorem 3.3 are used to impose Lemma 3.4. In general, any characteristic functions that satisfying (3.9), (3.10) and (3.11) will imply \( \varphi(S) = \varphi(S) \) are appropriate.
3.3 Reducing the Three Distances Equalities in the Theorem

In applying the theory of section 3.2, it would be helpful to characterize the equality of two multivariate distributions problem by simply establishing the equality of two distance distributions. In this section, we will consider the $h$ as the Euclidean distance or the maximum coordinates distance in Theorem 3.3. Remark 3.4 mentioned that none of the equations in (3.6) can be dropped. However, there are some conditions that can be imposed to check whether

$$h(X_1, X_2) \overset{d}{=} h(Y_1, Y_2) \iff F = G \quad (3.14)$$

or

$$h(X_1, X_2) \overset{d}{=} h(Y_1, X_3) \iff F = G. \quad (3.15)$$

A trivial counter example for (3.14) is that $X$ is from $N(0,1)$ and $Y$ is from $N(1,1)$. Another counter example in $k$-dimensional space is by taking $Y$ as the orthogonal transform of $X$ as well as the shift transformation, i.e. $Y = BX + A$, where $B^T B = I$ and $A$ is a relocation vector, $h$ is the Euclidean distance. Therefore, one has

$$h(Y_1, Y_2) \overset{d}{=} (BX_1 + A - BX_2 - A)^T ((BX_1 + A - BX_2 - A) \overset{d}{=} (X_1 - X_2)B^T B(X_1 - X_2)$$

$$\overset{d}{=} (X_1 - X_2)^T (X_1 - X_2) \overset{d}{=} h(X_1, X_2). \text{ But } X \not\equiv Y$$

Remark 3.4 gave a counter example for (3.15). The following theorem shows that (3.15) is valid when the characteristic functions of $X$'s and $Y$'s never vanish.
**Theorem 3.4.** With the same setting in Theorem 3.3, if the characteristic function of X, \( \phi(S) \), never vanishes, then

\[
\rho(X_1, X_2) \overset{d}{=} \rho(X_3, Y_1) \text{ iff } F = G. \quad (3.16)
\]

**PROOF.** From (3.9) and (3.10) in Theorem 3.3, one has \( \varphi(-S) \phi(S) = |\phi(S)|^2 \). Therefore, \( \varphi(S) = \phi(S) \) if \( \phi(S) \) never vanishes for any \( S \). ■

The counter example for (3.15) as we mentioned above (\( Y \) degenerate at 0 and \( X \) is from a Bernoulli with probability 1/2), does have the property of a vanishing characteristic function for some \( S \).

**Theorem 3.5.** With the same setting as in Theorem 3.3, for the characteristic functions \( \varphi(S) \) and \( \phi(S) \) satisfy that \( |\varphi(S)|^2 = |\phi(S)|^2 \) implies \( \varphi(S) = \phi(S) \), one has

\[
\rho(X_1, X_2) \overset{d}{=} \rho(Y_1, Y_2) \text{ iff } F = G. \quad (3.17)
\]

**PROOF.** From (3.10) and (3.11) in Theorem 3.3, one has \( |\varphi(S)|^2 = |\phi(S)|^2 \) which implies \( \varphi(S) = \phi(S) \). Therefore, \( \varphi(S) = \phi(S) \). ■

The counter example for (3.14) as mentioned before, \( Y = X + a \), satisfies \( |\varphi(S)|^2 = |\phi(S)|^2 \), although we do not have \( \varphi(S) = \phi(S) \) unless \( a = 0 \). Similarly, the shift and orthogonal transform example that was previously mentioned also satisfies \( |\varphi(S)|^2 = |\phi(S)|^2 \), although we do not have \( \varphi(S) = \phi(S) \).
Remark 3.7. Theorem 3.4 is also true for $\rho(X_1, X_2) \overset{d}{=} \rho(X_1, Y_1)$ iff $F=G$.

It may be useful for possible application in developing a goodness-of-fit criterion or a two-sample test which has the same reference point in the distances.
Chapter IV

The Proposed Simulation-based Estimation Method

Introduction

This chapter includes a proposed criterion for the comparison of simulated data with experimental data, and a proposed random function minimization algorithm. The proposed criterion is motivated by the theory of Chapter III, by ease of computation motivation, and by a particular application to prey-predator modeling. It has a chi-square type formula. The proposed optimization algorithm is designed for the random criterion. Finally, sections 4.4 and 4.4 are to compare the proposed simulation-based estimation method to methods in the literature.

4.1 The Proposed Random Criterion

The one sample multivariate goodness-of-fit test is an appropriate fitting criterion if it is possible to simulate a large amount of data within a short time, since in this case the distributional properties of the simulation model become
essentially known. Otherwise, we may use the two-sample test formula as the criterion. However, it is important that a criterion has the small sample property and also can test the general alternative. Therefore, this section proposes a criterion which has these features.

For illustration, Figure 4.1 gives an example where the experimental data are observed counts of spider mite populations living on lima bean plants in a greenhouse, and simulated data are from a complicated model established to mimic these populations. In Figure 4.1, this data is time sequence data with iid replication. The experimental replication size is 3 while the simulation was replicated 4 times in this example. Since both are time sequence data, one will need a method of mapping k time space points into a single point in k dimensions. Figure 4.2 illustrates the mapping technique in the k=2 case. Let \( X_{ij} \) be the count for replication i at time j, in the greenhouse, \( i=1,2,3, j=1,...,k \) and similarly take \( Y_{ij} \) to be the simulated count for the ith replication at time j. Thus, from Figure 4.2 one can map these points into data in k-dimensional space \( X_i = (X_{i1}, X_{i2},..., X_{ik}) \) and \( Y_i = (Y_{i1}, Y_{i2},..., Y_{ik}) \) which can then be used for the data comparison.

Using the theory of Chapter III, the proposed criterion checks if the Y's are as close to X's as the X's are to themselves. The proposed criterion is then based on a series of concentric "grids" center around the experimental data. For example, these "grids" are balls for data in three dimensions and are circles in two dimensions, see Figure 4.3, for the comparison of the proportion of simulation points within the ball to the expected proportion.
Partial data from Horn (1991) USDA project 2SSM female adults population at 20°C on lima bean plant

**Figure 4.1** Example of two sets of data for comparison
Figure 4.2 Illustration of mapping the time sequence observed data to points in multi-dimensions
The formulation of the proposed criterion is as follows:

Let \( X_1, X_2, \ldots, X_m \) be the observed data iid from \( F \) and let \( Y_1, Y_2, \ldots, Y_n \) be the simulated data iid from \( G \). Also, taking each \( X \) in turn as the reference point:

\[ X_j^* = \text{the reference point} \]

\[ D(A,B) = \text{distance between } A \text{ and } B, \]

\[ D_k^j = k^{th} \text{ order statistic amongst } D(X_1, X_j^*), D(X_2, X_j^*), \ldots, D(X_m, X_j^*) \text{ excluding the distance of } X_j^* \text{ to itself.} \]

Define sets:

\[ R_k^j = \begin{cases} \{ (X_1 X_2 \ldots X_m \text{ and } Y) : D_k^j < D(Y, X_j^*) < D_{k+1}^j \} & \text{for } k=1 \text{ to } m-1 \\ \{ (X_1 X_2 \ldots X_m \text{ and } Y) : D_m^j < D(Y, X_j^*) < \infty \} & \text{for } k = m \end{cases} \]

and let

\[ O_k^j = \text{observed number of } Y's \text{ in set } R_k^j \text{ (see Figure 4.3), and} \]

\[ p_k^j = P(Y \in R_k^j), \pi_k^j = \text{observed proportion of } Y's \text{ in set } R_k^j. \]

\[ E_k^j = \text{expect number of } Y's \text{ in } R_k^j = n \frac{p_k^j}{\pi_k^j}. \]

Then, the criterion is:

\[ C_{m,n} = \sum_{j=1}^{m} \sum_{k=1}^{m} \frac{(\pi_k^j - p_k^j)^2}{p_k^j}. \]
The mean and variance of $C_{m,n}$ can be approached by the following (for fixed $m$ and $n$):

$$C_{m,n} = \frac{1}{n} \sum_{j=1}^{m} \sum_{k=1}^{m} \frac{(O_{k}^{j} - E_{k}^{j})^2}{E_{k}^{j}}$$

let $H_{m,n}^{i} = \frac{m}{\sum_{i=1}^{m} (O_{i}^{i} - np_{i}^{i})^2}$, then,

$$E(H_{m,n}^{i}) = m-1,$$

$$\text{var}(H_{m,n}^{i}) = 2(m-1) + \frac{1}{n} \left[ (\sum_{i=1}^{m} \frac{1}{p_{i}^{j}}) - m^2 - 2m + 2 \right].$$ (Cramer, 1946)
Therefore,

\[ E(C_{m,n}) = E\left( \sum_{j=1}^{m} H_{m,n}^j \right) = \frac{m(m-1)}{n}, \]

\[ \text{var}(C_{m,n}) = \text{var}\left( \sum_{j=1}^{m} H_{m,n}^j \right) \]

\[ = \frac{1}{n^2} \sum_{j=1}^{m} \text{var}(H_{m,n}^j) + \sum_{j\neq j'} \text{cov}(H_{m,n}^j, H_{m,n}^{j'}) \]

\[ = \frac{1}{n^2} [2m(m-1) + \frac{1}{n} \left( \sum_{j=1}^{m} \sum_{i=1}^{m} \frac{1}{p_i^j} \right) - m^3 - 2m^2 + 2m] + \sum_{j\neq j'} \text{cov}(H_{m,n}^j, H_{m,n}^{j'}). \]

with

\[ \text{cov}(H_{m,n}^j, H_{m,n}^k) = \frac{2}{n} - 2 + \frac{2}{n} \sum_{l=1}^{m} \sum_{h=1}^{m} \frac{p_{lh}^{jk}}{p_l h} + \frac{1}{n} \sum_{l=1}^{m} \sum_{h=1}^{m} \frac{p_{lh}^{jk}}{p_l h} - \frac{m^2}{n}, \]

where \( p_{lh} = p(Y \in R_i^j \text{ and } Y \in R_h^k) \).

It is usually very complicated to calculate \( p_k^j \) even for a one-dimensional distribution. Also they depend on the distributions, F and G. However, one can obtain a close estimate of \( p_k^j \) from the simulation model by assuming that the current best fit provides the null distribution and repeatedly generating data sets as to serve the observed sample. This might be time consuming in the process of estimating those \( p \)'s in the simulation estimation routine. In Lemma 4.1 and Lemma 4.2, some bounds for the \( p \)'s will be discussed along with the use of the value \( \frac{1}{m} \), which doesn't depend on the underlying distribution.
The proposed criterion has the advantage that the origins are from the observed data. One can define the balls simply from the available data. The next issue is to find the correct expected value, the P's, for the simulated data within the balls, although they won’t be distribution free. However, it is shown in Lemma 4.2, in the one-dimensional case, that the bounds for the P's will be very close to a constant if the observed sample size is large enough.

