Simultaneous estimation of risk in several 2 x 2 contingency tables : an empirical Bayes approach
by James Stanley Bergum

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Philosphy in Statistics
Montana State University


Abstract:
A set of g 2X2 contingency tables with the row margins fixed may be represented as g pairs of independent binomial variates, and x ^ » with corresponding parameters, and P ^ , and sample sizes, and n ^ » where i = 1,... ,g. Associated with table i is the parameter of interest, 0^, which is some specific function of P ^ and P ^ * In this thesis, Bayes and empirical Bayes estimators are derived to simultaneously estimate 9^, 0g . The following six choices for 6jL were studied in detail: P ^ , P ^ , P 11/(l - P ^ ) , P ^ / C l - P 12) , P 11(l - ^ 12^ *" P i l ^ ' an<* P il^P i2 * The last two choices, the odds ratio and the relative risk, are measures of association commonly estimated in epidemiological studies. The estimators are based on a bivariate prior distribution on the unit square which allows for a correlation between P ^ and P ^ * The Bayes estimators are shown to be asymptotically equivalent to the maximum likelihood estimators. An ad hoc confidence interval procedure based on the empirical Bayesian estimates is given. The posterior densities for each of the parameters are presented. A simulation study to compare the empirical Bayes estimator with the maximum likelihood estimator and to study the properties of the confidence interval procedure is presented. The results indicate that in almost all of the situations simulated, the empirical Bayes estimator is significantly better than maximum likelihood with respect to reducing mean squared error. The confidence interval procedure is at the nominal level and is accurate enough for practical purposes.

SIMULTANEOUS ESTIMATION OF RISK IN SEVERAL 2 X 2

CONTINGENCY TABLES: AN EMPIRICAL BAYES APPROACH

by

JAMES STANLEY BERGUM

A thesis submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Statistics

Approved:

Head, Major Department

Chairman, Examining Committee

Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

October, 1981

## ACKNOWLEDGMENT

I wish to thank my thesis advisor Dr. Martin A. Hamilton for his advice and assistance throughout my graduate work and for suggesting the topic of this thesis. I would also like to thank Dr. John R. Lund for his assistance in evaluating some of the integrals as well as suggesting some of the numerical methods used. Appreciation is extended to Professors Richard Lund, Kenneth Tiahrt, Warren Esty, Evan Smouse, and Joe Johnson for serving on my graduate committee.

Thanks are also due to the Montana State University Research-Creativity Development Committee for funding part of the computer time necessary for the simulation.

I would like to acknowledge my parents for their encouragement and support throughout my education. Finally, I would like to thank my wife, Janet, for typing the thesis so efficiently.

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## ABSTRACT

A set of g 2×2 contingency tables with the row margins fixed may be represented as g pairs of independent binomial variates, $x_{i1}$ and $x_{i2}$, with corresponding parameters, $P_{i1}$ and $P_{i2}$, and sample sizes, $n_{i1}$ and $n_{i2}$, where $i = 1,\ldots,g$. Associated with table i is the parameter of interest, $\theta_i$, which is some specific function of $P_{i1}$ and $P_{i2}$. In this thesis, Bayes and empirical Bayes estimators are derived to simultaneously estimate $\theta_1, \theta_2,\ldots, \theta_g$. The following six choices for $\theta_i$ were studied in detail: $P_{i1}$, $P_{i2}$, $P_{i1}/(1 - P_{i1})$, $P_{i2}/(1 - P_{i2})$, $P_{i1}(1 - P_{i2})/(P_{i2}(1 - P_{i1}))$, and $P_{i1}/P_{i2}$. The last two choices, the odds ratio and the relative risk, are measures of association commonly estimated in epidemiological studies. The estimators are based on a bivariate prior distribution on the unit square which allows for a correlation between $P_{i1}$ and $P_{i2}$. The Bayes estimators are shown to be asymptotically equivalent to the maximum likelihood estimators. An ad hoc confidence interval procedure based on the empirical Bayesian estimates is given. The posterior densities for each of the parameters are presented.

A simulation study to compare the empirical Bayes estimator with the maximum likelihood estimator and to study the properties of the confidence interval procedure is presented. The results indicate that in almost all of the situations simulated, the empirical Bayes estimator is significantly better than maximum likelihood with respect to reducing mean squared error. The confidence interval procedure is at the nominal level and is accurate enough for practical purposes.

# 1. INTRODUCTION

Cross-classified tables of counts, often called contingency tables, are common in the biological and social sciences. If the data is cross-classified according to two dichotomous categorical variables, then the table is called a 2 × 2 or fourfold contingency table. An example might be the presence or absence of a risk factor, such as smoking, cross-classified with the presence or absence of a disease, such as lung cancer. Another might be the presence or absence of a head injury cross-classified with whether or not the accident victim was wearing a seat belt.

Suppose that data is to be collected and then cross-classified by two dichotomous categorical variables A and B. A 2 × 2 contingency table can be generated by three sampling designs.

1.) The total number of observations is fixed and then each observation is classified according to the presence or absence of A or B.

2.) A fixed number of observations is randomly drawn from a population possessing A and then another fixed number of observations is randomly drawn from a population not possessing A. At the conclusion of the experiment, the observations are classified as to their presence or absence of B.

3.) The sampling units are randomly assigned to A or not A with a fixed number of units in each group.

At the conclusion of the experiment, the observations are classified as to their presence or absence of B.

The difference between designs two and three is that in design two the sampling units are not randomly assigned to A and not A. They are sampled from populations where A is known to be present or absent. Friedman (1980) calls a study with sampling design two an observational study because, other than randomly selecting the subjects from the two populations, the experimenter just observes the study. He calls a study based on sampling design three an experimental study since the experimenter intervenes in the study by randomly assigning the subjects to treatment A or not A.

Epidemiological studies based on design two provided the motivation for this thesis, but the methods developed in this thesis are also applicable to designs of type three. Design two can be thought of as drawing two random samples of fixed size, one from the population possessing A and another from the population not possessing A and then counting the number of observations in each sample possessing characteristic B. The following notation will be convenient:

$X_1$ = count of observations possessing characteristic B in the sample drawn from the population possessing characteristic A;

$P_1$ = probability an observation possesses characteristic B given that it possesses characteristic A;

$n_1$ = fixed sample size drawn from population possessing A;

$X_2$ = count of observations possessing characteristic B in the sample drawn from the population not possessing characteristic A;

$P_2$ = probability an observation possesses characteristic B given that it does not possess characteristic A;

$n_2$ = fixed sample size drawn from population not possessing A.

The data from the study can be presented in a 2 × 2 table as shown in Table 1.1.

Table 1.1 THE BASIC 2 × 2 CONTINGENCY TABLE

| | characteristic B | | sample size |
|---|---|---|---|
| characteristic A | present (B) | absent (not B) | |
| present (A) | $X_1$ | $n_1 - X_1$ | $n_1$ |
| absent (not A) | $X_2$ | $n_2 - X_2$ | $n_2$ |

Note that $X_1$ and $X_2$ are independent, binomially distributed random variables.

Retrospective (case-control) and prospective (cohort) studies make use of the second sampling design. Consider the example concerning smoking and lung cancer. A retrospective study would consider cancer deaths as characteristic A and non-cancer deaths as not A and then look back in records or interview relatives to determine if the person had smoked. A prospective study would consider smokers as A and non-smokers as not A and then wait to determine if they die of cancer. Discussions of retrospective and prospective studies are found in Fleiss (1981) and Mantel and Haenszel (1959).

There are many functions of the binomial parameters which are of interest in a 2 × 2 contingency table. The following are considered in this paper:

$O_1 = P_1/(1 - P_1)$, the individual odds for B, given A is present;

$O_2 = P_2/(1 - P_2)$, the individual odds for B, given A is not present;

$OR = O_1/O_2$, the odds ratio;

$RR = P_1/P_2$, the relative risk.

$P_1$ and $O_1$ are the probability and odds, respectively, of B being present in a randomly chosen individual in the population where A is present. Similarly, $P_2$ and $O_2$ are the probability and odds of B being present in the population where A is not present.

In the smoking-lung cancer example with A being the population of smokers, a $P_1 = .04$ would have the interpretation that one out of every twenty-five smokers would die of lung cancer. The interpretation of the associated individual odds would be that for every smoker dying of lung cancer, twenty-four will not die of lung cancer. Expressing the risk as an odds gives a different emphasis in the interpretation even though the statements are equivalent.

Often a comparison between two frequencies of occurrence ($P_1$ and $P_2$) in the two populations is desired. This is common in tables where one population is a control group. RR and OR are the most common parameters used for this kind of comparison. RR is the ratio of the individual probabilities and OR is a ratio of the individual odds. If $P_1 > P_2$ then OR and RR are greater than 1. If $P_1 < P_2$ then OR and RR are less than 1. If $P_1$ and $P_2$ are equal then OR and RR equal 1. An odds ratio of 2 in the smoking-lung cancer example, with A being the population of smokers, indicates that the odds of a smoker dying of cancer is twice that of a non-smoker.

In this thesis, the symbol $\theta$ is generic for an unknown parameter. Thus $\theta$ stands for one of $P_1$, $P_2$, $O_1$, $O_2$, OR, or RR. The particular parameter of interest depends on the specific study.

The problem considered in this paper is the estimation of $\theta$ simultaneously for each of several 2 X 2 contingency tables. The problem can arise in many different situations, including:

1.) the population may have been stratified with respect to another important variable and each stratum yields a separate fourfold table;

2.) the same two variables might be cross-classified for several different populations with each population producing a separate fourfold table.

A set of g 2 X 2 contingency tables with the row margins fixed may be represented as g pairs of independent binomial variates, $X_{i1}$ and $X_{i2}$, with corresponding parameters, $P_{i1}$ and $P_{i2}$, and sample sizes, $n_{i1}$ and $n_{i2}$, where $i = 1,\ldots,g$.

The combining of information from several 2 × 2 contingency tables to calculate a pooled estimate of $\theta$ has been given considerable attention (Gart 1962, 1970; Radhakrisha 1965; McKinlay 1978; Mantel and Haenszel 1959). However, the combining of information is based on the assumption of an equal value for $\theta$, the risk parameter under consideration, across all the tables. If the investigator knows that $\theta$ differs from table to table or a hypothesis of homogeneous $\theta$ is rejected, then a combined, single estimate of $\theta$ may not be meaningful. The proper method of analysis is to give a separate estimate of $\theta$ for each table. The usual estimates given are the maximum likelihood estimates.

The maximum likelihood estimate of $\theta$ for a specific table is not influenced by the data in the other tables. It seems reasonable, however, to believe that information in the other tables might be used to advantage. If the estimates ($p_1$ and $p_2$) were plotted, a pattern indicating a correlation between $P_1$ and $P_2$ may be present. Suppose that several states each performed a retrospective study of smoking versus lung cancer and that the measure of association is different from state to state. It is reasonable to expect that the proportions of smokers in the cancer and non-cancer groups will vary together from state to state. For instance, a state like Utah may have a lower proportion of smokers in both groups while a state like New Jersey may have a high proportion of smokers in both groups due to state to state variation in socio-economic class, religion, etc. The maximum likelihood method completely ignores this possible association between $P_1$ and $P_2$. The method of analysis proposed in this paper incorporates the correlation by assuming a bivariate distribution on $P_1$ and $P_2$ and then constructs estimates of the odds ratios for the various states.

Two estimation techniques can be used to incorporate the correlation. The techniques are called Bayes and empirical Bayes and will be discussed in later sections.

The goal of this paper is to apply maximum likelihood, Bayes, and empirical Bayes estimation procedures to estimate $\theta$, which is one of the parameters $P_1$, $P_2$, $O_1$, $O_2$, OR, or RR, simultaneously over the g contingency tables.

The estimators of $\theta$ based on each estimation procedure are given in Chapter 2. A simulation study comparing the empirical Bayes estimators with the maximum likelihood estimators is presented in Chapter 3. Problems and extensions associated with the empirical Bayes estimation procedure are given in Chapter 4.

The method and results presented in Chapters 2 and 3 are original. The author believes at least some of the methods given in Chapter 4 based on a prior which is the product of two independent beta distributions can be found in the literature, but was unable to find a refererce.