Lemma 4.1. Let $X_1, X_2, \ldots, X_m$, and $Y$ be univariate random samples iid from $F$ and $X(1), X(2), \ldots, X(m)$ be the order statistics from $X_1, X_2, \ldots, X_m$. Then, for $r < s$,

$$P(X(r) < Y < X(s)) = \frac{s - r}{m + 1}.$$ 

Proof. Note that $P(X(r) < Y < X(s))$

$$= \int_0^1 \int_0^{F(v)} \frac{m!}{(r - 1)! (s - r - 1)! (m - s)!} F(u)^{r-1} [F(v) - F(u)]^{s-r-1} [1 - F(v)]^{m-s} dF(u) dF(v)$$

$$= \frac{m!}{(r - 1)! (s - r - 1)! (m - s)!} \int_0^1 F(v)^{r-1} [F(v) - F(u)]^{s-r-1} [1 - F(v)]^{m-s} dF(u) dF(v)$$

$$= \frac{m!}{(r - 1)! (s - r - 1)! (m - s)!} \frac{(r-1)! (s-r)! s! (m-s)!}{s! (m+1)!}$$

$$= \frac{s - r}{m + 1}. \quad \square$$

Remark 4.1. For the same setting as in Lemma 4.1, $P(Y \leq X(s)) = \frac{s}{m + 1}$.
Lemma 4.2. Let $X_1, X_2, \ldots, X_m$, and $Y$ be univariate random variables iid from $F$ and $D(1,i), D(2,i), \ldots, D(m-1,i)$ be the order statistics from $|X_1 - X(i)|, |X_2 - X(i)|, \ldots, |X_{j-1} - X(i)|, |X_{j+1} - X(i)|, \ldots, |X_m - X(i)|$ with $X_j (= X(i))$ being the center of the distance, $i$ from 1 to $m$. Then, for all $s$ from 1 to $m-1$,

$$\frac{s}{m+1} \leq P(|Y - X(i)| \leq D(s,i)) \leq \frac{s + 1}{m + 1}.$$

PROOF. When the center is $X(1)$, one can use the inequality that $-\infty \leq 2X(1) - X(s) \leq X(1)$ and follow the same proof as in Lemma 4.1 to show (4.1) is true. Similarly, when the center is $X(m)$, the proof can be followed by using the inequality $X(m) \leq 2X(m) - X(m-s) \leq \infty$. The following proof is for the case when center is $X(2)$. The proof for other center points allow by analogy.

For $s=1$, $D(1,2)$ is the shortest distance from $X(2)$ with the nearest neighbor being either $X(3)$ or $X(1)$. For $D(1,2) = X(3) - X(2)$, one has

$$P(|Y - X(2)| \leq D(1,2)) = P(|Y - X(2)| \leq X(3) - X(2)) = P(2X(2) - X(3) \leq Y \leq X(3)) = P(Y \leq X(3)) - P(Y < 2X(2) - X(3))$$

Since $X(1) \leq 2X(2) - X(3) \leq X(2)$, one has $\frac{1}{m+1} \leq P(Y < 2X(2) - X(3)) \leq \frac{2}{m+1}$. Therefore, $\frac{1}{m+1} \leq P(|Y - X(2)| \leq D(1,2)) \leq \frac{2}{m+1}$. Similarly, for the nearest neighbor to $X(2)$ is $X(1)$, i.e., $D(1,2) = X(1) - X(2)$, one can show that it is also true that

$$\frac{1}{m+1} \leq P(|Y - X(2)| \leq D(1,2)) \leq \frac{2}{m+1}.$$
For $s=2$, $D_{(2),2}$ is the second shortest distance from $X_{(2)}$ with the second nearest neighbor being either $X_{(4)}$ or $X_{(1)}$. For $D_{(2),2} = X_{(4)} - X_{(2)}$, one has

$$P\left( |Y - X_{(2)}| \leq D_{(2),2} \right) = P( |Y - X_{(2)}| \leq X_{(4)} - X_{(2)})$$

$$= P(2X_{(2)} - X_{(4)} \leq Y \leq X_{(4)})$$

$$= P(Y \leq X_{(4)}) - P(Y < 2X_{(2)} - X_{(4)})$$

Since $X_{(1)} \leq 2X_{(2)} - X_{(4)} \leq X_{(2)}$, one has $\frac{1}{m+1} \leq P(Y < 2X_{(2)} - X_{(3)}) \leq \frac{2}{m+1}$. Therefore, $\frac{2}{m+1} \leq P( |Y - X_{(2)}| \leq D_{(1),2}) \leq \frac{3}{m+1}$.

Similarly, for $D_{(2),2} = X_{(1)} - X_{(2)}$, one can show that it is also true that

$$\frac{2}{m+1} \leq P( |Y - X_{(2)}| \leq D_{(2),2}) \leq \frac{3}{m+1}.$$  

By similarly argument, one can show that it is also true for all $s$ in the case of center is $X_{(2)}$. Therefore, $\frac{k}{m+1} \leq P( |Y - X_{(2)}| \leq D_{(k),2}) \leq \frac{k+1}{m+1}$. In general, one can obtain the same bounds in (4.1) for all $s$ and all $i$. 

**Remark 4.2** In the same setting as above, one has $\frac{1}{m+1} \leq p_{k}^{j} \leq \frac{2}{m+1}$ for all $k$ and for all $j$. Therefore, the bias of using $\frac{1}{m}$ as the estimated value will be eliminated as the observed sample size goes to infinity.

Since being distribution-free is a desirable property for a good criterion, another approach is to generate the origins such that they are independent of both the observed data and the simulated data. Therefore, a potential modified version
of the proposed criterion is based on this concern about the bias of using incorrect expected proportions. The formulation of the modified criterion is as follows:

Let \( X_1, X_2, \ldots, X_m \) be iid from \( F \) and \( Y_1, Y_2, \ldots \) iid from \( G \)

\[
\rho(A, B) = \text{distance between } A \text{ and } B
\]

\[
\rho_k(X, C) = k^{\text{th}} \text{ order distance among } \rho(X_1, C), \rho(X_2, C), \ldots, \rho(X_m, C),
\]

\( k = 0, 1, 2, \ldots, m \). Let \( \rho_0(X, C) = 0 \) and \( \rho_{m+1}(X, C) = \infty \), where \( C \) is the corresponding origin.

Define sets:

\[
\mathcal{R}_k = \{ (X_1, X_2, \ldots, X_m, Y, C) \mid \rho_k(X, C) \leq \rho(Y, C) < \rho_{k+1}(X, C) \} \quad k = 0 \text{ to } m.
\]

Therefore, we have disjoint classes \( \mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_m \), where \( C \) can be obtained from prior information or can be a random sample from the simulation.

let

\[
I_{k, x, c} = \begin{cases} 
1 & \text{if } (X_1, X_2, \ldots, X_m, Y, C) \in \mathcal{R}_k \\
0 & \text{otherwise}
\end{cases}
\]

\[
\pi_k(x) = E_C \{ I_{k, x, c} \},
\]

\[
P_k = P\{ (X_1, X_2, \ldots, X_m, Y, C) \in \mathcal{R}_k \} = E_x \{ \pi_k(x) \} = E_x \{ E_C(I_{k, x, c}) \}.
\]

We will have \( P_k = \frac{1}{m+1} \) under the hypothesis that \( F = G \). This is because

\[
p(T(k) \leq U < T(k+1)) = \frac{1}{m+1},
\]

where \( T(k) \) is the \( k^{\text{th}} \) order statistics from \( T_1, T_2, \ldots, T_m \) and \( U, T \) are iid.

Therefore,

\[
P_k = \text{prob} \{ \rho_k(X, C) \leq \rho(X, C) < \rho_{k+1}(X, C) \} \quad \text{for } k = 0, 1, 2, \ldots, m
\]

\[
= \int p(T(k) \leq U < T(k+1)) f(c) \, dc
\]
The modified criterion is then

$$C_m = \sum_{k=0}^{m} m \left[ \pi_k(x) - \frac{1}{m+1} \right]^2.$$ 

The primary advantage of the above method is that one will be able to apply the criterion even when the sample size is small. Since the expected value of a $Y$ in any ring is always $\frac{1}{m+1}$, it saves the time of estimating those $p$'s in the simulation estimation procedure. Comparing the original proposed criterion to the modified criterion; the original criterion uses observed data as the origins to define the "grids", which is intuitively the best choice from the present available information. However, there is bias if one tries to use $\frac{1}{m}$ as the values of $p$'s, particularly with a small sample of observations. The modified criterion uses $\frac{1}{m+1}$ which is the correct expected value, but the origins are generated from the simulation model with the parameter values constantly changing, which may not define the "grids" well. In addition, the user must specify the number of origins that will affect the efficiency and accuracy of measuring the goodness-of-fit.

4.2 Proposed Optimization Algorithm for the Random Criterion

A general algorithm which is designed specifically for a random criterion can be described as follows:
Step 1. starting from initial guess $\theta_0$,

Step 2. decide whether to make a global jump (with probability $\zeta$ and generate jump size and jump direction from the full parameter space) or carry out a local search and use local jump direction and local jump size,

Step 3. if it is a global jump, accept the new $\theta'$ if the simulation shows improvement,

Step 4. if it is a local jump, perform the local search loop until the local minimum is found. A local search will be guided by available deterministic differential equations and a local regression to search the correct direction and appropriate jump size,

Step 5. repeat step 2 to step 4 until the convergence of the algorithm.

In step 2, the global jump direction and the jump size should be free to be chosen from the entire parameters space. In step 3 and step 4, with respect to the "improvement", since there is variation in simulating the stochastic process at given parameter values for a finite number of simulations, the concept of "improvement" is different from the deterministic minimization approach. We decide that there has been an improvement in the simulation estimation process and accept the new parameter value if the estimated objective function is lower than $(100\alpha)^{th}$ percentile of the criterion values that were created by the old candidate.

In step 4, the direction chosen for the local search is random with the distribution being weighted forward by a quick approximation method such as a set of deterministic differential equations. The jump direction distribution is also decided by a local regression model and by the previous jump direction. Those tools for choosing the best jump direction are important, particularly for a search in a high
dimensional parameter space and for reducing the time of finding the optimization solution. Those three elements: differential equations, local regression, and previous jump direction are used in combination depending on the efficiency of solving the differential equations, and generating the data for the local regression. The search looks in directions specified by the differential equations with a probability that is a monotone function of the $R^2$ from a regression model relating locally estimated values of the criterion and the corresponding values of the criterion. Further, let

$$\frac{\text{cost of simulation}}{\text{cost of differential equation}} = \lambda.$$  

The conditions of using differential equation and regression model in the local jump routine are:

a) if $\lambda \leq c_1$, then use the differential equation and regression model with all cross-product terms,

b) if $c_2 \leq \lambda < c_1$, then use the differential equation and regression model only once per local search,

c) if $\lambda < c_2$, then don't use differential equation at all.

Where $c_1$ and $c_2$ are user specified constants. For the local jump size, let

$$\theta_i^* = \theta_i + \text{sign}_i \xi_i, \ i=1 \text{ to } J,$$

where

$J = \text{number of parameters to be estimated.}$
\( \theta^*_i \) is the next candidate for i-th parameter with \( \text{sign}_i \) being the direction of the proposed change, \( \text{sign}_i = (-,0,+) \), and \( \xi_i \) being the jump size. \( \theta_i \) is the current n-th step, i-th parameter value. \( \xi_i \) is generated from a mixture of uniform(0, \( w_n \)) distribution and point mass at 0 (with probability \( \pi \)). Here \( \varepsilon_i \) is the width of the confidence boundary for parameter \( \theta_i \). \( w_0 \) = maximum number of steps allowed in the local search, and \( w_n \) = current local search step number.