# 2.1 ESTIMATION PROCEDURES

## 2.1 The Maximum Likelihood Procedure

### 2.1.1 Introduction

Suppose $\vec{x}$ is the realized value of a set of observations having joint density $f(\vec{x};\vec{\theta})$, where $\vec{\theta} = (\theta_1,\ldots,\theta_q)$ is the parameter vector belonging to a specified subset $\Theta$ of Euclidean q space. The likelihood of $\vec{\theta}$ given the observations is defined to be a function of $\vec{\theta}$:

$$L(\vec{\theta}|\vec{x}) \propto f(\vec{x};\vec{\theta}).$$

The maximum likelihood (ML) estimator of $\vec{\theta}$ is $\vec{\hat{\theta}} = (\hat{\theta}_1,\ldots,\hat{\theta}_q)$, where

$$L(\vec{\hat{\theta}}|\vec{x}) = \sup_{\vec{\theta}\in\Theta} L(\vec{\theta}|\vec{x}).$$

It is often convenient to use the log likelihood, $\ell(\vec{\theta}|\vec{x}) = \log L(\vec{\theta}|\vec{x})$, in which case $\hat{\theta}$ satisfies the equation

$$\ell(\vec{\hat{\theta}}|\vec{x}) = \sup_{\vec{\theta}\in\Theta} \ell(\vec{\theta}|\vec{x}).$$

When the supremum is attained at an interior point of $\Theta$ and $\ell(\vec{\theta}|\vec{x})$ is a differentiable function of $\vec{\theta}$, then the ML estimator is defined as any solution of the maximum likelihood equations

$$\partial\ell(\vec{\theta}|\vec{x})/\partial\theta_i = \vec{0}, \quad i = 1,\ldots,q.$$

If $\vec{\hat{\theta}}$ is the ML estimator of $\vec{\theta}$ and $h(\theta)$ is a function of $\vec{\theta}$, then the ML estimator of $h(\vec{\theta})$ is $h(\vec{\hat{\theta}})$ (Graybill 1976).

### 2.1.2 The Maximum Likelihood Estimators

Given a set of g 2 × 2 contingency tables with row totals fixed, the likelihood function is

$$L(P_{11},P_{12},P_{21},P_{22},\ldots,P_{g1},P_{g2};x_{11},x_{12},x_{21},x_{22},\ldots,x_{g1},x_{g2}) =$$

$$\prod_{i=1}^{g} \binom{n_{i1}}{x_{i1}} \binom{n_{i2}}{x_{i2}} P_{i1}^{x_{i1}} (1 - P_{i1})^{n_{i1}-x_{i1}} P_{i2}^{x_{i2}} (1 - P_{i2})^{n_{i2}-x_{i2}} ,$$

where $x_{i1}$ denotes the realized value of a binomial variate ($X_{i1}$) for row 1 in table i based on $n_{i1}$ trials with corresponding parameter $P_{i1}$, $x_{i2}$ denotes the realized value of a binomial variate ($X_{i2}$) for row 2 in table i based on $n_{i2}$ trials with corresponding parameter $P_{i2}$. Solving the maximum likelihood equations, the ML estimators for parameters $P_{i1}$ and $P_{i2}$ are $\tilde{P}_{i1} = x_{i1}/n_{i1}$ and $\tilde{P}_{i2} = x_{i2}/n_{i2}$, respectively. Since the other functions to be estimated are functions of $P_{i1}$ and $P_{i2}$, their ML estimators are:

$$\tilde{O}_{i1} = x_{i1}/(n_{i1} - x_{i1}), \qquad \widetilde{OR}_i = x_{i1}(n_{i2} - x_{i2})/(x_{i2}(n_{i1} - x_{i1})),$$

$$\tilde{O}_{i2} = x_{i2}/(n_{i2} - x_{i2}), \qquad \widetilde{RR}_i = (x_{i1}/n_{i1})(n_{i2}/x_{i2}).$$

### 2.1.3 Asymptotic Distributions for ML Estimators

As $n_i$ goes to infinity, the probability distribution of the ML estimator is closely approximated by a normal distribution. The asymptotic (large sample) distributions for the maximum likelihood estimates of $P_{i1}$ and $P_{i2}$, suitably standardized, are

$$\sqrt{n_{i1}}\left[\tilde{P}_{i1} - P_{i1}\right] \xrightarrow{L} N(0, P_{i1}(1 - P_{i1}))$$

and $$\sqrt{n_{i2}}\left[\tilde{P}_{i2} - P_{i2}\right] \xrightarrow{L} N(0, P_{i2}(1 - P_{i2}))$$

(Bishop, Fienberg and Holland 1975, ex. 14.33). Applying the δ-method (Bishop, Fienberg and Holland 1975, sec. 14.6), the asymptotic distritions for the individual odds, suitably standardized, are

$$\sqrt{n_{i1}}\left[\tilde{O}_{i1} - O_{i1}\right] \xrightarrow{L} N(0, P_{i1}/(1 - P_{i1})^3)$$

and $$\sqrt{n_{i2}}\left[\tilde{O}_{i2} - O_{i2}\right] \xrightarrow{L} N(0, P_{i2}/(1 - P_{i2})^3).$$

Let $n_i = n_{i1} + n_{i2}$ and $n_{i1}/n_i \to a_i$, $0 < a_i < 1$. Then the asymptotic distributions for OR and RR can be found by applying the multivariate version of the δ-method (Bishop, Fienberg, and Holland 1975, sec. 14.6). Since

$$\sqrt{n_i}\left[\begin{pmatrix}\tilde{P}_{i1}\\ \tilde{P}_{i2}\end{pmatrix} - \begin{pmatrix}P_{i1}\\ P_{i2}\end{pmatrix}\right] \xrightarrow{L} \mathrm{MVN}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}P_{i1}(1 - P_{i1})/a_i & 0\\ 0 & P_{i2}(1 - P_{i2})/(1 - a_i)\end{pmatrix}\right)$$

then by the multivariate δ-method,

$$\sqrt{n_i}(\widetilde{OR}_i - OR_i) \xrightarrow{L} N\left(0, \frac{P_{i1}(1 - P_{i2})}{P_{i2}^2(1 - P_{i1})^2}\left(\frac{1 - P_{i2}}{(1 - P_{i1})a_i} + \frac{P_{i1}}{P_{i2}(1 - a_i)}\right)\right) \text{ and}$$

$$\sqrt{n_i}(\widetilde{RR}_i - RR_i) \xrightarrow{L} N\left(0, \frac{P_{i1}}{P_{i2}^2}\left(\frac{1 - P_{i1}}{a_i} + \frac{P_{i1}(1 - P_{i2})}{P_{i2}(1 - a_i)}\right)\right).$$

## 2.2 The Bayesian Procedure

### 2.2.1 Introduction

The Bayesian approach is to incorporate prior information about the parameters into the estimation procedure. This is done by specifying a prior distribution function, $G(\vec{\theta})$, on the parameters. A loss function, $L(\vec{\theta},\vec{\hat{\theta}})$, is also specified to define the loss incurred when estimating $\vec{\theta}$ with the estimator $\vec{\hat{\theta}}$. Given a specified prior distribution, the Bayes estimator is the value of $\vec{\hat{\theta}}$ that minimizes the Bayes risk defined as

$$r(\vec{\theta},\vec{\hat{\theta}}) = \int R(\vec{\theta},\vec{\hat{\theta}})\ dG(\vec{\theta})$$

where $R(\vec{\theta},\vec{\hat{\theta}}) = \int L(\vec{\theta},\vec{\hat{\theta}})\ dF(\vec{x}|\vec{\theta})$ and $F(\vec{x},\vec{\theta})$ is the distribution function of the data given $\theta$.

The Bayes estimate also minimizes $\int L(\vec{\theta},\vec{\hat{\theta}})\ dF(\vec{\theta}|\vec{x})$ where $F(\vec{\theta}|\vec{x})$ denotes the distribution function of $\vec{\theta}$ given the data (Ferguson 1967). The density, if it exists, is denoted by $f(\vec{\theta}|\vec{x})$ and is called the posterior density (DeGroot 1970). The Bayes estimators in this thesis will be found using the quadratic loss function defined as $L(\vec{\theta},\vec{\hat{\theta}}) = (\vec{\hat{\theta}} - \vec{\theta})'(\vec{\hat{\theta}} - \vec{\theta})$ and then applying Rule #1 from Ferguson (1967, p. 46).

The rule states that using quadratic loss to estimate $\vec{\theta}$, a Bayes estimate, $\vec{\hat{\theta}}$, with respect to a given prior distribution for $\vec{\theta}$, is the mean of the posterior distribution of $\vec{\theta}$, given the observations.

The prior distributions, posterior distributions, and the Bayes estimates for each of the six functions will be presented in the remainder of this section. For comprehensive discussions of Bayesian inference see DeGroot (1970) and Box and Tiao (1973).

### 2.2.2 The Prior Distribution for $P_1$ and $P_2$

Several methods of constructing a bivariate distribution on the unit square were considered to be used as a prior distribution. Mardia (1970) and Ord (1972) were the primary sources for the methods. The two problems most frequently encountered when applying the methods were that the algebra required to derive the Bayes estimators became unmanageable and the range of the possible correlations between $P_1$ and $P_2$ was too small. The bivariate prior chosen is a special case of the van Uven equation (van Uven 1947) given as

$$f(x,y) = K_0(ax + b)^{e_1}(cy + d)^{e_2}(a_1x + b_1y + c_1)^{e_3}.$$

Let the density, $g^-$, be defined as

$$g^-(p_1,p_2) = K_0 p_1^{\alpha_1-1}\, p_2^{\alpha_2-1}\, (1 - \beta_1 p_1 - \beta_2 p_2)^{\alpha_3-1}$$

where $0 \le p_1 \le 1$, $\alpha_1,\alpha_2,\alpha_3 > 0$, $\alpha_3 \in$ {positive integers},

$0 \le p_2 \le 1$, $\beta_1, \beta_2 > 0$, with $\beta_1 + \beta_2 \le 1$.

The normalizing constant $K_0$ is

$$1/\left[\sum_{k=0}^{\alpha_3-1} \sum_{j=0}^{\alpha_3-k-1} \frac{(-1)^{k+j}(\alpha_3 - 1)!\, \beta_1^k\, \beta_2^j}{k!\, j!\, (\alpha_3 - k - j - 1)!\, (\alpha_1 + k)(\alpha_2 + j)}\right].$$

Substituting $(1 - p_2)$ for $p_2$ in $g^-$, define $g^+$ as

$$g^+(p_1,p_2) = K_0\, p_1^{\alpha_1-1}\, (1 - p_2)^{\alpha_2-1}\, (1 - \beta_1 p_1 - \beta_2(1 - p_2))^{\alpha_3-1}$$

where $K_0$ and the constraints remain the same. Since $\alpha_3$ is a positive integer, we can expand and rewrite $g^-$ and $g^+$ as

$$g^-(p_1,p_2) = K_0 \sum_{k=0}^{\alpha_3-1} \sum_{j=0}^{\alpha_3-k-1} \frac{(-1)^{k+j} (\alpha_3 - 1)!\, \beta_1^k \beta_2^j p_1^{\alpha_1+k-1} p_2^{\alpha_2+j-1}}{k!\, j!\, (\alpha_3 - k - j - 1)!} \quad \text{and}$$

$$g^+(p_1,p_2) = K_0 \sum_{k=0}^{\alpha_3-1} \sum_{j=0}^{\alpha_3-k-1} \frac{(-1)^{k+j} (\alpha_3 - 1)!\, \beta_1^k \beta_2^j p_1^{\alpha_1+k-1} (1 - p_2)^{\alpha_2+j-1}}{k!\, j!\, (\alpha_3 - k - j - 1)!} .$$

The prior $g^-$ is appropriate when it is known that $P_1$ and $P_2$ are negatively correlated and $g^+$ is appropriate for a known positive correlation. Note that for $\alpha_3 = 1$, the prior distribution for $P_1$ and $P_2$ is such that $P_1$ and $P_2$ are independent random variables.

To simplify notation throughout the remainder of this chapter, let

$$\Lambda = \frac{(-1)^{k+j} (\alpha_3-1)!\, \beta_1^k \beta_2^j}{k!\, j!\, (\alpha_3 - k - j - 1)!} .$$

Note that $\Lambda$ is a function of k, j, $\alpha_3$, $\beta_1$ and $\beta_2$. The indices on the double sum $\sum_{k=0}^{\alpha_3-1} \sum_{j=0}^{\alpha_3-k-1}$ will also be suppressed throughout the remainder of Chapter 2.

The marginal distributions for $g^+$ and $g^-$ are

$$g_1^-(p_1) = g_1^+(p_1) = K_0 \sum\sum \Lambda \frac{p_1^{\alpha_1+k-1}}{\alpha_2 + j} ,$$

$$g_2^-(p_2) = K_0 \; \Sigma\Sigma \; \Lambda \frac{p_2^{\alpha_2+j-1}}{\alpha_1 + k}, \quad \text{and}$$

$$g_2^+(p_2) = K_0 \; \Sigma\Sigma \; \Lambda \frac{(1 - p_2)^{\alpha_2+j-1}}{\alpha_1 + k} .$$

The moments of the prior distributions are

$$E_{g^-}(p_1^M \; p_2^N) = K_0 \; \Sigma\Sigma \; \Lambda \; /((\alpha_1 + k + M)(\alpha_2 + j + N)) \quad \text{and}$$

$$E_{g^+}(p_1^M \; p_2^N) = K_0 \; \Sigma\Sigma \; \Lambda \; B(N + 1, \; \alpha_2 + j)/(\alpha_1 + k + M),$$

where $B(\cdot,\cdot)$ is the beta function defined as

$$B(a,b) = \int_0^1 t^{a-1} \; (1 - t)^{b-1} \; dt.$$

The conditional means are

$$E_{g^-}(p_2|p_1) = \frac{\Sigma\Sigma \; \Lambda \; p_1^k/(\alpha_2 + j + 1)}{\Sigma\Sigma \; \Lambda \; p_1^k/(\alpha_2 + j)} \quad \text{and}$$

$$E_{g^+}(p_2|p_1) = \frac{\Sigma\Sigma \; \Lambda \; p_1^k/((\alpha_2 + j + 1)(\alpha_2 + j))}{\Sigma\Sigma \; \Lambda \; p_1^k/(\alpha_2 + j)} .$$

Figures 2.1 and 2.2 are examples of contour plots for the prior distributions $g^-$ and $g^+$. The conditional means are also shown.