The above confidence region for \( \theta \), to define \( \varepsilon_i \), is taken as below,

\[
\{ \theta : |C(\theta) - C(\theta^0)| < D^Y_n \},
\]

(4.1)

where \( C(\theta^0) \) is the current best criterion at the local search. \( D^Y_n \) is the cut off point for the (1-\( \gamma \)) confidence region, see Figure 4.4. In order to use (4.1), we need a method which specifies how changes in the criterion are related to changes in the parameters. One such method is to use the Taylor approximation formula,

\[
C(\theta) = C(\theta_0) + \sum_i (\theta_i - \theta_{0i}) \frac{\partial C}{\partial \theta_i}.
\]

An alternative way to implement (4.1) is to use the regression model for the confidence region. Then, \( D^Y_n = C_{1-\gamma}(\theta) - C_\alpha(\theta) \) which can be obtained by the local regression involving the differential equations. \( \gamma \) and \( \alpha \) are user supplied values for the precision of the region. \( \alpha \) is associated with the level of accuracy in defining the "improvement".
If the local jump direction is solely determined by previous attempts for an improvement without using the differential equations and regression model, the distribution of assigning the three possible signs (-, 0, +) to the parameter $\theta_i$ is

$$Z_1 = P(\text{next step sign} = 0) = \pi,$$
$$Z_2 = P(\text{next step sign} = \text{last successful attempt direction}) = (1 - \frac{w_n}{w_o})(1 - \pi),$$
$$Z_3 = P(\text{next step sign} \neq \text{last successful attempt direction}) = 1 - Z_1 - Z_2.$$

Noting that the chance we continue to search in the same direction is a decreasing function of how long it has taken as to be successful in that direction ($W_n$).

When the efficiency conditions of using the differential equations and the regression model are satisfied, one can approach the jump direction by the following:
let
\[ q = \text{the level of reliance of the differential equation, } 1 \geq q \geq 0, \]
\[ R^2 = \text{square of correlation coefficient of the regression model}. \]

Then,
\[ Z_1 = P(\text{next step sign} = 0) = \pi \]
\[ Z_2 = P(\text{next step sign} = \text{last successful attempt direction}) = Z_2 + qR^2(1-Z_2) \]
\[ Z_3 = P(\text{next step sign} \neq \text{last successful attempt direction}) = 1 - \pi - Z_2 + qR^2(1-Z_2). \]

Where in \( Z_2 \), it means the differential equation agrees with the last successful attempted direction, and in \( Z_3 \), it means the differential equation does not agree with the last successful attempt direction. The regression with cross-product terms,
\[
C(\theta) = a_0 + \sum_i a_i(\theta_i - \theta_i^n) + \sum_i \sum_j a_{ij}(\theta_i - \theta_i^n)(\theta_j - \theta_j^n),
\]
picks the direction of parameter \( \theta_i \) as \( \text{sign}\{a_i + \sum_j a_{ij}(\theta_j - \theta_j^n)\} \).

In between step 1 and step 5, the cross-validation sampling is applied frequently to check if there is any improvement in the random criterion. This procedure is described as in the following:

1) determine \( k_1 \times k_2 = \text{total number of simulations for each run to investigate the previous best parameter estimate} \). Suppose each quick local search used \( k_2 \) simulations.
2) calculate $C_1(\theta_n^{k_2}), C_2(\theta_n^{k_2}), ..., C_{k_1}(\theta_n^{k_2})$ for each group of simulation at current best choice of parameter $\theta^n$.

3) look for next candidate by checking whether the new criterion is lower than the $\alpha$-th percentile of $C_1(\theta_n^{k_2}), C_2(\theta_n^{k_2}), ..., C_{k_1}(\theta_n^{k_2})$.

Also in step 5, an important aspect of the stopping rules is to check if the number of group, $k_1$, is large enough and without any improvement in the search. Otherwise when $k_1$ is small, there may be a chance of improvement due to the randomness of the criterion. To ensure the global minimization for the random criterion, it is necessary to have a large number of $k_1$ and a large number of trials in the global search. The above algorithm is summarized in following flowchart, see Figure 4.5.

### 4.3 Advantages of the Proposed Random Criterion

The simulation estimation technique is a very general method which can be applied to any situation that can be simulated. Therefore, the random criterion should have all the nice properties of a good multivariate goodness-of-fit test. Some desirable properties of a good multivariate goodness-of-fit test are:

(i) the test is consistent against general alternative,

(ii) the test procedure should be distribution free,

(iii) the test should have small sample property,

(iv) the test should be easy to compute.
Figure 4.5 Flow chart of optimization method for the simulation-based estimator
Comparing with multivariate tests in the literature, the classical one-sample procedures such as the Hotelling $T^2$ test do not have the distribution-free property. Furthermore, in order to use an one-sample multivariate test as the goodness-of-fit criterion in the current case, one has to simulate a very large amount of data. Therefore, classical one-sample tests are inappropriate in the simulation. In two-sample testing, Friedman and Rafsky's two-sample test is computationally too slow for finding the minimum spanning tree, especially in a minimization algorithm that repeatedly simulates data. Other two-sample tests such as the nearest neighbor method of Schilling (1986) has a very simple form and distribution-free properties. But to find all the nearest neighbors of each corresponding origin from the whole data set is not easy, particularly in a large data set.

The advantages of the proposed criterion are, first, the test is consistent against the general alternative ($X$ and $Y$ are not equal in distribution). Second, the criterion is distribution free as shown in section 4.2. Finally, the most important part of the proposed criterion is that it is computationally easy. One only needs to count the number of $Y$'s in the "rings".

4.4 Advantages of the Proposed Algorithm

Since in most cases, there will be no closed form solution through minimizing the criterion, it is very important to have a good optimization algorithm. Practically, a good algorithm should have the following nice properties:

(i) the final solution should not depend on the initial starting point,
(ii) it should avoid getting trapped in local minimum,
(iii) the algorithm to find the solution should be fast computationally including an appropriate jump size, a correct search direction and a good stopping rule.

The criterion is random due to the experimental process and the simulation process. Therefore, the algorithm in the simulation estimation should be designed particularly for a random function instead of for a deterministic objective function.

Most optimization algorithm in the literature involved deterministic search. The problems with deterministic searches are i) the final solution depends on the starting point and ii) the deterministic methods tend to get trapped in local optimum. Unfortunately, simulated annealing methods and others are very general and slow, also they are not designed for a random criterion function. Therefore, they are not readily adapted to the technique of simulation estimation. Designing an algorithm for minimizing a random criterion becomes the primary task in practicing the simulation-based estimation method. The proposed algorithm (in step 2) has both a global and a local search strategy which occurs with an user specified probability. This helps avoid being trapped in a local solution in a manner similar to the way in which the simulating annealing method is designed. Also, an appropriate jump size based on a confidence ellipsoid and a search direction guided by deterministic equations and a regression model are factors to speed up the computation. Finally, the stopping rule can be determined from the null distribution if the null distribution is known. Another approach is to stop the searching if the simulation size reaches an user specified maximum without any improvement after a specified number of steps.

Combining both the proposed random criteria and algorithm, the simulation-based estimation method has the following significant advantages:
(i) it can be used without a likelihood function,
(ii) it can be applied for any type of data,
(iii) it is easy to apply.

The importance of (i) is that, with a complicated data structure, sometimes the likelihood function is not available or is very complicated. Sometimes, one may have a problem with no closed-form solution. Furthermore, it is very easy to make a mistake in finding the second derivative, especially if the likelihood is complicated and has many parameters. To see the importance of (ii), one frequently has to deal with the problem of missing data. The simulation estimation method can fit and compare any type of data as long as one can simulate them. Particularly, in the case of missing data and incomplete data, all that is needed is a model which created the missing values. For (iii), the estimation can be applied to data from any distribution and data in any dimensions as long as the simulation model has been built.
Chapter V

Confidence Interval for S.B.E. and Its Empirical Study

Introduction

The idea of obtaining a level of accuracy about an estimated parameter by the simulation-based estimation technique is quite different from the classical confidence region approach. In this chapter, an approximated confidence region is defined and the consistency of the simulation estimators are discussed. Section 5.3 includes the empirical study of S.B.E. from several popular distributions. Section 5.4 concerns a simulation study of S.B.E. interval estimation.

5.1 Consistency of S.B.E.

A general central limit theorem is proved by Pakes (1989) for estimator defined by the minimization of the length of a vector-valued, random-criterion function , \( \| G(\cdot) \| \). Let \( n \) = number of observations, \( s \) = number of simulation draws per
observation, and \( \theta_o = \) the true value of a parameter vector at which the criterion is minimized. One way to estimate \( \theta_o \) is to construct a sequence \( \{G_n\} \) of random functions that converge to \( G \) in some sense, then find \( \hat{\theta}_n \) that makes \( G_n(\hat{\theta}_n) \) as close to zero as possible.

The following theorems, due to Pakes (1989), provide a consistency theorem and a central limit theorem for a simulation-based estimator that minimizes the length, \( \|G_n(\cdot)\| \), of a random function. This function is defined on a subset \( \Theta \) of some \( \mathbb{R}^k \). It should be thought of as an estimate of a deterministic, vector-valued function \( G(\cdot) \) that is also defined on \( \Theta \). The true value \( \theta_o \) is defined implicitly as the unique point in \( \Theta \) for which \( G(\theta_o) = 0 \).

**Theorem 5.1** Under the following conditions \( \hat{\theta}_n \) converges in probability to \( \theta_o \):

(i) \( \|G_n(\hat{\theta}_n)\| \leq o_p(1) + \inf_{\theta \in \Theta} \|G_n(\cdot)\| \),

(ii) \( G_n(\theta_o) = o_p(1) \),

(iii) \( \sup_{|\theta - \theta_o| < \delta} \|G_n(\theta)\|^{-1} = O_p(1) \) for each \( \delta > 0 \).

Conditions for strong consistency of \( \hat{\theta}_n \) could be obtained by replacing the \( o_p(\cdot) \) by \( O_p(\cdot) \) in (i), (ii) and use almost sure analogue in (iii).

**Theorem 5.2** Let \( \hat{\theta}_n \) be a consistent estimator of \( \theta_o \). If

(i) \( \|G_n(\hat{\theta}_n)\| \leq o_p(n^{-1/2}) + \inf_{\theta \in \Theta} \|G_n(\cdot)\| \),

(ii) \( G(\cdot) \) is differentiable at \( \theta_o \) with a derivative matrix \( \Gamma \) of full rank,

(iii) for every sequence \( \{\delta_n\} \) of positive numbers that converges to zero

\[
\sup_{|\theta - \theta_o| < \delta_n} \frac{\|G_n(\theta) - G(\theta) - G_n(\theta_o)\|}{n^{-1/2} + \|G_n(\theta)\| + \|G_n(\theta)\|} = o_p(1),
\]
(iv) $\sqrt{n} \ G_n(\theta_0) \to N(0, V)$,
(v) $\theta_0$ is an interior point of $\Theta$, then

$$\sqrt{n} \ \left( \hat{\theta}_n - \theta_0 \right) \to N(0, (\Gamma \Gamma')^{-1} \Gamma' \Gamma (\Gamma \Gamma')^{-1}). \quad (5.1)$$

The following two lemmas state conditions under which the Euclidean norm $\| \cdot \|$ can be replaced by random norms that depend on $\theta$, without disturbing the above limit theorems.