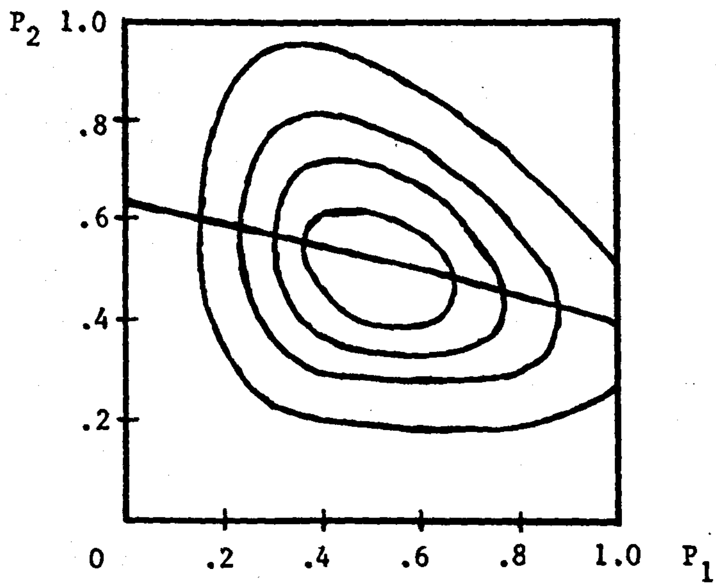



Figure 2.1 Contours and $E_{g^-}(p_2|p_1)$ with $\alpha_1 = 5$, $\alpha_2 = 6$, $\alpha_3 = 12$, $\beta_1 = .4$, $\beta_2 = .5$.

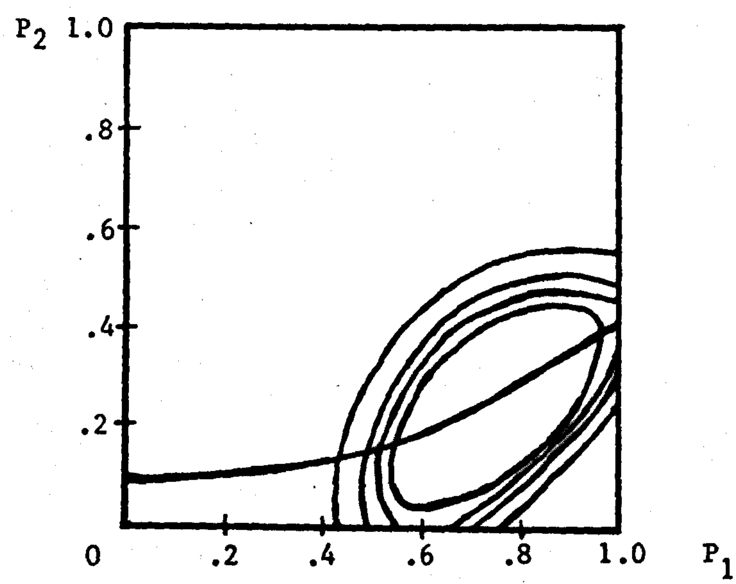



Figure 2.2 Contours and $E_{g^+}(p_2|p_1)$ with $\alpha_1 = 19$, $\alpha_2 = 19$, $\alpha_3 = 13$, $\beta_1 = .5$, $\beta_2 = .5$.

2.2.3 The Posterior Distributions

Let $P^-(p_1,p_2|x_1,x_2)$ and $P^+(p_1,p_2|x_1,x_2)$ denote the posterior distributions of $P_1$ and $P_2$ associated with $g^-$ and $g^+$, respectively.

$$P^-(p_1,p_2|x_1,x_2) = \frac{f(x_1,x_2|p_1,p_2)\ g^-(p_1,p_2)}{\iint f(x_1,x_2|p_1,p_2)\ g^-(p_1,p_2)\ dp_1\ dp_2}$$

$$= \frac{\Sigma\Sigma\ \Lambda\ p_1^{\alpha_1+x_1+k-1}\ (1-p_1)^{n_1-x_1}\ p_2^{\alpha_2+x_2+j-1}\ (1-p_2)^{n_2-x_2}}{\Sigma\Sigma\ \Lambda\ B(\alpha_1+x_1+k,\ n_1-x_1+1)\ B(\alpha_2+x_2+j,\ n_2-x_2+1)} \quad \text{and}$$

$$P^+(p_1,p_2|x_1,x_2) = \frac{f(x_1,x_2|p_1,p_2)\ g^+(p_1,p_2)}{\iint f(x_1,x_2|p_1,p_2)\ g^+(p_1,p_2)\ dp_1\ dp_2}$$

$$= \frac{\Sigma\Sigma\ \Lambda\ p_1^{\alpha_1+x_1+k-1}\ (1-p_1)^{n_1-x_1}\ p_2^{x_2}\ (1-p_2)^{n_2-x_2+\alpha_2+j-1}}{\Sigma\Sigma\ \Lambda\ B(\alpha_1+x_1+k,\ n_1-x_1+1)\ B(x_2+1,\ n_2-x_2+\alpha_2+j)}$$

where $f(x_1,x_2|p_1,p_2) = \binom{n_1}{x_1}\binom{n_2}{x_2} p_1^{x_1}\ (1-p_1)^{n_1-x_1}\ p_2^{x_2}\ (1-p_2)^{n_2-x_2}$,

$$x_1 = 0,\ldots,n_1,$$
$$x_2 = 0,\ldots,n_2,$$

is the density of the data given the parameters.

Let the denominator of $P^-(p_1,p_2|x_1,x_2)$ and $P^+(p_1,p_2|x_1,x_2)$ be defined as $D^-$ and $D^+$, respectively.

The posterior distribution for $P_1$ is the marginal distribution of $P^-$ or $P^+$ depending on the prior used. The posterior densities for

$P_1$ are

$$h^-_{P_1|x}(p_1) = \Sigma\Sigma\ \Lambda\ p_1^{\alpha_1+x_1+k-1}(1-p_1)^{n_1-x_1}B(\alpha_2+x_2+j,\ n_2-x_2+1)/D^-,$$
$$0 \le p_1 \le 1,$$

and

$$h^+_{P_1|x}(p_1) = \Sigma\Sigma\ \Lambda\ p_1^{\alpha_1+x_1+k-1}(1-p_1)^{n_1-x_1}B(x_2+1,\ n_2-x_2+\alpha_2+j)/D^+,$$
$$0 \le p_1 \le 1.$$

The posterior densities for $P_2$ are

$$h^-_{P_2|x}(p_2) = \Sigma\Sigma\ \Lambda\ p_2^{\alpha_2+x_2+j-1}(1-p_2)^{n_2-x_2}B(\alpha_1+x_1+k,\ n_1-x_1+1)/D^-,$$
$$0 \le p_2 \le 1.$$

and

$$h^+_{P_2|x}(p_2) = \Sigma\Sigma\ \Lambda\ p_2^{x_2}(1-p_2)^{n_2-x_2+\alpha_2+j-1}B(\alpha_1+x_1+k,\ n_1-x_1+1)/D^+,$$
$$0 \le p_2 \le 1.$$

The posterior distribution for $O_1$ and $O_2$ (individual odds) are found by using the change of variable

$$o_1 = \frac{p_1}{1-p_1}, \qquad o_2 = \frac{p_2}{1-p_2}$$

on $P^-$ and $P^+$ and then finding the marginal distributions of $o_1$ and $o_2$. The posterior densities of $O_1$ and $O_2$ are

$$h^-_{O_1|x}(o_1) = \Sigma\Sigma\ (\Lambda\ o_1^{\alpha_1+x_1+k-1}(1+o_1)^{-(n_1+\alpha_1+k+1)}$$
$$\times B(\alpha_2+x_2+j,\ n_2-x_2+1))/D^-, \quad 0 \le o_1 < \infty,$$

$$h^{+}_{O_1|x}(o_1) = \Sigma\Sigma\ (\Lambda\ o_1^{\alpha_1+x_1+k-1}\ (1 + o_1)^{-(n_1+\alpha_1+k+1)}$$
$$\times\ B(x_2 + 1,\ n_2 - x_2 + \alpha_2 + j))/D^{+}, \quad 0 \le o_1 < \infty,$$

$$h^{-}_{O_2|x}(o_2) = \Sigma\Sigma\ (\Lambda\ o_2^{\alpha_2+x_2+j-1}\ (1 + o_2)^{-(n_2+\alpha_2+j+1)}$$
$$\times\ B(\alpha_1 + x_1 + k,\ n_1 - x_1 + 1))/D^{-}, \quad 0 \le o_2 < \infty,$$

$$h^{+}_{O_2|x}(o_2) = \Sigma\Sigma\ (\Lambda\ o_2^{x_2}\ (1 + o_2)^{-(n_2+\alpha_2+j+1)}$$
$$\times\ B(\alpha_1 + x_1 + k,\ n_1 - x_1 + 1))/D^{+}, \quad 0 \le o_2 < \infty.$$

The posterior distribution for OR (odds ratio) is found by using the change of variable

$$or = p_1(1 - p_2)/(p_2(1 - p_1)) \quad \text{and} \quad \psi = p_2/(1 - p_2)$$

on $P^-$ and $P^+$ and then finding the marginal distribution for or. The posterior densities for OR are

$$h^{-}_{OR|x}(or) = \Sigma\Sigma\ (\Lambda\ or^{\alpha_1+x_1+k-1} \int_0^{\infty} \psi^{\alpha_1+\alpha_2+x_1+x_2+k+j-1}\ (1 + \psi)^{-(\alpha_2+n_2+j+1)}$$
$$\times\ (1 + or\cdot\psi)^{-(\alpha_1+n_1+k+1)}\ d\psi)/D^{-}, \quad 0 \le or < \infty \quad \text{and}$$

$$h^{+}_{OR|x}(or) = \Sigma\Sigma\ (\Lambda\ or^{\alpha_1+x_1+k-1} \int_0^{\infty} \psi^{\alpha_1+x_1+x_2+k}\ (1 + \psi)^{-(n_2+\alpha_2+j+1)}$$
$$\times\ (1 + or\cdot\psi)^{-(n_1+\alpha_1+k+1)}\ d\psi)/D^{+}, \quad 0 \le or < \infty.$$

The posterior distribution for RR (relative risk) is found by using the change of variable

$$rr = p_1/p_2 \qquad \text{and} \qquad \psi = p_2$$

on $P^-$ and $P^+$ and then finding the marginal density for rr. The posterior densities for RR are

$$h^-_{RR|x}(rr) = \Sigma\Sigma\,(\Lambda\, rr^{\alpha_1+x_1+k-1}\int_0^1 (1 - rr\cdot\psi)^{n_1-x_1}\,\psi^{\alpha_1+\alpha_2+x_1+x_2+k+j-1}$$

$$\times\,(1 - \psi)^{n_2-x_2}\,d\psi)/D^-, \quad \text{if } 0 \le rr \le 1,$$

$$= \Sigma\Sigma\,(\Lambda\, rr^{\alpha_1+x_1+k-1}\int_0^{1/rr} (1 - rr\cdot\psi)^{n_1-x_1}\,\psi^{\alpha_1+\alpha_2+x_1+x_2+k+j-1}$$

$$\times\,(1 - \psi)^{n_2-x_2}\,d\psi)/D^-, \quad \text{if } rr > 1 \quad \text{and}$$

$$h^+_{RR|x}(rr) = \Sigma\Sigma\,(\Lambda\, rr^{\alpha_1+x_1+k-1}\int_0^1 (1 - rr\cdot\psi)^{n_1-x_1}\,\psi^{\alpha_1+x_1+x_2+k}$$

$$\times\,(1 - \psi)^{n_2-x_2+\alpha_2+j-1}\,d\psi)/D^+, \quad \text{if } 0 \le rr \le 1,$$

$$= \Sigma\Sigma\,(\Lambda\, rr^{\alpha_1+x_1+k-1}\int_0^{1/rr} (1 - rr\cdot\psi)^{n_1-x_1}\,\psi^{\alpha_1+x_1+x_2+k}$$

$$\times\,(1 - \psi)^{n_2-x_2+\alpha_2+j-1}\,d\psi)/D^+, \quad \text{if } rr > 1.$$

### 2.2.4 The Bayes Estimators

Applying the rule from Ferguson given in 2.2.1, the means of the posterior distribution are the Bayes estimators.