**Lemma 5.1** Let $\{ A_n(\theta) : \theta \in \Theta \}$ be a family of sequences of nonsingular random metrics for which

(a) $\| A_n(\theta_0) \| = o_p(1),$  
(b) $\sup_{\theta \in \Theta} \| A_n(\theta)^{-1} \| = o_p(1).$

If $G_n(\cdot)$ satisfies conditions (ii) and (iii) of Theorem 5.1, then these conditions also hold with $G_n(\theta)$ replaced by $A_n(\theta)G_n(\theta)$.

**Lemma 5.2** Let $\{ A_n(\theta) : \theta \in \Theta \}$ be a family of sequences of nonsingular random metrics for which there exists a nonsingular, non random matrix $A$ such that

$$\sup_{|\theta - \theta_0| < \delta_n} \| A_n(\theta) - A \| = o_p(1),$$

whenever $\{ \delta_n \}$ is a sequence of positive numbers that converges to zero. If conditions (ii), (iii) and (iv) of Theorem 5.2 are satisfied by $G_n(\cdot)$ and $G(\cdot)$, then one can replace $G_n(\theta)$ by $A_n(\theta)G_n(\theta)$, $G(\theta)$ by $AG(\theta)$, $V$ by $AVA'$, and $\Gamma$ by $A\Gamma$. In addition, one can replace $G_n(\theta)$ by $A_n(\theta)G_n(\theta)$, $G(\theta)$ by $AG(\theta)$, $V$ by
AVA' and $\Gamma$ by $A\Gamma$ in Theorem 5.1 and Theorem 5.2 where $A_n(\theta)$ is a nonsingular random matrix.

5.2 Confidence Interval of S.B.E.

Although Theorem 5.2 gives the limiting distribution of the simulation-based estimator, the derivative matrix $\Gamma$ is very difficult to get. In most cases, the formula for the derivative matrix is unavailable. Therefore, one cannot obtain the approximate confidence region for the parameters.

In this section, the confidence region is approximated by the relationship between the parameters and the criterion. From section 4.1, we have seen that the proposed criterion has a non-central chi-square distribution if the parameter value does not equal $\theta_0$. Therefore, the variation of the criterion will tend to increase as the values of parameters move away from $\theta_0$. The relationship between the random criterion and parameters is unknown. However, under a smoothness assumption, one may build up a model for the relationship. For example, Thompson (1987) uses a quadratic regression model for the relationship between criterion and parameters around the final solution $\hat{\theta}$, i.e., $C(X,Y;\theta) = \beta_0 + \beta_1 \theta + \beta_2 \theta^T\theta + \epsilon$ for $\theta$ around $\hat{\theta}$. A $(1-\alpha)$ approximate confidence interval was obtained from the relationship

$$P\{ C_l(X,Y;\hat{\theta}) \leq C(X,Y;\theta) \leq C_u(X,Y;\hat{\theta}) \} \leq \alpha,$$
where $C^L(X,Y;\hat{\theta})$ and $C^U(X,Y;\hat{\theta})$ are the empirical $\frac{\alpha}{2}$th percentile and $(1-\frac{\alpha}{2})$th percentiles from the simulation data at $\hat{\theta}$. Then, the $1-\alpha$ approximated confidence can be obtained by inverting the inequality,

$$C^L(X,Y;\hat{\theta}) \leq C(X,Y;\theta) \leq C^U(X,Y;\hat{\theta}).$$

The above method constitutes a very rough approach of the confidence interval, unless one has the following two important conditions: i) with a good model, for example, one must have high $R^2$ in the regression model, ii) invertible relationship between parameters and criterion. For instance, in i), one usually has low $R^2$ if the size of the simulation is small. Also, it is difficult to specify the range of parameters in the regression model around the global minimum, especially in multiple parameters case. As regards ii), the inversion formula will not be available if the model is complicated. For example, a model with multiple parameters which includes all interaction terms may not have an inversion formula.

In addition to the randomness of the experimental process, the randomness of the simulation process involved in the random criterion will create a region for finding the solution of the simulation-based estimator. Figure 5.1 shows the shape of the proposed criterion for the observed data from a $N(0,1)$ distribution with observed sample size $= 64$. The criterion is increased when the simulated data are from $N(\mu,1)$ with $\mu$ being away from 0. The size of the simulated data is 256 and there are 10 replicates of simulation for each given $\mu$. Below, we show a method leading to a more accurate confidence region.
Figure 5.1  Values of the proposed criterion when the location parameter deviates from N(0,1)

Method 5.1. Based on the estimated values, one can simulate the observed data several times. For each replicated data set, a new estimated value is obtained by the simulation-based estimation routine. Therefore, the confidence region can be approximated by the new set of estimated values obtained from S.B.E. procedure. The algorithm for the simulation-based interval estimation is:
Step 1: simulate $n$ data sets from the estimated parameter value,

Step 2: run the simulation estimation procedure to get the new estimated value for each observed data set generated from the simulation model,

Step 3: complete step 2 for all the $n$ data sets, and obtain $n$ estimated values $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$,

Step 4: obtain the confidence interval from those estimated values.

In one parameter case, the $1-\alpha$ confidence interval for $\theta$ is then $(a, b)$ where $a$ and $b$ are the $\frac{\alpha}{2}$th percentile and $(1-\frac{\alpha}{2})$100th percentiles of the estimated values.

However, if the simulation model is very time consuming, it is practically difficult to apply the above method. Since the overall interval estimation procedure needs far more CPU time than it is required in a single run of simulation. One can approach the interval estimation as follows (which may result with less accurate estimated confidence region):

From $P\{ C(X,Y;\theta)|_{X=\Delta} \leq \Delta \} \leq \alpha$, one can obtain the region by inverting the inequality $C(X,Y;\theta)|_{X=\Delta} \leq \Delta$ into a region, say $\hat{\theta} \in S(\hat{\theta})$. This method is similar to Thompson's method. However, it gives more accurate estimates, since the confidence region results from an inversion of a one-sided inequality. It should include the small criterion values in obtaining the confidence region. The length of a simulation-based confidence interval depends on the form of the random criterion, the observed sample size, and the simulation sample size. One can reduce the length of the interval by increasing the size of simulation.
5.3 Empirical Study for the Simulation-based Estimator

Simulation-based estimation is particularly useful when the experiment is very complicated (which usually makes the likelihood function unavailable). However, the robustness of the simulation-based estimator can be studied from the empirical analyses in a variety of models. For example, from an univariate $N(\mu, \sigma^2)$ distribution, 100 sets of data were generated from IMSL routine with the size of each data set $m=16$. For each set of data, one can compare the estimated value for $(\mu, \sigma^2)$ from the results of the uniformly minimum variance unbiased estimator, $(\hat{\mu}, \hat{\sigma}^2)$, i.e., UMVU, and the results from the simulation-based estimator, $(\tilde{\mu}, \tilde{\sigma}^2)$, i.e., S.B.E., where $\hat{\mu} = \bar{X}$, and $\hat{\sigma}^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$. Two cases are studied to check the robustness of the S.B.E: one is data sets from $N(2,25)$ and the other is from $N(9,4)$. In the simulation-based estimation procedure, the proposed criterion was applied for the fitting, except that $\frac{1}{m}$ was used for the expected value in the rings. Figure 5.2 shows the values of UMVU and S.B.E. From those plots, S.B.E. catches the true mean and variance very well. Both plots of UMVU vs. true mean and variance, and S.B.E. vs. true mean and variance are quite similar. The bottom of Figure 5.2 indicates that there are slightly higher value of the estimated variance for the S.B.E as compared to UMVU. This is because the incorrect value, $\frac{1}{m}$, was used for the expected values of $Y$ in the rings. When the observed sample size is increased to 64, the bias is smaller as compared to the bias in Figure 5.2. Also, Figure 5.3 shows an improvement in estimating $\mu$ and $\sigma^2$ for both UMVU and S.B.E., since the observed sample size is increased to 64. Again, this is because the bias of using $\frac{1}{m}$ decreased as the observed sample size increased. By using $\frac{1}{m}$ as the expected value for the $p$'s will speed up the
simulation estimation process, simply by allowing to avoid the procedure of repeatedly estimating the p's in the simulation estimation procedure. The medians in the boxes that were plotted in Figure 5.2 and Figure 5.3 are listed in Table 5.1.

There is also a study for the observed sample size m=32, with value of μ changing from 2 to 9 and σ² changing from 0 to 7 in the N(μ,σ²) distribution. Figure 5.4 shows that both plots for the estimated mean by UMVU method and S.B.E. method are similar. There is slightly larger variation among the estimated values by S.B.E. for estimating the variance as compared to the values estimated by UMVU method. The CPU time of running the simulation-based estimation procedure for the m=64 case is about 9 seconds on CRAY-Y/8MP machine, and the random numbers were generated from IMSL RNNOA routine.

The empirical study of S.B.E. for other common discrete distributions includes binomial(30,p), poisson(τ) and discrete uniform(1,K). The results are listed in Table 5.2. The empirical analysis of S.B.E. includes the study for other common continuous density distributions, such as Cauchy(θ), beta(α,1) and exponential(λ). The results are listed in Table 5.3. For each distribution, 64 points are generated from IMSL routines. From those generated data, S.B.E.s are obtained by running the simulation-based estimation procedure, and the classical estimation methods used for comparison are MLE, method of moment and others which give a closed form solution. Formulas of the closed from estimator for each distribution in the empirical study are as follows:

\[
\begin{align*}
\text{binomial}(30,p) & \quad \hat{p} = \frac{\sum X_i}{30m} \\
\text{discrete uniform}(1,K) & \quad \hat{k} = \max(X_i) \\
\text{Poisson}(\tau) & \quad \hat{\tau} = \overline{X} \\
\text{Cauchy}(\theta) & \quad \hat{\theta} = \text{median}(X_i) \\
\text{beta}(\alpha,1) & \quad \hat{\alpha} = \frac{\overline{X}}{1 - \overline{X}} \\
\text{exponential}(\lambda) & \quad \hat{\lambda} = \overline{X}.
\end{align*}
\]
The data for the comparison of UMVU and S.B.E. were generated from $N(2,25)$ and $N(9,4)$, each box with 100 points, each point with 16 samples. Bottom of Figure 5.2 shows the comparison between S.B.E. and UMVU. The lines in those boxes are the $45^0$ lines which indicate equal values on both axes.

**Figure 5.2** UMVU and S.B.E. in $N(2,25)$ and $N(9,4)$ study (m=16)
The data for the comparison of UMVU and S.B.E were generated from $\mathcal{N}(2, 25)$ and $\mathcal{N}(9, 4)$, each box with 100 points, each point with 64 samples. Lines in bottom boxes indicate equal values on both axes. Compare these plots to Figure 5.2: S.B.E. in the present case is closer to the UMVU estimation results. This is because larger observed sample size was used in this study.

**Figure 5.3** UMVU and S.B.E. in $\mathcal{N}(2, 25)$ and $\mathcal{N}(9, 4)$ study (m=64)
Table 5.1 Medians in the boxes of Figure 5.1 and Figure 5.2

<table>
<thead>
<tr>
<th></th>
<th>sample size</th>
<th>for μ=2</th>
<th>for σ²=25</th>
<th>for μ=9</th>
<th>for σ²=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.B.E</td>
<td>16</td>
<td>1.56</td>
<td>34.93</td>
<td>8.97</td>
<td>5.57</td>
</tr>
<tr>
<td>S.B.E</td>
<td>64</td>
<td>1.91</td>
<td>27.12</td>
<td>9.01</td>
<td>4.37</td>
</tr>
<tr>
<td>UMVU</td>
<td>16</td>
<td>2.06</td>
<td>26.05</td>
<td>8.93</td>
<td>3.80</td>
</tr>
<tr>
<td>UMVU</td>
<td>64</td>
<td>1.93</td>
<td>25.04</td>
<td>8.98</td>
<td>3.89</td>
</tr>
</tbody>
</table>

From Table 5.2 and Table 5.3, it can be seen that the simulation-based estimator is robust against any discrete and continuous distribution. Furthermore, it is easy to apply without involving any formula calculation or worrying about the closed form solution problem. For example, in Table 5.3, there is no closed form solution for the maximum likelihood method in both the Cauchy distribution and beta distribution cases. Note that in the discrete parameter space such as the discrete uniform distribution case, the S.B.E. looks for the estimation result only in discrete values. Therefore, there is no risk that the final solution lies outside the parameter space. In addition, the S.B.E. method may be computationally more efficient if the optimization program for the numerical approach in the maximum likelihood method is too complicated.

Figure 5.5 contains the plots of the proposed criterion vs. parameters. The observed sample is generated from N(0,1) with size m=32. Then, values of criterion are obtained by generating simulated data sets with size=32 from N(μ,σ²) for different values of μ and σ². For each value of μ and σ², there are 9
replications. Figure 5.5 shows the relationship between the criterion and parameter. One will expect that the estimated value ($\hat{\mu}, \hat{\sigma^2}$) to be close to (0,1). Again, Figure 5.5 gives the positive answer.

Table 5.2 S.B.E. and classical estimators for discrete distributions, m=64

<table>
<thead>
<tr>
<th>true parameter</th>
<th>closed form estimator</th>
<th>S.B.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>binomial(30, p)</td>
<td>p = 0.5</td>
<td>0.512</td>
</tr>
<tr>
<td>Poisson(τ)</td>
<td>τ = 5.5</td>
<td>5.313</td>
</tr>
<tr>
<td>discrete</td>
<td>κ = 25</td>
<td>25</td>
</tr>
<tr>
<td>uniform(1, κ)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3 S.B.E. and classical estimators for continuous distributions, m=64

<table>
<thead>
<tr>
<th>true parameter</th>
<th>classical estimator</th>
<th>S.B.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauchy(θ)</td>
<td>θ = 1</td>
<td>0.78</td>
</tr>
<tr>
<td>beta(α, 1)</td>
<td>α = 2.5</td>
<td>2.38</td>
</tr>
<tr>
<td>exponential(λ)</td>
<td>λ = 5</td>
<td>4.87</td>
</tr>
</tbody>
</table>
For each point of these plots, the observed sample size equals 32. Lines in those boxes indicate equal values for both axes.

**Figure 5.4** UMVU and S.B.E. for $N(\mu, \sigma^2)$ study with $m=32$
Figure 5.5 Values of the proposed criterion for parameters $\mu$ and $\sigma^2$ deviate from $N(0,1)$, $m=32$, $n=96$
5.4 Empirical Study for Simulation-based Interval Estimation

The asymptotic normality of S.B.E. from the result of Theorem 5.2 cannot be used in this study since the derivative matrix $\Gamma$ is not available. In fact, the functional relationship between the criterion and parameter, $\| G_n(\theta) \|$ is not available in most cases. In this section, simulation-based interval estimation is studied from the $N(\mu,1)$ case with the observed data generated from $N(0,1)$. Let $X_1, X_2, \ldots, X_m$ be the observed random sample from $N(\mu,1)$ with $\mu = 0$. Then, the common estimate of $\mu$ is $\hat{\mu} = \bar{X}$. Let $\mu_1$ be the simulation-based estimator. The following studies compare $\text{var}(\hat{\mu})$ with $\text{var}(\mu_1)$.

The $\text{var}(\mu_1)$ can be obtained by the interval estimation algorithm in section 5.2. In short, the idea is that after S.B.E. was obtained, say $\mu_0$, $k$ sets of data with size $m$ were generated from $\mu_0$ presuming each data set is the original observed data. For each presumed original data set, the simulation estimation process was applied to obtain S.B.E., $\mu_i$. Therefore, $\text{var}(\mu_1)$ was estimated by

$$\text{var}(\mu_1) = \frac{\sum_{i=1}^{k}(\mu_i - \mu_0)^2}{k}.$$  

The $1-\alpha$ simulation-based interval of $\mu$ is then,

$$\tilde{\mu}_0 - Z_\alpha \sqrt{\hat{\text{var}}(\mu_1)} \leq \mu \leq \tilde{\mu}_0 + Z_\alpha \sqrt{\hat{\text{var}}(\mu_1)},$$

where $Z_\alpha$ is $(1 - \frac{1}{2\alpha})^{th}$ percentile of the standard normal distribution.
Table 5.4 gives the results of \( \text{var}(\mu) \) for \( m=10 \) and Table 5.5 gives the results of \( \text{var}(\mu) \) for \( m=40 \). The whole process is repeated 16 times. First column of Table 5.4 is \( \overline{X} \), second column is the corresponding ending solution of S.B.E, third column is \( \text{var}(\mu) \) for \( k=12 \), and similarly for the study of \( \text{var}(\mu) \) at \( m=40 \). The fourth column is to compare the mean square deviation from \( \overline{X} \) with the mean square deviation from \( \tilde{\mu}_0 \) in the third column. When the observed sample size is increased, the estimated value for \( \mu \) should be more accurate for any consistent estimator. Table 5.4 and Table 5.5 show that both \( \overline{X} \) and S.B.E. improved the estimation result along with larger observed sample size (see column 1 and column 2 in both tables). Also note that the mean square deviation for the simulation-based estimation results is smaller by calculating the deviation from \( \tilde{\mu}_0 \) than the deviation from \( \overline{X} \). Finally, the empirical study verifies the \( \sqrt{m} \)-consistency property in Theorem 5.1 and Theorem 5.2. From the data in column three of Table 5.4 and Table 5.5, let

\[
\sigma^2_{10} = \text{average of } \text{var}(\mu) \text{ in Table 5.4} = 0.038, \\
\text{(sample size } m_1 = 10 \text{)}
\]

\[
\sigma^2_{40} = \text{average of } \text{var}(\mu) \text{ in Table 5.5} = 0.163. \\
\text{(sample size } m_2 = 40 \text{)}
\]

Since

\[
\frac{\sigma^2_{10}}{\sigma^2_{40}} = \frac{0.403}{0.195} \approx \frac{m_2}{m_1} = \frac{40}{10}
\]

the \( \sqrt{m} \)-consistency of the simulation-based estimator is verified through the empirical study.
Table 5.4 Estimated variation of simulation-based estimator, m=10

<table>
<thead>
<tr>
<th>$\bar{X}$</th>
<th>S.B.E.</th>
<th>$\frac{\text{var}(\hat{\mu})}{\text{var}(\mu)}$</th>
<th>$\frac{\text{var}(\hat{\mu})^*}{\text{var}(\mu)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1018</td>
<td>0.2308</td>
<td>0.241</td>
<td>0.2576</td>
</tr>
<tr>
<td>0.5232</td>
<td>0.3473</td>
<td>0.3793</td>
<td>0.4102</td>
</tr>
<tr>
<td>0.0084</td>
<td>-0.1589</td>
<td>0.2157</td>
<td>0.2437</td>
</tr>
<tr>
<td>-0.0014</td>
<td>-0.3006</td>
<td>0.1454</td>
<td>0.235</td>
</tr>
<tr>
<td>-0.7923</td>
<td>-0.867</td>
<td>0.029</td>
<td>0.0345</td>
</tr>
<tr>
<td>-0.5685</td>
<td>-0.6008</td>
<td>0.0865</td>
<td>0.0875</td>
</tr>
<tr>
<td>-0.4574</td>
<td>-0.6839</td>
<td>0.1169</td>
<td>0.1682</td>
</tr>
<tr>
<td>-0.8866</td>
<td>-0.7497</td>
<td>0.0552</td>
<td>0.0740</td>
</tr>
<tr>
<td>0.1932</td>
<td>0.0838</td>
<td>0.2260</td>
<td>0.2380</td>
</tr>
<tr>
<td>0.3441</td>
<td>0.2375</td>
<td>0.1497</td>
<td>0.1611</td>
</tr>
<tr>
<td>0.3176</td>
<td>0.2853</td>
<td>0.1808</td>
<td>0.1818</td>
</tr>
<tr>
<td>0.9420</td>
<td>0.5702</td>
<td>0.1569</td>
<td>0.2951</td>
</tr>
<tr>
<td>0.5548</td>
<td>0.2678</td>
<td>0.1414</td>
<td>0.2238</td>
</tr>
<tr>
<td>0.6928</td>
<td>0.6526</td>
<td>0.0991</td>
<td>0.1007</td>
</tr>
<tr>
<td>-0.7592</td>
<td>-0.5256</td>
<td>0.1509</td>
<td>0.2054</td>
</tr>
<tr>
<td>0.5661</td>
<td>0.3037</td>
<td>0.2352</td>
<td>0.3041</td>
</tr>
</tbody>
</table>

Note that \( \text{var}(\mu) = \text{var}(\bar{X}) = \frac{1}{m} = 0.1 \)

\[
\text{var}(\hat{\mu})^* = \frac{\sum_{i=1}^{m} (\hat{\mu}_i - \bar{X})^2}{k}.
\]
Table 5.5  Estimated variation of simulation-based estimator, \( m=40 \)

<table>
<thead>
<tr>
<th>( \bar{X} )</th>
<th>S.B.E.</th>
<th>( \hat{\text{var}}(\hat{\mu}) )</th>
<th>( \hat{\text{var}}(\hat{\mu})^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.1994</td>
<td>-0.2945</td>
<td>0.0302</td>
<td>0.0392</td>
</tr>
<tr>
<td>-0.1514</td>
<td>-0.1097</td>
<td>0.0185</td>
<td>0.0202</td>
</tr>
<tr>
<td>-0.1750</td>
<td>-0.1294</td>
<td>0.0683</td>
<td>0.0704</td>
</tr>
<tr>
<td>0.3875</td>
<td>0.3238</td>
<td>0.0215</td>
<td>0.0255</td>
</tr>
<tr>
<td>0.0227</td>
<td>-0.0245</td>
<td>0.0319</td>
<td>0.0342</td>
</tr>
<tr>
<td>-0.1231</td>
<td>-0.2156</td>
<td>0.0196</td>
<td>0.0283</td>
</tr>
<tr>
<td>0.1747</td>
<td>0.0962</td>
<td>0.0332</td>
<td>0.0394</td>
</tr>
<tr>
<td>0.2035</td>
<td>0.2245</td>
<td>0.0196</td>
<td>0.0201</td>
</tr>
<tr>
<td>0.0567</td>
<td>-0.0091</td>
<td>0.038</td>
<td>0.0424</td>
</tr>
<tr>
<td>-0.1459</td>
<td>-0.2003</td>
<td>0.0241</td>
<td>0.0271</td>
</tr>
<tr>
<td>0.0219</td>
<td>-0.0038</td>
<td>0.0725</td>
<td>0.0731</td>
</tr>
<tr>
<td>-0.0798</td>
<td>-0.2782</td>
<td>0.0391</td>
<td>0.0785</td>
</tr>
<tr>
<td>-0.0060</td>
<td>-0.1026</td>
<td>0.0531</td>
<td>0.0624</td>
</tr>
<tr>
<td>0.1932</td>
<td>0.1714</td>
<td>0.0658</td>
<td>0.0663</td>
</tr>
<tr>
<td>-0.0835</td>
<td>-0.0708</td>
<td>0.0286</td>
<td>0.0287</td>
</tr>
<tr>
<td>-0.0346</td>
<td>-0.0591</td>
<td>0.0374</td>
<td>0.0380</td>
</tr>
</tbody>
</table>

Note that \( \text{var}(\hat{\mu}) = \text{var}(\bar{X}) = \frac{1}{m} = 0.25 \)

\[
\hat{\text{var}}(\hat{\mu})^* = \frac{\sum_{i=1}^{k}(\hat{\mu}_i - \bar{X})^2}{k}.
\]
Chapter VI

Application:
S.B.E. for Greenhouse Prey-predator Model

Introduction

The two-spotted spider mite (2SSM), *Tetranychus urticae Koch*, is the most widespread and serious pest among spider mites in greenhouse and outdoor crops. Because of pesticide resistance, it is desirable to replace chemical spraying by biological control. The predatory mites used against 2SSM is the *Phytoseiulus persimilis* (PPE). PPEs feed on all stages of the spider mite and have a sequence of developmental stages similar to that of the spider mites but take much less time for molting from one stage to the other. Both mites have relatively small size, less than 1 mm. The predator-prey system involves life history parameters (development, reproduction, mortality and sex ratio) and the predator-prey interaction (predation and dispersal behavior). The environmental factors such as temperature and humidity play a very important role in the population dynamics and influence several parameters like oviposition rate, maturation and predation
rate. For example, (see Sabelis, 1981), a rise in temperature enhances development, stimulates reproduction and increases food consumption.