The Bayes estimators using $g^-$ are:

$$\hat{P}_1^- = \sum\sum w^-_{k,j,n_1,n_2}\,(\tilde{P}_1,\tilde{P}_2)\,\frac{(\alpha_1 + x_1 + k)}{(\alpha_1 + n_1 + k + 1)}$$

$$\hat{P}_2^- = \sum\sum w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_2 + x_2 + j)}{(\alpha_2 + n_2 + j + 1)},$$

$$\hat{O}_1^- = \sum\sum w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(n_1 - x_1)},$$

$$\hat{O}_2^- = \sum\sum w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_2 + x_2 + j)}{(n_2 - x_2)},$$

$$\hat{OR}^- = \sum\sum w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(n_1 - x_1)} \cdot \frac{(n_2 - x_2 + 1)}{(\alpha_2 + x_2 + j - 1)}, \text{ and}$$

$$\hat{RR}^- = \sum\sum w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(\alpha_1 + n_1 + k + 1)} \cdot \frac{(\alpha_2 + n_2 + j)}{(\alpha_2 + x_2 + j - 1)},$$

where $w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2)$

$$= \frac{\Lambda\ B(\alpha_1 + n_1\tilde{P}_1 + k,\ n_1 - n_1\tilde{P}_1 + 1)\ B(\alpha_2 + n_2\tilde{P}_2) + j,\ n_2 - n_2\tilde{P}_2 + 1)}{D^-}.$$

Notice that $\sum\sum w^-_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) = 1.$

The Bayes estimators using $g^+$ are:

$$\hat{P}_1^+ = \sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(\alpha_1 + n_1 + k + 1)},$$

$$\hat{P}_2^+ = \sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(x_2 + 1)}{(\alpha_2 + n_2 + j + 1)},$$

$$\hat{O}_1^+ = \sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(n_1 - x_1)},$$

$$\hat{O}_2^+ = \sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(x_2 + 1)}{(\alpha_2 + n_2 - x_2 + j - 1)},$$

$$\hat{OR}^+ = \sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(n_1 - x_1)} \cdot \frac{(\alpha_2 + n_2 - x_2 + j)}{x_2}, \text{ and}$$

$$\hat{RR}^+ = \sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) \frac{(\alpha_1 + x_1 + k)}{(\alpha_1 + n_1 + k + 1)} \cdot \frac{(\alpha_2 + n_2 + j)}{x_2},$$

where $w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2)$

$$= \frac{\Lambda\, B(\alpha_1 + n_1\tilde{P}_1 + k,\ n_1 - n_1\tilde{P}_1 + 1)\, B(n_2\tilde{P}_2 + 1,\ n_2 - n_2\tilde{P}_2 + \alpha_2 + j)}{D^+}.$$

Notice that $\sum\sum w^+_{k,j,n_1,n_2} (\tilde{P}_1,\tilde{P}_2) = 1.$

### 2.2.5 Asymptotic Behavior of the Bayes Estimators

In this section, the Bayes estimators are shown to have the same asymptotic distribution as the maximum likelihood estimators. Therefore, for large sample size, one can use the ML estimates and construct confidence intervals based on the variances given for the ML estimators in Section 2.1.3. Several lemmas are necessary before proving the main theorem of the section.

Lemma 2.1

If $\lim\limits_{n\to\infty} \dfrac{\partial f_n(x)}{\partial x}$ exists, $f_n(x) \neq 0$ for each n, and $\lim\limits_{n\to\infty} f_n(x) \neq 0$,

then $\lim\limits_{n\to\infty} \dfrac{\partial \frac{1}{f_n(x)}}{\partial x}$ exists.

Proof:

Since $\lim\limits_{n\to\infty} f_n'(x)$ exists and $\lim\limits_{n\to\infty} f_n(x) \neq 0$, then $\dfrac{-\lim\limits_{n\to\infty} f_n'(x)}{\left(\lim\limits_{n\to\infty} f_n(x)\right)^2}$

exists. But $\dfrac{-\lim\limits_{n\to\infty} f_n'(x)}{\left(\lim\limits_{n\to\infty} f_n(x)\right)^2} = \lim\limits_{n\to\infty} \dfrac{-f_n'(x)}{(f_n(x))^2} = \lim\limits_{n\to\infty} \dfrac{\partial \frac{1}{f_n(x)}}{\partial x}$.

Lemma 2.2

Let $n_1/n \to a$, $n_2/n \to b$, then

$$\lim_{n\to\infty} \frac{\partial w^-_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_1}, \qquad \lim_{n\to\infty} \frac{\partial w^-_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2}$$

$$\lim_{n\to\infty} \frac{\partial w^+_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_1}, \qquad \lim_{n\to\infty} \frac{\partial w^+_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2}$$

exist for all $k + j \leq \alpha_3 - 1$.

Proof:

The proof is given for $w^-$. The proof for $w^+$ is similar.

Let $w^*_{k,j,n_1,n_2}(P_1,P_2) = \dfrac{1}{w^-_{k,j,n_1,n_2}(P_1,P_2)}$. The method of proof is to show that $\lim\limits_{n\to\infty} \dfrac{\partial w^*_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_1}$ and $\lim\limits_{n\to\infty} \dfrac{\partial w^*_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2}$ exist and then apply lemma 2.1.

Define $\Lambda_{r,s} = \dfrac{(\alpha_3 - 1)!\,(-1)^{r+s}\,\beta_1^r\,\beta_2^s}{r!\,s!\,(\alpha_3 - 1 - r - s)!}$ .

$$w^*_{k,j,n_1,n_2}(P_1,P_2) = \sum_{k'=0}^{\alpha_3-1}\ \sum_{j'=0}^{\alpha_3-1-k'} \frac{\Lambda_{k',j'}\,C_{k'}\,D_{j'}}{\Lambda_{k,j}}$$

where $C_{k'} = \dfrac{\Gamma(\alpha_1 + n_1P_1 + k')}{\Gamma(\alpha_1 + n_1P_1 + k)} \cdot \dfrac{\Gamma(\alpha_1 + n_1 + k + 1)}{\Gamma(\alpha_1 + n_1 + k' + 1)}$ and

$$D_{j'} = \frac{\Gamma(\alpha_2 + n_2P_2 + j')}{\Gamma(\alpha_2 + n_2P_2 + j)} \cdot \frac{\Gamma(\alpha_2 + n_2 + j + 1)}{\Gamma(\alpha_2 + n_2 + j' + 1)} .$$

If $k' = 0$, then

$$C_0 = 0, \qquad \text{for } k = 0,$$

$$= \frac{\Gamma(\alpha_1 + n_1P_1)}{\Gamma(\alpha_1 + n_1P_1 + k)} \cdot \frac{\Gamma(\alpha_1 + n_1 + k + 1)}{\Gamma(\alpha_1 + n_1 + 1)}$$

$$= \frac{(\alpha_1 + n_1 + k) \ldots (\alpha_1 + n_1 + 1)}{(\alpha_1 + n_1P_1 + k - 1) \ldots (\alpha_1 + n_1P_1)}, \qquad \text{for } k = 1,\ldots,\alpha_3\text{-}1.$$

If $j' = 0$, then

$$D_0 = 0, \qquad \text{for } j = 0.$$

$$= \frac{\Gamma(\alpha_2 + n_2P_2)}{\Gamma(\alpha_2 + n_2P_2 + j)} \cdot \frac{\Gamma(\alpha_2 + n_2 + j + 1)}{\Gamma(\alpha_2 + n_2 + 1)}$$

$$= \frac{(\alpha_2 + n_2 + j) \ldots (\alpha_2 + n_2 + 1)}{(\alpha_2 + n_2P_2 + j - 1) \ldots (\alpha_2 + n_2P_2)}, \qquad \text{for } j = 1,\ldots,\alpha_3\text{-}1.$$

Notice that $a_0$ and $b_0$ are the reciprocals of non-zero polynomials on $[0,1]$. Also, $\lim_{n\to\infty} C_0 = (1/P_1)^k$ and $\lim_{n\to\infty} D_0 = (1/P_2)^j$.

We then have

$$\lim_{n\to\infty} \frac{\partial C_0}{\partial P_1} = \frac{\partial \lim_{n\to\infty} C_0}{\partial P_1} = \frac{\partial\left(\frac{1}{P_1}\right)^k}{\partial P_1} = \frac{-k}{P_1^{k+1}} \qquad \text{and}$$

$$\lim_{n\to\infty} \frac{\partial D_0}{\partial P_2} = \frac{-j}{P_2^{j+1}} .$$

With the recursion relationships

$$C_{k'+1} = \frac{\alpha_1 + n_1P_1 + k'}{\alpha_1 + n_1 + k' + 1} C_{k'}, \qquad k' = 0,\ldots,\alpha_3-2$$

$$D_{j'+1} = \frac{\alpha_2 + n_2P_2 + j'}{\alpha_2 + n_2 + j' + 1} D_{j'}, \qquad j' = 0,\ldots,\alpha_3-2$$

and

$$\lim_{n\to\infty} \frac{\alpha_1 + n_1P_1 + k'}{\alpha_1 + n_1 + k' + 1} = P_1, \qquad \lim_{n\to\infty} \frac{\alpha_2 + n_2P_2 + j}{\alpha_2 + n_2 + j' + 1} = P_2,$$

$$\lim_{n\to\infty} \frac{\partial\left(\dfrac{\alpha_1 + n_1P_1 + k'}{\alpha_1 + n_1 + k' + 1}\right)}{\partial P_1} = 1, \text{ and } \lim_{n\to\infty} \frac{\partial\left(\dfrac{\alpha_1 + n_2P_2 + k'}{\alpha_2 + n_2 + j' + 1}\right)}{\partial P_2} = 1,$$

we have

$$\lim_{n\to\infty} \frac{\partial\, w^*_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_1} = \sum\sum \frac{\Lambda_{k',j'}}{\Lambda_{k,j}} \lim_{n\to\infty} \frac{\partial C_{k'}}{\partial P_1} \lim_{n\to\infty} D_{j'}$$

$$= \sum\sum \frac{\Lambda_{k',j'}}{\Lambda_{k,j}} \frac{-k}{P_1^{k+1}} \frac{P_2^{j'}}{P_2^{j}}$$

$$= \frac{-k}{\Lambda_{k,j}\, P_1^{k+1}\, P_2^{j}} \sum\sum \frac{(\alpha_3 - 1)!\, (-\beta_1)^{k'}\, (-P_1\beta_2)^{j'}}{k'!\, j'!\, (\alpha_3 - 1 - k' - j')!}$$

$$= \frac{-k}{\Lambda_{k,j}\, P_1^{k+1}\, P_2^{j}} (1 - \beta_1 - \beta_2P_2)^{\alpha_3-1}$$

and

$$\lim_{n\to\infty} \frac{\partial\, w^*_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2} = \frac{-j}{\Lambda_{k,j}\; P_1^k\; P_2^{j+1}} (1 - \beta_1 P_1 - \beta_2)^{\alpha_3 - 1}.$$

Thus $\lim\limits_{n\to\infty} \dfrac{\partial\, w^*_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_1}$ and $\lim\limits_{n\to\infty} \dfrac{\partial\, w^*_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2}$ exist.

Also

$$\lim_{n\to\infty} w^*_{k,j,n_1,n_2}(P_1,P_2) = \sum\sum \frac{\Lambda_{k',j'}}{\Lambda_{k,j}} \lim_{n\to\infty} C_{k'} \lim_{n\to\infty} D_{j'}$$

$$= \frac{1}{\Lambda_{k,j}\; P_1^k\; P_2^j} \sum\sum \frac{(\alpha_3 - 1)!\; (-\beta_1 P_1)^{k'} (-\beta_2 P_2)^{j'}}{k'!\; j'!\; (\alpha_3 - 1 - k' - j')!}$$

$$= \frac{(1 - \beta_1 P_1 - \beta_2 P_2)^{\alpha_3 - 1}}{\Lambda_{k,j}\; P_1^k\; P_2^j}$$

$$\neq 0.$$

So by lemma 2.1, the result follows.

Lemma 2.3

If $\sum\limits_i f_{i,n}(x_1,x_2)$ is a finite sum with $\lim\limits_{n\to\infty} \dfrac{\partial\, f_{i,n}(x_1,x_2)}{\partial x_1}$ and

$\lim\limits_{n\to\infty} \dfrac{\partial\, f_{i,n}(x_1,x_2)}{\partial x_2}$ existing for each i and $\lim\limits_{n\to\infty} g_{i,n}(x_1,x_2) = h(x_1,x_2)$, then $\lim\limits_{n\to\infty} \sum\limits_i \left[ \dfrac{\partial f_{i,n}(x_1,x_2)}{\partial x_j} \quad g_{i,n}(x_1,x_2) \right] = 0$, where $j = 1,2$.