To model such a complex system, even if it is composed by a string of deterministic pieces, the model predictions may be biased even if each component of the model gives unbiased predictions. Furthermore, one may be unable to decide whether experimental deviations from model predictions are due to bias or to the natural variability in the system. A stochastic model will allow for the complete testing to judge the usefulness of modeling the system. Pearl, Bartoszynski and Horn (1989) proposed a stochastic model for simulation of spider mite predator-prey interactions which has been successfully tested and continue to be refined, see Horn et al (1991,a). The key aspects of the predator-prey model are:

1. maturation times of egg and nymph for both 2SSM and PPE are from a symmetric distribution,
2. for sex ratio, a nymph has a probability P for 2SSM or P* for PPE of maturing into a female adult (this assumption is due to the fact that it is practically impossible to distinguish the sex of eggs or nymphs in the greenhouse),
3. between the time of hatching and maturation, the events of mortality, ovipositions and migrations follow a multidimensional Markov process.

The components of event intensities of 2SSM death are: natural death, death due to miticide and PPE predation, except that PPE nymph cannot eat 2SSM female adult. PPE mortality intensity includes natural death and death due to miticide, if it is not a resistant mite. Egg laying rates of both species depend on temperature,
humidity and their gut-contents. Migrations of both species depend on competition for available food. Other assumptions like modeling the plant growth, miticide decays, gut-emptying rates, and attempting migration rates were formulated in Horn et al (1991,b).

Some of the parameters of the model can be estimated from the specifically designed experiments. In addition, information about some of the parameters can be found in the literature. In this chapter, the new estimated values of the parameters in the predator-prey model from different experiments will be discussed. The simulation-based estimation technique is also applied in several data sets that are too complicated to get the likelihood formula for estimating the parameters.

6.1 Starvation Experiments:

Cases When Likelihood Functions Are Available

Under the model assumptions, the event of PPE female adult death follows a Poisson process. The data from Cheng (1993) observed number of PPE female adults \( N_{i,t} \) which survived at least \( t \) days without any food in an environmental chamber until all the mites are gone, 10 mites in one chamber with total of 10 replications (i=1 to 10). The event intensity under these conditions is \( \frac{s}{D e^{-t+\alpha}} \)

\( N_{i,t} \). The survival time of a mite is then exponentially distributed with rate, \( \frac{s}{D e^{-t+\alpha}} \), where \( s \) is the starvation coefficient, \( a \) is the reduction coefficient and \( t \) is the time. \( D \) is initial (time 0) gut-contents of a mite and \( r \) is the digesting rate of
the gut-content from other experiments, r=1.76 and D=5.1. Both s and a are the parameters to be investigated. Therefore,

\[ P_t = \text{prob( an individual alive after } t \text{ days)} = P( T > t ) \]

\[ = \int_0^\infty \exp \left\{ -\frac{s}{D} e^{-s/t} - a \right\} d\mu. \]

The likelihood function can be written as

\[ L = \prod_{i=1}^{10} \prod_{t=1}^{n} \left( \frac{N_{i,t-1}}{P_{t-1}} \right)^{N_{i,t}} \left( \frac{P_{t}}{P_{t-1}} \right)^{N_{i,t} - N_{i,t-1}} \]

where \[ \frac{P_t}{P_{t-1}} = \left( \frac{D+ae^r(t-1)}{D+ae^rt} \right)^{ar}. \]

First, the maximum likelihood estimator can be obtained by maximizing the above log likelihood function. Since there is no closed form solution, by calling a NAG FORTRAN subroutine to find the maximum, the MLE of s and a were \((\hat{s}, \hat{a}) = (0.1739, 0.289)\). To obtain confidence intervals for both parameters, we can use the asymptotic normality of MLE (Lehmann 1983). Under reparameterization \( \left( \frac{s}{a} = x, \ a = y \right) \), and letting

\[ f_y = \frac{1+dy \exp(rt-r)}{1+dy \exp(rt)}, \ \text{where } d = 1/D, \]

\[ m_t = \sum_i N_{i,t}, \ \ l_t = \sum_i (N_{i,t-1} - N_{i,t}), \ \text{and } k=1/r. \]

One can get the second derivatives in organized forms:
\[
\frac{\partial^2 \log L}{\partial x^2} = \sum_t -l_t k \log(\phi y) \frac{k_{fykx} \log(\phi y)}{(1-fykx)^2},
\]
\[
\frac{\partial^2 \log L}{\partial x \partial y} = \sum_t \left\{ m_t k \frac{\partial \phi y}{\partial y} \right\} - \frac{f_{fy} k x f_{fy}^{-kx-1} \log(f_{fy} + f_{fy}^{kx-1})(1-fykx) + k x \log(f_{fy}) f_{fy}^{2kx-1}}{(1-fykx)^2},
\]
\[
\frac{\partial^2 \log L}{\partial y^2} = \sum_t \left\{ m_t k \frac{f_{fy}^2 f_{fy}/\partial y^2 - (\partial f_{fy}/\partial y)^2}{(f_{fy})^2} \right\} - \frac{k l_t x [f_{x-1} f_{fy}^{kx-2} (\partial f_{fy}/\partial y)^2 + f_{fy}^2 f_{fy}/\partial y^2 f_{fy}^{kx-1} + k x f_{fy}^{2kx-2} (\partial f_{fy}/\partial y)^2]}{(1-fykx)^2}.
\]

Therefore,

\( (\hat{\mu} - \mu) \) is distributed approximately as from \( N(0, \Sigma) \) where \( \hat{\mu} = \left( \begin{array}{c} \hat{s} \\ \hat{\alpha} \end{array} \right) \) and

\[
\Sigma = (JJ')^{-1}, \text{ with } J = \left( \begin{array}{cc} -\partial^2 \log L/\partial x^2 & -\partial^2 \log L/\partial x \partial y \\ -\partial^2 \log L/\partial x \partial y & -\partial^2 \log L/\partial y^2 \end{array} \right) \text{ being the sample Fisher information matrix and } J = \left( \begin{array}{ccc} \partial x/\partial s & \partial y/\partial s \\ \partial x/\partial \alpha & \partial y/\partial \alpha \end{array} \right).
\]

The Bayesian estimator is somewhat difficult to calculate if the marginal distribution is too complicated. Geman and Geman (1984) introduced the Gibbs sampler to solve this problem. The priors of \( s \) and \( \alpha \) are assumed as uniform(0,\( b_1 \)) and uniform(0,\( b_2 \)) with \( b_1 \) and \( b_2 \) being known constants. Now,
the unknown variables are s and a. The algorithm of Gibbs sampling from Gelfand and Smith (1990) is as follows:

step (1) choose starting point $a^0$ (from prior),
step (2) generate $s^1$ from $f_1(s \mid N, a^0)$, and
step (3) generate $a^1$ from $f_2(a \mid N, s^1)$.

Iterate steps (2) and (3) $k$ times to obtain $s^k$ and $a^k$. Repeat (1),(2),(3) for $m$ times to get $(s^1, a^1), (s^2, a^2), \ldots, (s^m, a^m)$.

Marginal densities of s and a are $\frac{1}{m} \sum_{i=1}^{m} f_1(s \mid N, a_i)$ and $\frac{1}{m} \sum_{i=1}^{m} f_2(a \mid N, s_i)$ by the data argumentation method, (see Tanner and Wong, 1987). Convergence can be shown by the method similar to the simulated annealing. At steps (2) and (3), a random number must be generated from the corresponding distribution. Those functions are very sensitive to tiny changes of parameters because of the power $(\sum_{t} N_{i,t}$ and $\sum_{t} (N_{i,t-1} - N_{i,t})$ terms in the joint binomial distribution. This makes it very difficult to generate a random number by rejection method, which also requires to find the maximum for each step $i$ and for each posterior density. Two rejection methods were tried to generate a random number from the $i$-th step for each posterior. One is box rejection method and the other is triangle rejection method. When $k = 100$ and $m = 100$, the box rejection method took 22 seconds on the CRAY super-computer Y-8MP machine. However, it needs only 9 seconds if the triangle rejection method is used. Finding the maximum for each $i$-th step takes $1/3$ of the total computing time. Note that there are other ways of generating random numbers which may be more efficient, see Devroye (1986). From 100
sampling points, the posterior modes of $s$ and $a$ are 0.19 and 0.275, see Figure 6.1 for their posterior percentiles.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{histograms}
\caption{Histograms of Gibbs sampling posteriors}
\end{figure}

Finally, the simulation-based estimation technique is applied in this study. It took less than 1 second in CRAY time to reach the convergence with an estimated value of $(s, a)$ equal (0.19, 0.219). The results of these three methods are listed at Table 6.1, where S. B. E. is the simulation-based estimator.

Since Gibbs sampling heavily depends on the way of generating random numbers, it may not be efficient compared to technical methods like MLE. Also, for the application of Gibbs sampling, knowing how to choose appropriate $K$ and $M$ (the number of iterations and the number of replication) without theoretically figuring out the convergence rate is an interesting question. There are discussions about the inference for iterative simulation, and monitoring convergence of Gibbs sampler and Gibbs stopper by Gelman (1992). In addition, Ritter (1992)
proposed a Gibbs stopper technique for assessing convergence of the Gibbs sample for moderate sized problems.

**Table 6.1** Results of estimated values from different methods

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>Gibbs Sampler</th>
<th>S. B. E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starvation coefficient</td>
<td>0.174</td>
<td>0.190</td>
<td>0.191</td>
</tr>
<tr>
<td>Reduction constant</td>
<td>0.289</td>
<td>0.275</td>
<td>0.219</td>
</tr>
</tbody>
</table>

The simulation-based estimation seems to perform well. Furthermore, less computing time compared to the likelihood based technique is one of the surprise results. Figure 6.2 are the plots for the remaining PPE female adults after t days of starvation for both the experimental data and the simulated fits.

Similar experiments were conducted for PPE male adults and nymphs populations, and for the 2SSM female adult population. Final estimates of the MLEs are listed in Horn et al (1991,b).

### 6.2 2SSMs and PPEs on Lima Bean Plant Experiments

In the greenhouse, observing 2SSM and PPE populations for the predator-prey interaction experiments on lima bean plants is much simpler than observing
both populations on cucumber plants. The stochastic modeling of the spider mite predator-prey-plant system was developed by Pearl et al (1989) and had been modified for stochasticity due to egg, nymph maturation, and separating of adult population into male and female groups. In addition, it has been validated through experiments for the parameters refinement, see Horn et al (1991,b), which involved more than hundred parameters and required a variety of experiments for the estimation. In this section, the discussion will focus on the experiments and analyses which involved simulation estimation. In addition, mimic results from simulation will be plotted for the comparison to the observed data. Other small scale experiments to obtain accurate information on demographic events and several important biological parameters are discussed in Cheng (1993). The stochastic model was programmed on a CRAY Y-MP/864 computer with the program modified from the preliminary study version directed by Dennis K. Pearl on a Pyramid 90X computer. The CRAY version of program contains a vectorization technique and is carried out in a parallel fashion to speed the simulation. The supercomputer makes the simulation-based techniques feasible to estimate system parameters in a multi-dimensional setting.

6.2.1 Experiments at Different Temperatures

The data recorded 2SSM and PPE populations on single lima bean plant under different temperatures(20°C, 22.5°C, 25°C, 27.5°C and 30°C). Two female 2SSM adults were released at day 0 and two PPE female adults were released after the second generation of 2SSM appeared. The dynamic interactions between predator and prey becomes very complicated. Therefore, likelihood based
estimation techniques cannot be applied in this case since the likelihood function is not available. Although one might have some idea about the maturation time of egg and nymph from the early stages of development, it is not accurate since the exact times of maturation were not observed. Accurate estimates of maturation and egg laying rates for both species can be obtained by the simulation-based estimation procedure.

The model assumes intensities of egg laying are \( \frac{D_t}{D_t + tT} q_T \) for both 2SSM and PPE where \( D_t \) is the current gut-contents at time \( t \), \( q_T \) is the egg laying parameter which depends on temperature \( T \), \( r_T \) is the reduction constant. From above, one might expect shorter maturation time and higher egg-laying rate as the temperature increases. Let \( q_T^*, c_E, c_N, q_T^*, c_E^* \) and \( c_N^* \) be the corresponding parameters for egg-laying rate, egg-maturation time and nymph-maturation time of 2SSM and PPE. The final estimates from the simulation base estimation method are listed in Table 6.2.

The results of sensitivity analysis also showed that the predator-prey system is highly sensitive to small change in those parameters. Particularly, this is very important in the study of environmental stochasticity associated with changing temperature and humidity. From the S.B.E. results, one can obtain the mimic data set as in the laboratory. For these five different temperature experiments, two females adults 2SSM were released at day 0 on single lima bean plant. 2SSM populations were recorded at different days. After second generations of 2SSM appeared, 2 female adults PPE were released to control the pest population.
Figure 6.2 PPE female adults survival under starvation
Therefore, both populations were recorded several times until PPE populations were significantly increased and 2SSM populations were decreased.

Table 6.2  S.B.E. results for egg laying, egg maturation, and nymph maturation of 2SSM and PPE populations

<table>
<thead>
<tr>
<th></th>
<th>$q_T$</th>
<th>$c_E$</th>
<th>$c_N$</th>
<th>$q_T^*$</th>
<th>$c_{E^*}$</th>
<th>$c_{N^*}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20°C</td>
<td>4.08</td>
<td>5.16</td>
<td>5.78</td>
<td>2.35</td>
<td>2.8</td>
<td>2.81</td>
</tr>
<tr>
<td>22.5°C</td>
<td>5.50</td>
<td>3.73</td>
<td>5.15</td>
<td>3.22</td>
<td>2.65</td>
<td>2.06</td>
</tr>
<tr>
<td>25°C</td>
<td>6.14</td>
<td>2.80</td>
<td>5.38</td>
<td>4.05</td>
<td>1.91</td>
<td>2.43</td>
</tr>
<tr>
<td>27.5°C</td>
<td>6.99</td>
<td>3.06</td>
<td>4.32</td>
<td>4.5</td>
<td>1.86</td>
<td>1.71</td>
</tr>
<tr>
<td>30°C</td>
<td>7.38</td>
<td>2.77</td>
<td>4.30</td>
<td>3.5</td>
<td>2.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Unit for $q_T$ and $q_T^*$ is g/day, for $c_E$, $c_N$, $c_{E^*}$, and $c_{N^*}$ is day.

Figures 6.3, 6.4, 6.5, 6.6, and Figure 6.7 show the fitted data from simulated vs. experimental data for 20°C, 22.5°C, 25°C, 27.5°C, and 30°C experiments. The plots compare pest populations without PPE in the system. Figures 6.8, 6.9, 6.10, 6.11, and Figure 6.12 allow to compare simulated results to the experimental data for 2SSMs for the 20°C, 22.5°C, 25°C, 27.5°C, and 30°C experiments under the presence of PPE populations. The number of 2SSM female adults is adjusted to 20 on each plant at the day of releasing PPE. Figures 6.13, 6.14, 6.15, 6.16, and Figure 6.17 show the graphs for the PPE populations from those temperature experiments. More detailed descriptions of those experiments are given in Cheng (1993). Also, note that in Figure 6.3 to Figure 6.7, there are 16 replications (plants), i.e., 16 samples in each box plot, for observing mites.
population on single plant, eight of them are control plants and eight plants are experimental plants. Therefore, in Figure 6.8 to Figure 6.12, there are eight samples in each box, and there are eight samples in Figure 6.13 to Figure 6.17. In addition, PPE female adults were released at day 17, 14, 11, 10, and 9 for the 20°C, 22.5°C, 25°C, 27.5°C, and 30°C experiments respectively. This is because second generations of 2SSM appear sooner as the temperature in experiment is higher in view of higher egg laying rate, shorter egg and nymph maturation time, as expected from the S.B.E. results in Table 6.2. The estimates of maturation time and oviposition rate at different temperatures provided information to improve the predictive ability of the model in different environmental conditions. Therefore, this information can be incorporated into the model to reflect the effect of temperature on development and reproduction of mites. All those figures show a good match between experimental data and simulated data. Particularly, the stochastic model fits not only the average numbers of the populations but also captures the inherent dynamic variability within the populations.
Figure 6.3 2SSM populations at 20°C before 2 PPE females released
Figure 6.4 2SSM populations at 22.5°C before 2 PPE females released
Figure 6.5 2SSM populations at 25°C before 2 PPE females released
Figure 6.6 2SSM populations at 27.5°C before 2 PPE females released
Figure 6.7 2SSM populations at 30°C before 2 PPE females released
Figure 6.8  2SSM populations at 20°C after 2 PPE females released
Figure 6.9 2SSM populations at 22.5°C after 2 PPE females released
Figure 6.10 2SSM populations at 25°C after 2 PPE females released
Figure 6.11  2SSM populations at 27.5°C after 2 PPE females released
Figure 6.12 2SSM populations at 300°C after 2 PPE females released
Figure 6.13  PPE populations at 20°C
Figure 6.14  PPE populations at 22.5°C
Figure 6.15  PPE populations at 25°C
Figure 6.16  PPE populations at 27.5°C
Figure 6.17  PPE populations at 30°C
6.2.2 Four Lima Bean Plants Experiment

Previous single plant experiments emphasized the estimation of life history parameters of both mites. The focus of the four-plant experiment is on the study of dispersal tendency for 2SSM populations, which plays an important role in the predator-prey model and in the pest control. In this experiment, since the lima bean plant is relatively small, with each plant limited to five leaves, the simulation model fits the mite's populations counted on an entire plant, and considers only the migration of mites between plants. Since the classical likelihood-based estimation method cannot be applied in this complicated experiment, the simulation-based estimation was used to refine the migration parameters, their reduction, and the probabilities of death during migration. Table 6.3 contains the simulation-based estimation results.

Table 6.3 Results of migration parameters from the 4 plants experiment

<table>
<thead>
<tr>
<th>parameters</th>
<th>2SSM male</th>
<th>2SSM female</th>
<th>PPE nymph</th>
<th>PPE male</th>
<th>PPE female</th>
</tr>
</thead>
<tbody>
<tr>
<td>migration</td>
<td>0.006</td>
<td>0.011</td>
<td>0.01</td>
<td>0.17</td>
<td>0.201</td>
</tr>
<tr>
<td>reduction</td>
<td>26.53</td>
<td>11.51</td>
<td>62.5</td>
<td>76.1</td>
<td>79.2</td>
</tr>
<tr>
<td>P(death)*</td>
<td>0.0053</td>
<td>0.0123</td>
<td>0.03</td>
<td>0.051</td>
<td>0.06</td>
</tr>
</tbody>
</table>

P(death)* = probability of death during migration.
Unit for migration parameter is g/day, for reduction constant is g.
The simulated fits in the four plants experiment were plotted in Figures 6.18, 6.19, 6.20 and 6.21 for the 2SSM egg, nymph, male adult and female adult populations. Note that there are eight replications in this experiment and it was conducted in a controlled environment chamber at 25°C and relative humidity 75±5%. More detail of this experiment are discussed in Cheng (1993).

From those figures, on day 1 and day 6, the distribution of simulated and experimental output match well for the initial infestation. By day 11, a small infestation has begun on neighboring plants. Simulated results catch this distribution well. By day 14, substantial infestations occur on neighboring plants and population distributions in simulated and experimental data match well again.
Figure 6.18 2SSM egg populations on 4 lima bean plants
Figure 6.19 2SSM nymph populations on 4 lima bean plants
Figure 6.20 2SSM male populations on 4 lima bean plants
Figure 6.21  2SSM female populations on 4 lima bean plants
6.3 Experiments on Cucumber Plants

Since the goal of our predator-prey simulation is to develop an optimal pest management system for cucumber production, the next step is to carry out the commercial setting experiments on cucumber plants. In the previous section, we had obtained the estimated values of our model parameters and fitted the data well for mites' populations on lima bean plant. However, several experiments were conducted to detect if there is any difference for some important parameters in lima bean model and in cucumber model. We found that the oviposition rate of 2SSM female feeding on lima bean leaf is larger than the oviposition rate of 2SSM female feeding on cucumber leaf. However, there is no difference for the maturation time of a 2SSM feeding on lima bean leaves and a 2SSM feeding on cucumber leaves. Similar experiments for PPEs were also conducted. The results were further discussed and listed in Cheng (1993).

6.3.1 Cucumber Leaf Growth Model

Since there are several issues which depend on the edible leaf surface in our simulation model, such as 2SSM migration and 2SSM gut-content, it is important to estimate the parameters in the cucumber leaf growth model. From our model assumption, a cucumber leaf grows logistically with

\[
\frac{d S_j}{dt} = \gamma S_j (M_j - S_j),
\]

\[j=1 \text{ to } 20 \text{ (the total leaf number allowed in a single plant)},\]
where $S_j = \text{leaf } j \text{ surface}$, $\gamma = \text{cucumber leaf growth rate}$, and $M_j = \text{maximum leaf surface of leaf } j$, with leaf $j$ indicating the $j$-th leaf from the bottom of a cucumber plant. In addition, the maximum cucumber leaf surface is assumed from a Hockey Stick model with

$$M_{ij} = \beta_0 + \beta_1 j + \epsilon_{ij}, \quad \text{for leaf } j \text{ from 1 to 8, plant } i \text{ from 1 to 2},$$

and

$$M_{ij} = \mu + \epsilon_{ij}, \quad \text{for leaf } j \text{ from 9 to 20, plant } i \text{ from 1 to 2},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$.