Proof:

$$\lim_{n\to\infty} \sum_i \left[ \frac{\partial f_{i,n}(x_1,x_2)}{\partial x_j} \quad g_{i,n}(x_1,x_2) \right]$$

$$= \sum_i \left[ \lim_{n\to\infty} \frac{\partial f_{i,n}(x_1,x_2)}{\partial x_j} \lim_{n\to\infty} g_{i,n}(x_1,x_2) \right] = \sum_i \left[ \lim_{n\to\infty} \frac{\partial f_{i,n}(x_1,x_2)}{\partial x_j} h(x_1,x_2) \right]$$

$$= h(x_1,x_2) \lim_{n\to\infty} \frac{\partial \sum\limits_i f_{i,n}(x_1,x_2)}{\partial x_j} = h(x_1,x_2) \lim_{n\to\infty} \frac{\partial c}{\partial x_j} = 0.$$

Lemma 2.4

If $f(x,y)$ has continuous first partials, then the Taylor expansion of $f$ at $(a,b)$

$$f(x,y) = f(a,b) + \frac{\partial f(a,b)}{\partial a} (x - a) + \frac{\partial f(a,b)}{\partial b} (y - b) + R$$

is such that

$$R = \left[ \left. \frac{\partial f(a,b)}{\partial a} \right|_{\substack{a=a^* \\ b=b^*}} - \frac{\partial f(a,b)}{\partial a} \right] (x - a) + \left[ \left. \frac{\partial f(a,b)}{\partial b} \right|_{\substack{a=a^* \\ b=b^*}} - \frac{\partial f(a,b)}{\partial b} \right] (y - b),$$

where $a^*$ is between $x$ and $a$ and $b^*$ is between $y$ and $b$ and $R = o(\max \{|x - a|, |y - b|\})$ as $x \to a$ and $y \to b$.

Proof:

There exists an $a^*$ and $b^*$ such that for $a^*$ between x and a and $b^*$ between y and b, the following equality holds for the Taylor expansion of f(x,y) about the point (a,b).

$$f(x,y) = f(a^*,b^*) + \left.\frac{\partial f(a,b)}{\partial a}\right|_{\substack{a=a_*^*\\ b=b^*}} (x - a) + \left.\frac{\partial f(a,b)}{\partial b}\right|_{\substack{a=a_*^*\\ b=b^*}} (y - b)$$

$$\Rightarrow f(x,y) = f(a^*,b^*) + \frac{\partial f(a,b)}{\partial a} (x - a) + \frac{\partial f(a,b)}{\partial b} (y - b) + R$$

where

$$R = \left[\left.\frac{\partial f(a,b)}{\partial a}\right|_{\substack{a=a_*^*\\ b=b^*}} - \frac{\partial f(a,b)}{\partial a}\right](x - a) + \left[\left.\frac{\partial f(a,b)}{\partial b}\right|_{\substack{a=a_*^*\\ b=b^*}} - \frac{\partial f(a,b)}{\partial b}\right](y-b).$$

Now

$$0 \leq \frac{|R|}{\max\{|x - a|, |y - b|\}}$$

$$\leq \left\|\left[\left.\frac{\partial f(a,b)}{\partial a}\right|_{\substack{a=a_*^*\\ b=b^*}} - \frac{\partial f(a,b)}{\partial a}\right] - \left[\left.\frac{\partial f(a,b)}{\partial b}\right|_{\substack{a=a_*^*\\ b=b^*}} - \frac{\partial f(a,b)}{\partial b}\right]\right\|,$$

but since f has continuous first partials,

$$\left.\frac{\partial f(a,b)}{\partial a}\right|_{\substack{a=a_*^*\\ b=b^*}} \longrightarrow \frac{\partial f(a,b)}{\partial a} \quad \text{and} \quad \left.\frac{\partial f(a,b)}{\partial b}\right|_{\substack{a=a_*^*\\ b=b^*}} \longrightarrow \frac{\partial f(a,b)}{\partial b}$$

as $x \to a$ and $y \to b$ so that the right side of the inequality approaches zero and therefore

$$R = o(\max\{|x - a|, |y - b|\}).$$

Theorem 2.1

The Bayes estimators have the same asymptotic distribution as the maximum likelihood estimators.

Proof:

A proof is given for $\hat{OR}^-$. Proofs for the other Bayes estimators follow similar arguments. Assume that $n_1/n \to a$ and $n_2/n \to b$ $(a + b = 1)$.

$$\hat{OR}^- = g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} = \Sigma\Sigma\, w^-_{k,j,n_1,n_2}(\tilde{P}_1,\tilde{P}_2)\,\frac{(\tilde{P}_1 + (\alpha_1 + k)/n_1)(1 - \tilde{P}_2 + 1/n_2)}{(1 - \tilde{P}_1)(\tilde{P}_2 + (\alpha_2 + j - 1)/n_2)}$$

where $\tilde{P}_1 = x_1/n_1$ and $\tilde{P}_2 = x_2/n_2$.

We need to show

$$\sqrt{n}\left[g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} - \frac{P_1(1 - P_2)}{P_2(1 - P_1)}\right] \xrightarrow{L} N(0,\sigma^2_{P_1,P_2})$$

where $\sigma^2_{P_1,P_2} = \dfrac{P_1(1 - P_2)}{P_2^2(1 - P_1)^2}\left(\dfrac{1 - P_2}{(1 - P_1)a} + \dfrac{P_1}{P_2 b}\right)$.

The method of proof is to apply the following lemma (Rao 1973, p. 122).

Lemma 2.5

If

i) $\sqrt{n}\left[g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} - g\begin{pmatrix}P_1\\ P_2\end{pmatrix}\right] \xrightarrow{L} N(0,\sigma^2_{P_1,P_2})$

ii) $\sqrt{n}\left[\dfrac{P_1(1 - P_2)}{P_2(1 - P_1)} - g\begin{pmatrix}P_1\\ P_2\end{pmatrix}\right] \xrightarrow{P} 0$

then $\sqrt{n}\left[g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} - \dfrac{P_1(1 - P_2)}{P_2(1 - P_1)}\right] \xrightarrow{L} N(0,\sigma^2_{P_1,P_2})$.

Proof of Theorem 2.1:

First to show (i) of Lemma 2.5.

i) The Taylor expansion of g about $\begin{pmatrix}P_1\\ P_2\end{pmatrix}$ is

$$g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} = g\begin{pmatrix}P_1\\ P_2\end{pmatrix} + [\mathit{1} + \mathit{2}]\ (\tilde{P}_1 - P_1) + [\mathit{3} + \mathit{4}]\ (\tilde{P}_2 - P_2) + \text{Remainder } (R)$$

where

$$\mathit{1} = \Sigma\Sigma\ w^{-}_{k,j,n_1,n_2}\ (P_1,P_2)\ \frac{(1 + (\alpha_1 + k)/n_1)}{(1 - P_1)^2} \cdot \frac{(1 - P_2 + 1/n_2)}{(P_2 + (\alpha_2 + j - 1)/n_2)},$$

$$\mathit{2} = \Sigma\Sigma\ \frac{\partial w^{-}_{k,j,n_1,n_2}\ (P_1,P_2)}{\partial P_1}\ \frac{(P_1 + (\alpha_1 + k)/n_1)}{(1 - P_1)}\ \frac{(1 - P_2 + 1/n_2)}{(P_2 + (\alpha_2 + j - 1)/n_2)},$$

$$\mathit{3} = \Sigma\Sigma\ w^{-}_{k,j,n_1,n_2}(P_1,P_2)\ \frac{(-1 - (\alpha_2 + j)/n_2)}{(P_2 + (\alpha_2 + j - 1)/n_2)^2} \cdot \frac{(P_1 + (\alpha_1 + k)/n_1)}{(1 - P_1)},$$

$$\mathit{4} = \Sigma\Sigma\ \frac{\partial w^{-}_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2}\ \frac{(1 - P_2 + 1/n_2)}{(P_2 + (\alpha_2 + j - 1)/n_2)}\ \frac{(P_1 + (\alpha_1 + k)/n_1)}{(1 - P_1)},$$

$$\sqrt{n}\left[g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} - g\begin{pmatrix}P_1\\ P_2\end{pmatrix}\right]$$

$$= [\mathit{1} + \mathit{2}]\sqrt{n}\ (\tilde{P}_1 - P_1) + [\mathit{3} + \mathit{4}]\sqrt{n}\ (\tilde{P}_2 - P_2) + \sqrt{n}\,R.$$

From lemma 2.2, $\lim_{n\to\infty} w^{-}_{k,j,n_1,n_2}(P_1,P_2)$, $\lim_{n\to\infty} \dfrac{\partial\, w^{-}_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_1}$,

and $\lim_{n\to\infty} \dfrac{\partial\, w^{-}_{k,j,n_1,n_2}(P_1,P_2)}{\partial P_2}$ exist for any $P_1$ and $P_2$ in the open interval

(0,1).

Recalling $\Sigma\Sigma\; w^{-}_{k,j,n_1,n_2}(P_1,P_2) = 1$,

$$\lim_{n\to\infty} \mathit{1} = \frac{(1 - P_2)}{(1 - P_1)^2 P_2} \qquad \text{and} \qquad \lim_{n\to\infty} \mathit{3} = \frac{-P_1}{P_2^2(1 - P_1)} \quad .$$

Applying lemma 2.3,

$$\lim_{n\to\infty} \mathit{2} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \mathit{4} = 0.$$

By lemma 2.4 $g\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix}$ has continuous first partials on (0,1). Noting

that the gamma function is continuous and differentiable,

$$R = o_p(\max\ \{|\tilde{P}_1 - P_1|, |\tilde{P}_2 - P_2|\})$$

$$\sqrt{n}R = o_p(\max\ \{|\ n(\tilde{P}_1 - P_1)|, |\sqrt{n}(\tilde{P}_2 - P_2)|\}).$$

But $\sqrt{n}\,(\tilde{P}_1 - P_1\ ) = O_p(1)$ and $\sqrt{n}\,(\tilde{P}_2 - P_2) = O_p(1)$

(Bishop, Fienberg, and Holland 1975) so $\sqrt{n}\,R = o_p(O_p(1)) = o_p(1)$.

We now have that

$$[1 + 2] \xrightarrow{P} \frac{(1 - P_2)}{(1 - P_1)^2 P_2} \quad \text{and} \quad [3 + 4] \xrightarrow{P} \frac{-P_1}{P_2^2(1 - P_1)} \quad .$$

Since $\sqrt{n}(\tilde{P}_1 - P_1) \xrightarrow{L} N(0,(P_1(1 - P_1))/a)$ and

$\sqrt{n}(\tilde{P}_2 - P_2) \xrightarrow{L} N(0,(P_2(1 - P_2))/b)$, we can apply Slutsky's Theorem (Rao 1973, p. 122). Thus

$$\sqrt{n}\; g\left[\begin{pmatrix}\tilde{P}_1\\ \tilde{P}_2\end{pmatrix} - g\begin{pmatrix}P_1\\ P_2\end{pmatrix}\right] \xrightarrow{L} N(0,\sigma^2_{P_1,P_2}).$$

Now to show (ii) of Lemma 2.5.

$$\text{ii)}\quad \lim_{n\to\infty} \sqrt{n}\left[\frac{P_1(1 - P_2)}{P_2(1 - P_1)} - g\begin{pmatrix}P_1\\ P_2\end{pmatrix}\right]$$

$$= \lim_{n\to\infty} \sqrt{n}\left[\frac{P_1(1 - P_2)}{P_2(1 - P_1)} - \Sigma\Sigma\left(\bar{w}_{k,j,n_1,n_2}(P_1,P_2)\,\frac{(P_1 + (\alpha_1 + k)/n_1)}{(1 - P_1)}\right.\right.$$

$$\left.\left.\times \frac{(1 - P_2 + 1/n_2)}{(P_2 + (\alpha_2 + j - 1)/n_2)}\right)\right]$$

$$= \lim_{n\to\infty} \sqrt{n}\;\; \Sigma\Sigma\left[\bar{w}_{k,j,n_1,n_2}(P_1,P_2)\left(\frac{P_1(1 - P_2)}{P_2(1 - P_1)} - \frac{(P_1 + (\alpha_1 + k)/n_1)}{(1 - P_1)}\right.\right.$$

$$\left.\left.\times \frac{(1 - P_2 + 1/n_2)}{(P_2 + (\alpha_2 + j - 1)/n_2)}\right)\right]$$

$$= \frac{1}{(1 - P_1)P_2} \Sigma\Sigma \left[ \left\{ \lim_{n\to\infty} w^{-}_{k,j,n_1,n_2} (P_1,P_2) \left[ \frac{P_1(\alpha_2 + j - 1 - P_2)}{n_2/n} \right. \right. \right.$$

$$\left. \left. \left. - \frac{P_2(\alpha_1 + k)}{n_1/n} \left( (1 - P_2) + \frac{1}{n(n_2/n)} \right) \right] \right\} \Bigg/ \left[ \sqrt{n} \left( P_2 + \frac{(\alpha_2 + j - 1)}{n_2/n} \right) \right] \right].$$

Since $\lim_{n\to\infty} w^{-}_{k,j,n_1,n_2} (P_1,P_2)$ exists for any $P_1$ and $P_2$ in (0,1),

the $\lim_{n\to\infty} \sqrt{n} \left[ \frac{P_1(1 - P_2)}{P_2(1 - P_1)} - g \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \right] \xrightarrow{P} 0.$

## 2.3 Empirical Bayes Procedure

### 2.3.1 Introduction

If the hyperparameters $\alpha_1$, $\alpha_2$, $\alpha_3$, $\beta_1$, and $\beta_2$ of the prior density are known, then the estimators derived using the Bayesian procedure will minimize the Bayes risk under squared error loss. If one is unwilling to specify the hyperparameters, then empirical methods can be used. Robbins (1956) was the first to suggest an empirical Bayes approach to the problem of estimation when the prior is not completely known. Since Robbins' paper, other empirical methods have been proposed. The method used in the present paper is to assume a $g^-$ or $g^+$ family of priors with the parameters unknown. The hyperparameters are then estimated by some kind of non-Bayesian technique. Then the empirical Bayes (EB) formulas are the same as the Bayesian procedure with the exception that the hyperparameters are replaced by their estimates. Good (1980) suggests that the method be called Semi-Empirical Bayesian, or Hyperparametric Empirical Bayesian, while Bishop, Fienberg, and Holland (1975) suggest the name Pseudo-Bayesian.

### 2.3.2 Estimation of Hyperparameters

Let the family of prior densities be of the family $g^+$ or $g^-$ with hyperparameters $\alpha_1$, $\alpha_2$, $\alpha_3$, $\beta_1$, and $\beta_2$.

The marginal likelihood functions for the hyperparameters are

$$L^-(\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2;\ (x_{i1},x_{i2}),\ i = 1,\ldots,g) = \prod_{i=1}^{g} m^-(x_{i1},x_{i2})$$

for prior $g^-$ and

$$L^+(\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2;\ (x_{i1},x_{i2}),\ i = 1,\ldots,g) = \prod_{i=1}^{g} m^+(x_{i1},x_{i2})$$

for prior $g^+$, where

$$m^-(x_{i1},x_{i2}) = K_0 \ \Sigma\Sigma\ \Lambda \binom{n_{i1}}{x_{i1}}\binom{n_{i2}}{x_{i2}} B(\alpha_1 + x_{i1} + k,\ n_{i1} - x_{i1} + 1)$$
$$\times\ B(\alpha_2 + x_{i2} + j,\ n_{i2} - x_{i2} + 1) \quad \text{and}$$

$$m^+(x_{i1},x_{i2}) = K_0 \ \Sigma\Sigma\ \Lambda \binom{n_{i1}}{x_{i1}}\binom{n_{i2}}{x_{i2}} B(\alpha_1 + x_{i1} + k,\ n_{i1} - x_{i1} + 1)$$
$$\times\ B(x_{i2} + 1,\ n_{i2} - x_{i2} + \alpha_2 + j),$$

which are the marginal densities of $x_{i1}$ and $x_{i2}$ for table i. One can use $L^-$ and $L^+$ to find the ML estimator of the hyperparameters.

A closed form solution for the hyperparameters was not found. A method for finding the ML estimate is given in Section 3.3.

### 2.3.3 Asymptotic Results

The Bayes estimates were shown to be asymptotically equivalent to the maximum likelihood estimates in Section 2.2.5. Thus the effect of the prior, regardless of the values of $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$, is lost as the sample size gets larger (i.e., the data dominates the prior as the sample size increases).

The empirical Bayes estimates have the hyperparameters replaced by their maximum likelihood estimates. The maximum likelihood estimates are consistent estimators and so with respect to $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$, the empirical Bayes estimators are converging to the Bayes estimators. This leads the author to believe that the empirical Bayes estimates are at least consistent and perhaps asymptotically equivalent to the ML estimators but he is unable to give a rigorous proof.

### 2.3.4 Confidence Intervals

In an attempt to construct a confidence interval based on the empirical Bayes estimators that would have a frequency interpretation, the $x_1$ and $x_2$ counts in each table are adjusted so that the ML estimate for the adjusted table is identical to the EB estimate for the original, unadjusted frequencies.

For example, given a table with $x_1 = 40$, $x_2 = 10$, $n_1 = 50$, and $n_2 = 50$, suppose the EB estimates are

$$\hat{P}_1 = .81 \qquad \hat{P}_2 = .19 \qquad \hat{O}_1 = 4.4$$

$$\hat{O}_2 = .22 \qquad \hat{OR} = 15.4 \quad \text{and} \quad \hat{RR} = 3.8.$$

Then the adjusted $x_1,x_2$ values for constructing the confidence interval for a specific parameter are determined by finding the $x_1,x_2$ values which give the ML estimate that is equal to the EB estimate. To estimate $P_1$, $x_1(\text{adj}) = 40.5$; to estimate $P_2$, $x_2(\text{adj}) = 9.5$; to estimate $O_1$, $x_1(\text{adj}) = 40.74$; and to estimate $O_2$, $x_2(\text{adj}) = 9.016$. Note that the adjusted counts in a particular table depend on what function of $P_1$ or $P_2$ is to be estimated. To provide unique adjusted counts given an EB estimate for OR or RR, the sum of the observed counts, $x_1$ and $x_2$, is assumed fixed.

Continuing with the example, assume $x_1 + x_2 = 50$ is fixed, then to estimate OR, we find that

$$\hat{OR} = \frac{x_1(n_2 - x_2)}{x_2(n_1 - x_1)} \quad \Rightarrow \quad 15.4 = \frac{x_1(50 - x_2)}{x_2(50 - x_1)}$$

$$\Rightarrow \quad x_1(\text{adj}) = 36.819 \text{ and } x_2(\text{adj}) = 13.181.$$

Similarly, to estimate RR, $x_1(\text{adj}) = 39.583$ and $x_2(\text{adj}) = 10.417$.

Confidence intervals based on large row totals can be constructed using the asymptotic results from Chapter 2.

For small row totals, confidence intervals based on the ML and EB estimators are determined using confidence interval techniques described below.

The confidence intervals for $P_1$ and $P_2$ are found by using the tabled exact 95% confidence interval for the binomial distribution in Snedecor and Cochran (1980). Linear interpolation is used for the adjusted counts.

The confidence intervals for $O_1$ and $O_2$ are found by transforming the estimate of $O_1$ and $O_2$ back to a $P_1$ or $P_2$ estimate and then using the tabled values as for $P_1$ and $P_2$.

The confidence intervals for OR are found by using the Cornfield Approximation Algorithm (Fleiss 1979). The Cornfield Approximation Algorithm is an iterative procedure to solve a quartic equation. The initial guess is to use the confidence interval suggested by Woolf (1955). Let

$$L = \ln (OR) = \ln \left[ \frac{(x_1 + .5)(n_2 - x_2 + .5)}{(x_2 + .5)(n_1 - x_1 + .5)} \right];$$

then L has standard error

$$S(L) = \left( \frac{1}{x_1 + .5} + \frac{1}{n_1 - x_1 + .5} + \frac{1}{x_2 + .5} + \frac{1}{n_2 - x_2 + .5} \right)^{\frac{1}{2}}.$$

The 95% confidence interval for ln (OR) is $L \pm 1.96\ S(L)$, so that the initial guess for the endpoints is $\exp(L \pm 1.96\ S(L))$.

The confidence intervals for RR are found by using a method proposed by Katz, Baptista, Azen and Price (1978). They state that $\ln (\tilde{P}_1/\tilde{P}_2)$ is approximately

$$N(\ln (P_1/P_2),\ (1 - \tilde{P}_1)/x_1 + (1 - \tilde{P}_2)/x_2),$$

so the approximate confidence interval endpoints are

$$\exp (\ln (\tilde{P}_1/\tilde{P}_2) \pm 1.96\ ((1 - \tilde{P}_1)/x_1 + (1 - \tilde{P}_2)/x_2)^{\frac{1}{2}}).$$

# 3. SIMULATION

## 3.1 Introduction

A simulation was conducted to compare the empirical Bayes (EB) estimator to the maximum likelihood (ML) estimator. The computations were performed on a Honeywell Level 66 main frame, operating under the Control Program Six (CP-6) operating system. The computer programs are written in Fortran.

## 3.2 Simulation Method

Fifteen different design configurations on the unit square are chosen for the simulation. Each configuration represents the g pairs of underlying binomial probabilities for a set of 2 × 2 contingency tables. Factors considered when choosing a particular design configuration are:

1.) $g$ = number of $(P_1,P_2)$ pairs (i.e. number of tables to be simulated);

2.) location of the $(P_1,P_2)$ points on the unit square;

3.) the correlation between $P_1$ and $P_2$.

The $(P_1,P_2)$ pairs are sufficiently spread out so that if they are treated as sample proportions based on the sample sizes used in the simulation, the hypothesis test for homogeneous odds ratios given in Fleiss (1981) would be rejected at the .05 level. In the simulation, each data set generated was not tested for homogeneous odds ratio. Plots of the fifteen configurations are presented in Figure 3.1. A listing of the $(P_1,P_2)$ pairs for each configuration is presented in Appendix A.

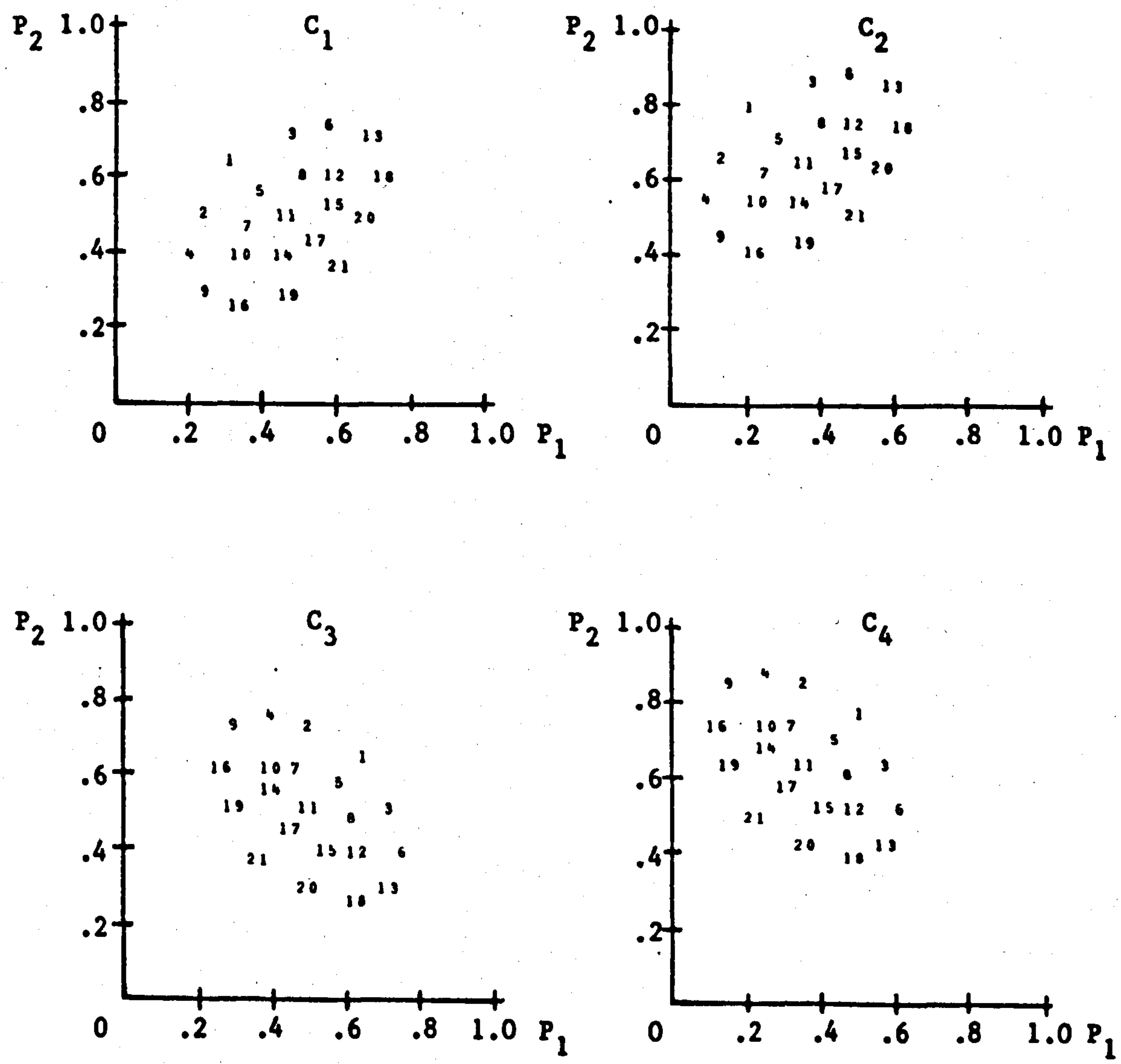



Figure 3.1 Configurations $(C_i = 1,\ldots,15)$ of $(P_1,P_2)$. The points in each configuration are identified by number for later reference.