There are three parameters to be estimated, $\beta_1$, $\mu$, and $\sigma^2$ with $\hat{\beta}_0 = \hat{\mu} - \frac{9}{10}\hat{\beta}_1$. The maximum leaf surface were measured at day 30, with 8 leaves already existed at day 0. Thus, the likelihood function is:

$$L = \prod_{i=1}^{2} \left( \prod_{j=1}^{8} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(M_{ij} - \beta_0 - \beta_1)^2}{2\sigma^2}} \right) \left( \prod_{j=9}^{20} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(M_{ij} - \mu)^2}{2\sigma^2}} \right).$$

The values of the MLE of ($\mu$, $\beta_1$, $\sigma^2$) are (483.87, 38.82, 1972.67). Figure 6.23 shows the fitted Hockey Stick model for the maximum cucumber leaf area. The logistic leaf growth model is:

$$\log\left( \frac{S_i}{M_i - S_i} \right) = \gamma M_{it} + C_i,$$

where $C_i$ is the half-time leaf surface of leaf $i$ (the time that a leaf will grow to the half size of the maximum surface). The estimated values for $\gamma M$ and $C$, for leaf 6
Figure 6.23  Hockey Stick model for cucumber maximum leaf surface

to leaf 20 were listed in Table 6.3.  Average $R^2$ for the regressions is 86.6%.
Figure 6.24 plotted the observed cucumber leaf area.
Table 6.3 Cucumber leaf growth rate and half-time leaf surface

<table>
<thead>
<tr>
<th>leaf number</th>
<th>growth rate (γM)</th>
<th>half-time (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.1124</td>
<td>8.30</td>
</tr>
<tr>
<td>7</td>
<td>0.1448</td>
<td>7.24</td>
</tr>
<tr>
<td>8</td>
<td>0.1430</td>
<td>12.07</td>
</tr>
<tr>
<td>9</td>
<td>0.1813</td>
<td>13.18</td>
</tr>
<tr>
<td>10</td>
<td>0.1811</td>
<td>13.39</td>
</tr>
<tr>
<td>11</td>
<td>0.2261</td>
<td>13.91</td>
</tr>
<tr>
<td>12</td>
<td>0.2337</td>
<td>15.07</td>
</tr>
<tr>
<td>13</td>
<td>0.2956</td>
<td>18.08</td>
</tr>
<tr>
<td>14</td>
<td>0.2430</td>
<td>17.90</td>
</tr>
<tr>
<td>15</td>
<td>0.2342</td>
<td>18.12</td>
</tr>
<tr>
<td>16</td>
<td>0.3323</td>
<td>19.93</td>
</tr>
<tr>
<td>17</td>
<td>0.2247</td>
<td>15.62</td>
</tr>
<tr>
<td>18</td>
<td>0.1445</td>
<td>20.58</td>
</tr>
<tr>
<td>19</td>
<td>0.1907</td>
<td>16.46</td>
</tr>
<tr>
<td>20</td>
<td>0.3187</td>
<td>26.89</td>
</tr>
</tbody>
</table>

6.3.2 2SSM on Cucumber Plants

In this section, the analysis is focused on the 2SSM migration in order to manage the pest infestation. The first experiment was conducted by releasing 2 female 2SSM adults on leaf number 4 of each plant. There are 16 cucumber plants.
separated into 4 groups, each group with 4 replications (plants). Leaves and mites of those 4 groups were frozen at 4 different days in order to count the mites. Experiment 2 was conducted similar to experiment 1 except 2 females were released on leaf 6 of each plant, with total number of leaves restricted to 20 in a single plant. The first group of experiment 2 was frozen 15 days after those 2SSMs were released. For the rest three groups, 10 female PPE adults were released on one edge plant of each group at day 15. Experiment 3 was designed similarly to experiment 2 but with total of 20 plants separated into 4 groups.

2SSM was allowed to migrate up 3 leaves and down 3 leaves in the cucumber predator-prey model. Based on the refined model, and the parameters estimated in lima bean experiments and cucumber experiments, the simulation-based estimation method was applied in the estimation of the 2SSM cucumber plant migration. Those migration related parameters are:

\[ \begin{align*}
\text{hw} & : \text{2SSM female adult migration}, \\
\text{hm} & : \text{2SSM male adult migration}, \\
P_{U1} & : \text{Prob(2SSM migrates up 1 leaf | the 2SSM migrates),} \\
P_{U2} & : \text{Prob(2SSM migrates up 2 leaf | the 2SSM migrates),} \\
P_{U3} & : \text{Prob(2SSM migrates up 3 leaf | the 2SSM migrates),} \\
P_{D1} & : \text{Prob(2SSM migrates down 1 leaf | the 2SSM migrates),} \\
P_{D2} & : \text{Prob(2SSM migrates down 2 leaf | the 2SSM migrates),} \\
P_{D3} & : \text{Prob(2SSM migrates down 3 leaf | the 2SSM migrates),} \\
\delta & : \text{Prob(2SSM stays on leaf 20 | the 2SSM migrates),} \\
\end{align*} \]

with the constrains that \( P_{Ui} = \alpha P_{Di}, \ i=1,2,3 \), and \( \sum_{i=1}^{3} (P_{Ui} + P_{Di}) = 1. \)
From those experiments, we found that there were thousands of mites after
day 30. Since the early stages of infestation is particularly important in estimating
mite's migration, only the early frozen groups in those experiments were
considered in the S.B.E fitting. Note that with this experimental data of moderate
size, it took more than 2.5 hours of CRAY CPU time to finish a S.B.E.
procedure. A single simulation took 30 seconds in CRAY since the events
happened in the stochastic predator-prey process become more intensive. The
S.B.E. of those migration parameters (hw, hm, P_{U1}, P_{U2}, P_{U3}, \delta) were
(1.24, 0.12, 0.28, 0.26, 0.42, 0.94). Thus, this gave us the values of the
migration parameters in our system of differential equations for mimic the large
populations in later stages. Figures 6.25, 6.26, 6.27, 6.28, and 6.29 showed the
comparison between the experimental data and the simulated data. Note that
although we were able to mimic the right size of the 2SSM populations, the early
arrivals of 2SSMs on leaf 20 in Figure 6.28 and 6.29 indicate that it may be
possible for a mite to migrate up k leaves directly with k>3. The PPE migration in
cucumber plants is under investigation. Both 2SSM and PPE migration
experiments will provide the information to develop the optimal biological controls
in the greenhouse cucumber production for the appropriate timing of releasing
predators and the appropriate amount of predators to be released.
Figure 6.24 Cucumber leaf growth curve
Figure 6.25  Average 2SSMs at day 13, N=4 (experiment 1 vs. simulation)
Figure 6.26  Average 2SSMs at day 20, N=4 ( experiment 1 vs. simulation )
Figure 6.27  Average 2SSMs at day 15, N=4 ( experiment 2 vs. simulation )
Figure 6.28  Average 2SSMs at day 23, N=4 ( experiment 2 vs. simulation )
Figure 6.29  Average 2SSMs at day 17, N=5 (experiment 3 vs. simulation)
Chapter VII

Conclusion

The studies of simulation-based estimation methodology were motivated by the procedure of refinement of a sophisticated predator-prey-plant model (Horn et al., 1991, b). When a system is so complicated, such as the predator-prey model, the likelihood function for the data of all the populations in the system is not available. Intuitively, adjusting the parameters values in the simulation model to match experimental data is the easiest way of estimating parameters. At the earlier stage of developing the S.B.E. technique, I encountered the problem of how to choose the appropriate parameters values such that the simulation model will produce better matched data. Therefore, I and our research team tried to adjust the parameters values from previous simulation outcomes to a new candidate which is anticipated to result in better matched data. This process yields several questions: First, how to deal with variation among simulations and variation among experiments. Second, how to adjust the parameter value to produce better matched data? Third, how to transform time-sequence data into multi-dimensional data. Fourth, how to compare two multi-dimensional data sets. Finally, how to define an appropriate goodness-of-fit criterion for the matching.
Chapter III gives the theoretical background of why the two-sample matching problem can be translated to the comparison of interpoint distances among samples and distances between samples. Characterizing a multi-dimensional two-sample comparison problem into an one-dimensional distance function practically allows us to calculate the fitting criterion. As in Chapter II, an ideal two-sample test for the fitting criterion must have distribution-free, small sample size, and easy calculation properties. The theorem in Chapter III does have the distribution-free property. This gives the theoretical support of applying S.B.E. technique to other simulation models, for example, in estimating parameters in stochastic cancer model. In addition, since the proof of the theorem doesn't use the triangle inequality, which is one of the properties in a distance function, this yields more flexibility in defining the one-dimensional function as the matching criterion for reducing multi-dimensional data.

Chapter IV includes the proposed S.B.E. criterion and algorithm to solve the problem of finding appropriate next candidate parameter value in the searching process and for the matching of two samples. Both algorithm and criterion are new in terms of minimizing a random objective function and defining the "grids" by the observed data for two-sample comparison. These together make the S.B.E. technique easy to apply and easy to use for data of any type and of any dimensions. In Chapter V, the S.B.E. consistency and asymptotic normality properties were shown from literature theorems. But the confidence interval is implemented by a proposed simulation-based interval estimation scheme. Its empirical study also validates the $\sqrt{n}$-consistency result from the theorem. Since the S.B.E. is designed for estimating parameters in any distribution and any model, the robustness of S.B.E. is studied in Chapter V for several common
distributions. The results show that S.B.E. ends up with very close solution as compared to standard estimation method.

Finally, the S.B.E. is applied to the predator-prey model on pest populations of *Tetranychus urticae* and its predator populations, *Phytoseiulus persimilis*. Although the interest of the USDA project by David Horn, Dennis K. Pearl and Robert Bartoszyński is in developing an optimal pest-management program for greenhouse cultural systems, a model refinement is essential to have the ability to reflect the inherent dynamics of populations both in mean and variance, in particular, for the occasional outbreaks of the two-spotted spider mite in many agricultural systems. Through the S.B.E. technique, the predator-prey model had been validated with the ability to dynamically mimic the complexities of environmental realities. The S.B.E. methodology also can be applied to the experiments conducted for practical production system in a greenhouse commercial setting to reach the goal of finding the optimal strategy of integrating biological and chemical control efforts in a greenhouse system.

There are some places that the S.B.E. can be improved in estimating parameters. For example, one can design a minimization algorithm depending on the size of simulation for more efficient performance. This is an open area which I believe will attract substantial interest in the future, just as the problem of minimizing a deterministic objective function. In addition, one can design a fitting criterion that has more power in detecting the difference between two samples such as in a location difference and scale difference problem.

Finally, as Thompson(1987) said, simulation can be used to proceed rapidly from the axioms to the estimation of the characterizing parameters. This concept is, on the one hand, a drastic departure from precomputer age methodology. On the other hand, it simply extends the one-dimensional goodness-of-fit concept.
The simulation-based estimation technique enables us to estimate parameters in models far more complex than the classical approach. Hopefully, this is a step in that direction.
REFERENCES