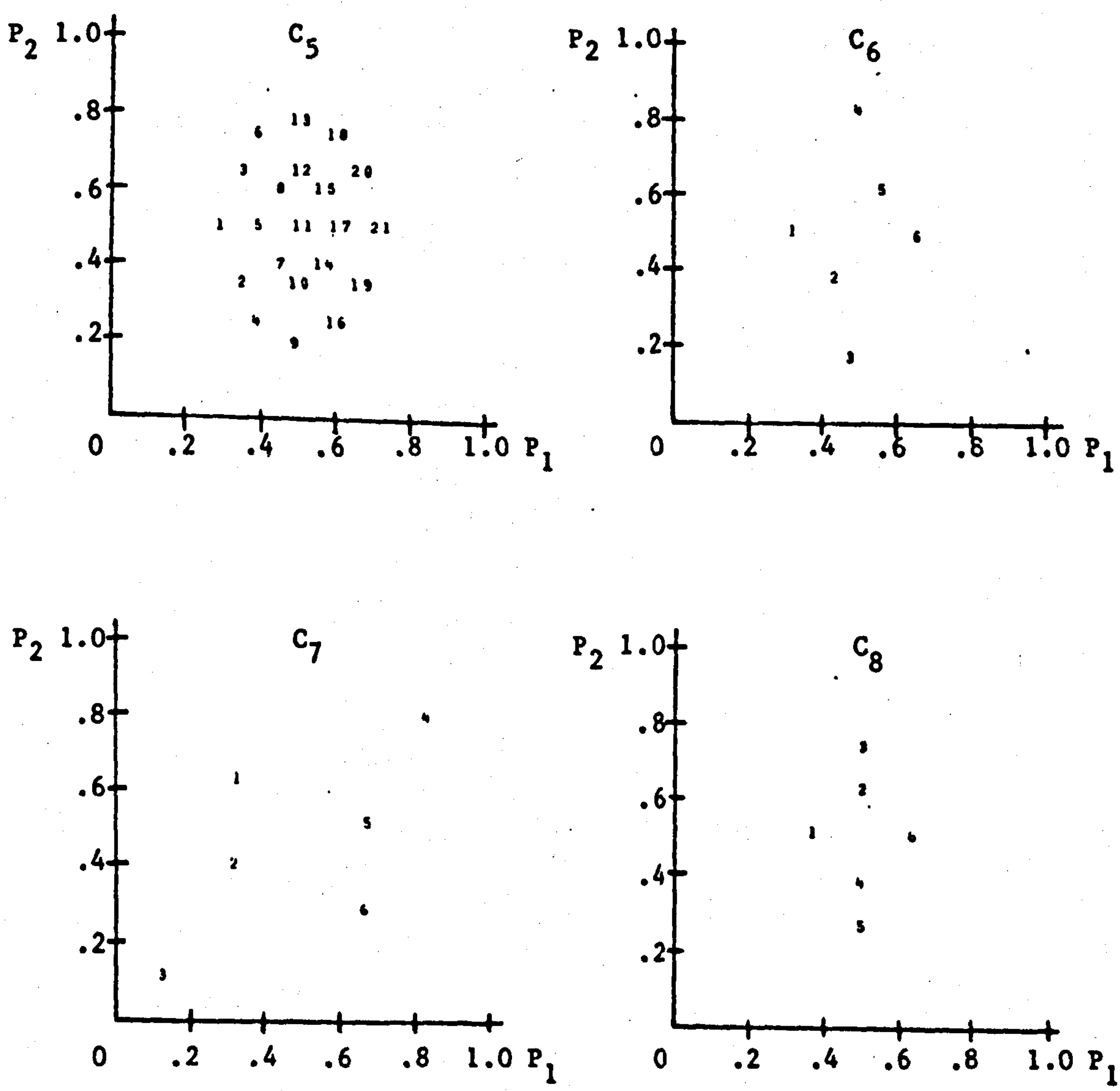



Figure 3.1 (continued)

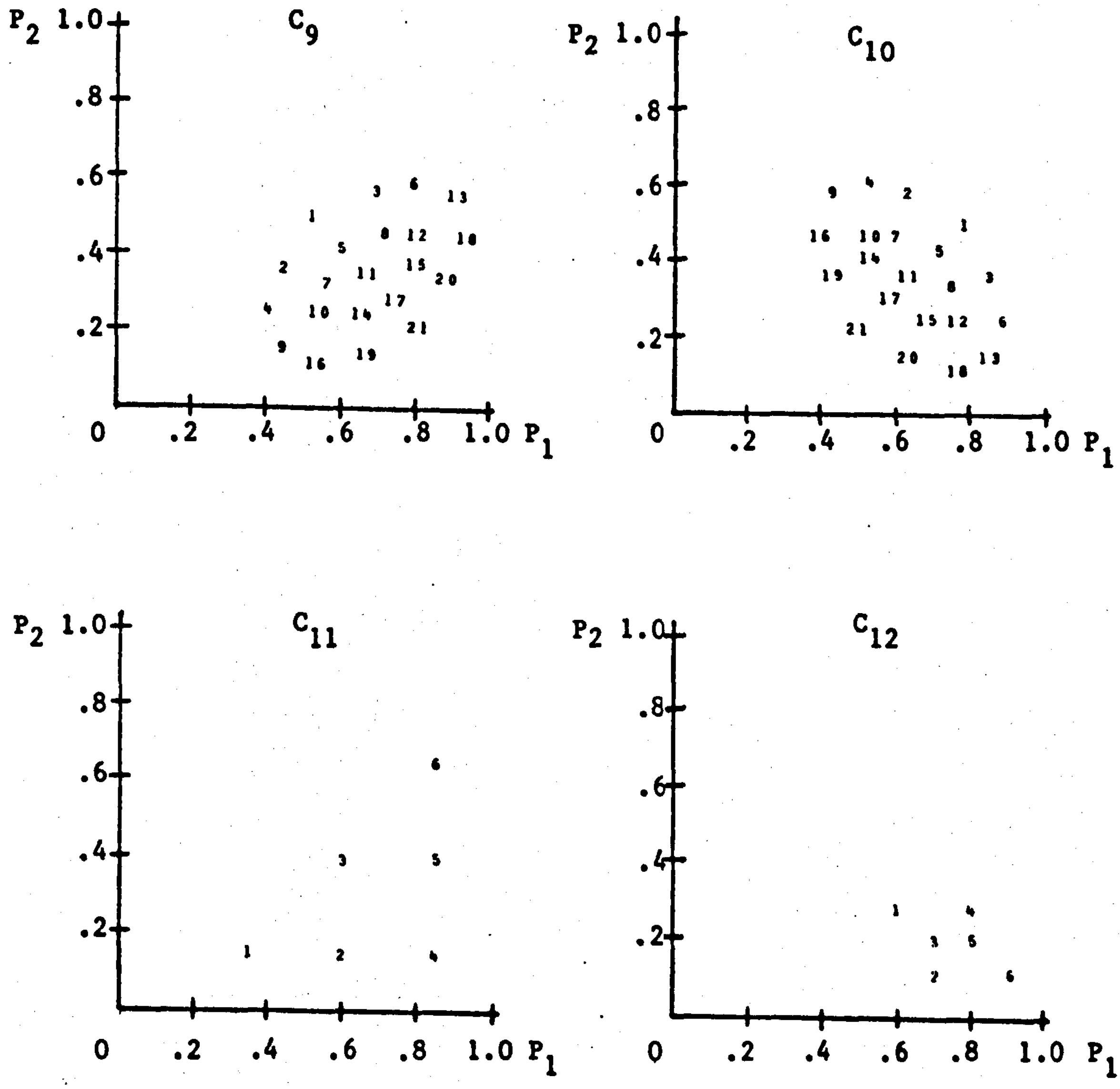



Figure 3.1 (continued)

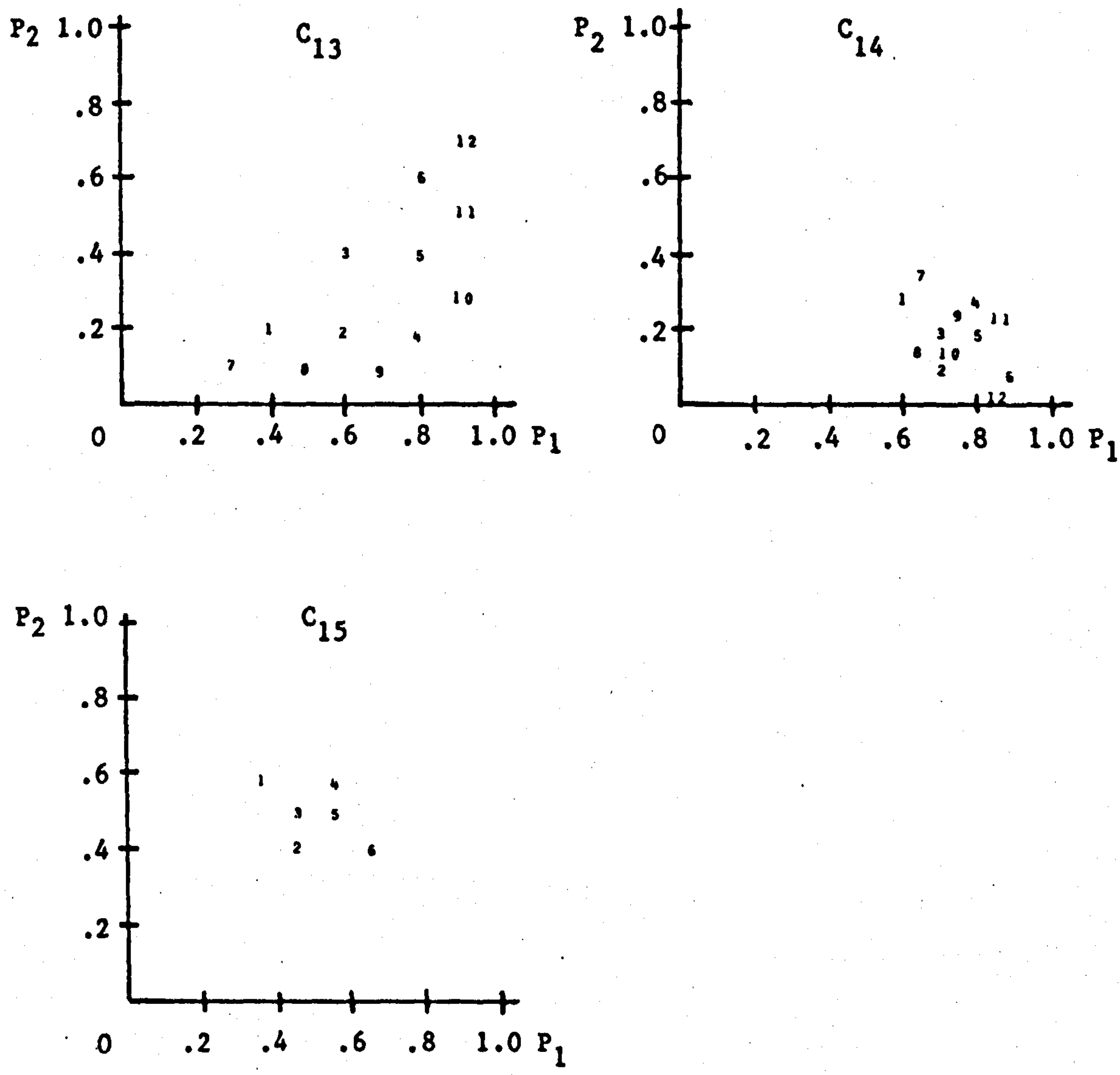



Figure 3.1 (continued)

Let each $(P_1,P_2)$ ordered pair, together with the sample sizes $n_1$ and $n_2$, be called a parameter point. Let the set of parameter points for each configuration be called a parameter set. The twenty-two parameter sets chosen for the simulation are defined in Table 3.1 and Table 3.2.

Table 3.1 PARAMETER SETS USED IN SIMULATION STUDY

Except for parameter sets 13 and 14, all parameter points in a given parameter set have equal sample size. For each parameter point, the sample size associated with $P_1$ is equal to the sample size associated with $P_2$.

| parameter set | configuration (see Fig. 3.1) | sample size |
|---|---|---|
| 1 | $C_1$ | 50 |
| 2 | $C_2$ | 50 |
| 3 | $C_3$ | 50 |
| 4 | $C_4$ | 50 |
| 5 | $C_1$ | 25 |
| 6 | $C_2$ | 25 |
| 7 | $C_3$ | 25 |
| 8 | $C_4$ | 25 |
| 9 | $C_5$ | 50 |
| 10 | $C_6$ | 50 |
| 11 | $C_7$ | 50 |
| 12 | $C_7$ | 25 |
| 13 | $C_1$ | see Table 3.2 |
| 14 | $C_7$ | see Table 3.2 |
| 15 | $C_8$ | 50 |
| 16 | $C_9$ | 50 |
| 17 | $C_{10}$ | 50 |

Table 3.1 (Continued)

| parameter set | configuration (see Fig. 3.1) | sample size |
|---|---|---|
| 18 | $C_{11}$ | 50 |
| 19 | $C_{12}$ | 50 |
| 20 | $C_{13}$ | 50 |
| 21 | $C_{14}$ | 50 |
| 22 | $C_{15}$ | 50 |

Table 3.2 PARAMETER POINTS FOR PARAMETER SETS 13 AND 14
The parameter points have sample sizes such that the sample size associated with $P_2$ is twice the sample size associated with $P_1$. The sample size for $P_1$ is picked at random from a uniform distribution on the integers 15 through 40, inclusive.

| parameter set 13 | | |
|---|---|---|
| | sample size | |
| point | $n_1$ | $n_2$ |
| 1 | 16 | 32 |
| 2 | 32 | 64 |
| 3 | 40 | 80 |
| 4 | 35 | 70 |
| 5 | 32 | 64 |
| 6 | 22 | 44 |
| 7 | 18 | 36 |
| 8 | 25 | 50 |
| 9 | 20 | 40 |
| 10 | 18 | 36 |
| 11 | 25 | 50 |
| 12 | 19 | 38 |

Table 3.2 (Continued)

| | sample size | |
|---|---|---|
| point | $n_1$ | $n_2$ |
| | parameter set 13 | |
| 13 | 16 | 32 |
| 14 | 34 | 68 |
| 15 | 21 | 42 |
| 16 | 34 | 68 |
| 17 | 23 | 46 |
| 18 | 19 | 38 |
| 19 | 19 | 38 |
| 20 | 28 | 56 |
| 21 | 24 | 48 |
| | parameter set 14 | |
| 1 | 37 | 74 |
| 2 | 18 | 36 |
| 3 | 32 | 64 |
| 4 | 21 | 42 |
| 5 | 30 | 60 |
| 6 | 18 | 36 |

Two hundred sets of 2 × 2 contingency tables are generated for each parameter set. The binomial counts are obtained by generating pseudo-random numbers on the interval [0,1] and comparing the number to the true parameter value. The pseudo-random numbers are generated using the linear congruential method (Knuth 1969). A 9-digit odd integer is used as a seed and the recursion

$$x_{n+1} = 65539x_n$$

is used to generate the sequence of numbers with the lower order bits being retained. The number was then scaled to the interval [0,1]. If the parity changed, then the largest number on the computer is added so that the number will be positive.

## 3.3 Analysis Method

### 3.3.1 Introduction

The maximum likelihood estimates of the hyperparameters for each set of tables are found by using the complex method of Box (Box 1965). Appendix B contains a description of the complex method of Box. The computer program is based on an algorithm given in Beveridge and Schecter (1970).

The integer-valued hyperparameter $\alpha_3$ creates two problems. The first is that when $\alpha_3$ becomes larger than 20, the complex method sometimes fails to find the maximum of the likelihood function. The second is that finding the $\alpha_3$ which maximizes the likelihood function consumes too much computer time. In an attempt to alleviate these problems, five different values of $\alpha_3$ are chosen ($\alpha_3 = 1, 5, 10, 15, 20$). For each value of $\alpha_3$, the likelihood function is maximized with respect to the other four hyperparameters for all two hundred sets of tables in each parameter set. It is hoped that if the above problems can not be overcome, perhaps the simulation results can help in eliciting a guess as to what $\alpha_3$ might be for a given pattern of data points.

For each set of tables and for each of the six parameters ($P_1$, $P_2$, $O_1$, $O_2$, OR, and RR) ML and EB estimates are found for each of the g tables in the parameter set.

### 3.3.2 Estimation When a Given Table Contains a Zero

Both the ML and EB estimates of $P_1$ and $P_2$ are well-defined when a zero occurs in a cell of any given table. However, the estimators for $O_1$, $O_2$, OR, and RR can be undefined when a zero occurs. The ML estimators will contain zeroes in the denominator and the EB estimators can have undefined beta integrals. Whenever a zero occurs, .5 is added to each of the four cells in the table prior to estimating $O_1$, $O_2$, OR, and RR. A quadrature rule based on the sinc function (Stenger 1981) is used to evaluate the beta integrals in the special case where .5 is added to each cell. (The quadrature rule is given in Appendix C.)

### 3.3.3 Mean Squared Error (MSE) Comparison

The sum of the g individual table MSE's is the simultaneous loss criterion used for comparison of the ML and EB procedures. For each parameter of interest this total mean squared error (TMSE) is calculated for each estimation procedure and a test for a significant difference between the TMSE's is conducted.

Let

$N$ = total number of runs,

$\theta_j$ = the parameter of interest (a function of $P_{j1}$ and $P_{j2}$) for table j,

$\tilde{\theta}_{jk}$ = the ML estimate of $\theta_j$ on run k,

$\hat{\theta}_{jk}$ = the EB estimate of $\theta_j$ on run k,

where $j = 1,\ldots,g$ and $k = 1,\ldots,N$.

The MSE for table j using the ML and EB estimators are

$$\tilde{MSE}_{j.} = \sum_{k=1}^{N} (\tilde{\theta}_{jk} - \theta_j)^2$$

and
$$\hat{MSE}_{j.} = \sum_{k=1}^{N} (\hat{\theta}_{jk} - \theta_j)^2, \text{ respectively.}$$

The estimate of TMSE for the ML and EB are

$$T\tilde{MSE}_{..} = \sum_{j=1}^{g} \tilde{MSE}_{j.}$$

and
$$T\hat{MSE}_{..} = \sum_{j=1}^{g} \hat{MSE}_{j.}, \text{ respectively.}$$

The test for significant differences in TMSE is constructed by letting

$$Y_{jk} = (\tilde{\theta}_{jk} - \theta_j)^2 - (\hat{\theta}_{jk} - \theta_j)^2.$$

By the central limit theorem

$$\frac{\sqrt{N}\,(\bar{Y}_{j.} - E(Y_{jk}))}{\sigma_{Y_j}} \xrightarrow{L} N(0,1)\,.$$

Since

$$S_{Y_j}^2 = \frac{\sum_{k=1}^{N} (Y_{jk} - \overline{Y}_{j.})^2}{N - 1}$$

is consistent for $\sigma_{Y_j}$, we have by Slutsky's theorem (Rao 1973, p. 122)

$$\frac{\sqrt{N}\,(\overline{Y}_{j.} - E(Y_{jk}))}{S_{Y_j}} \xrightarrow{L} N(0,1).$$

The test statistic

$$Z = \frac{\sqrt{N}\,\overline{Y}_{j.}}{S_{Y_j}}$$

is calculated to test the hypothesis $E(Y_{jk}) = 0$.

The test for a difference in TMSE is constructed by letting

$$Y_{.k} = \sum_{j=1}^{g} \left[(\tilde{\theta}_{jk} - \theta_j)^2 - (\hat{\theta}_{jk} - \theta_j)^2\right]$$

then by the central limit theorem and Slutsky's theorem,

$$\frac{\sqrt{N}\,(\overline{Y}_{..} - E(Y_{.k}))}{S_{Y.}} \xrightarrow{L} N(0,1)$$

where

$$S_{Y.}^2 = \frac{\sum_{k=1}^{N} (Y_{.k} - \overline{Y}_{..})^2}{N - 1}.$$

The test statistic

$$Z = \frac{\sqrt{N}\,\overline{Y}_{..}}{S_Y}$$

is calculated to test the hypothesis $E(Y_{.k}) = 0$.

It is evident that the sample size of 200 used in the simulation study is sufficiently large for Z to be compared to standard normal critical values.

### 3.4 Comparison of Confidence Intervals

The confidence interval based on the EB estimate described in 2.3.6 is compared to the ML based confidence intervals in two ways. The first is an accuracy plot. Accuracy at $\theta$ is defined as the probability that a confidence interval excludes $\theta$ (Ferguson 1967). Accuracy estimates for selected multiples of the true parameter value are found for each confidence interval procedure. The accuracy at the actual parameter value should be about $\alpha$ where $(1 - \alpha)$ is the nominal confidence coefficient. The accuracy curve should increase to one as the absolute difference between $\theta$ and the actual parameter value gets larger. The second is to compare the confidence interval widths. Let

$N$ = number of estimates of $\theta$,

$\theta_{Uij}$ = upper endpoint of the confidence interval for $\theta$ in table i, run j, and $\theta_{Lij}$ = lower endpoint of the confidence interval for $\theta$ in table i, run j.

Then the arithmetic mean confidence interval width (AMW) for table i is

$$AMW_i = \sum_{j=1}^{N} (\theta_{Uij} - \theta_{Lij})/N$$

and the geometric mean ratio (GMR) of upper to lower confidence interval endpoints for table i is

$$GMR_i = \exp\left[\sum_{j=1}^{N} (\ln \theta_{Uij} - \ln \theta_{Lij})/N\right] .$$

The arithmetic mean confidence interval widths as well as the geometric mean of the ratio of upper to lower confidence interval endpoints are calculated. The reason for the ratio of endpoints is that an odds ratio or relative risk with interval endpoints (3,4) or (1/4,1/3), for example, have the save relative meaning but their arithmetic lengths are quite different. Both comparisons are found for each table and a sum across all g tables is found for each parameter set.

Narrower interval widths is a desirable property of any confidence interval procedure. However, the primary goal is to have a procedure which has higher accuracy at points other than the actual parameter value and be at the nominal value at the actual parameter value.

## 3.5 Results

### 3.5.1 MSE Comparison

Table 3.3 is a summary of total mean squared errors for the maximum likelihood (ML) and empirical Bayes (EB) estimation procedures. The entries in the table are TMSE's for each parameter set (PS). The TMSE for the EB procedure is given for priors based on $\alpha_3$ being fixed at 1, 5, 10, 15, and 20. The plus (or minus) signs following the TMSE indicate the direction and magnitude of significance for the Z-test discussed in 3.3.3.

Table 3.3 TMSE COMPARISON

Plus (+) or minus (-) signs indicate a significant reduction or increase, respectively, in TMSE when using the EB procedure. A blank following the TMSE indicates no significant difference. One, two, or three plus (minus) signs indicate that the absolute value of the test statistic is in the upper .01, .001, or .0001 quantile, respectively, of the standard normal distribution.

| | | Estimated Parameter | | | | | |
|---|---|---|---|---|---|---|---|
| PS | $\alpha_3$ | $P_1$ | $P_2$ | $O_1$ | $O_2$ | OR | RR |
| 1 | ML | .097 | .094 | 5.734 | 5.201 | 12.40 | 1.53 |
| | 1 EB | .090+++ | .087+++ | 6.446--- | 5.158 | 16.14--- | 1.70--- |
| | 5 EB | .085+++ | .082+++ | 5.431+++ | 3.767+++ | 11.12+++ | 1.33+++ |
| | 10 EB | .081+++ | .078+++ | 4.392+++ | 2.888+++ | 7.77+++ | 1.09+++ |
| | 15 EB | .079+++ | .077+++ | 3.745+++ | 2.573+++ | 5.98+++ | 0.93+++ |
| | 20 EB | .079+++ | .078+++ | 3.364+++ | 2.509+++ | 5.48+++ | 0.91+++ |
| 2 | ML | .086 | .085 | 1.150 | 195.032 | 0.76 | 0.35 |
| | 1 EB | .081+++ | .080+++ | 1.242--- | 694.551 | 0.92--- | 0.36--- |
| | 5 EB | .076+++ | .077+++ | 1.002+++ | 147.837 | 0.71++ | 0.32+++ |
| | 10 EB | .074+++ | .073+++ | 0.931+++ | 36.100+ | 0.56+++ | 0.28+++ |
| | 15 EB | .073+++ | .071+++ | 0.846+++ | 25.838+ | 0.48+++ | 0.26+++ |
| | 20 EB | .073+++ | .072+++ | 0.800+++ | 22.901+ | 0.45+++ | 0.26+++ |
| 3 | ML | .099 | .100 | 5.787 | 5.676 | 44.08 | 2.48 |
| | 1 EB | .092+++ | .094+++ | 6.414--- | 6.290--- | 48.18--- | 2.37+++ |
| | 5 EB | .087+++ | .088+++ | 5.470+++ | 5.178+++ | 35.97++ | 1.85+++ |
| | 10 EB | .083+++ | .084+++ | 4.465+++ | 4.354+++ | 26.53+++ | 1.54+++ |
| | 15 EB | .082+++ | .082+++ | 3.915+++ | 3.760+++ | 21.23+++ | 1.39+++ |
| | 20 EB | .082+++ | .083+++ | 3.546+++ | 3.505+++ | 18.34+++ | 1.37+++ |
| 4 | ML | .087 | .085 | 1.143 | 202.796 | 1.91 | 0.46 |
| | 1 EB | .082+++ | .077+++ | 1.246--- | 226.196--- | 1.95--- | 0.43+++ |
| | 5 EB | .076+++ | .075+++ | 1.020+++ | 207.628-- | 1.65+++ | 0.39+++ |
| | 10 EB | .074+++ | .073+++ | 0.829+++ | 186.765 | 1.35+++ | 0.37+++ |
| | 15 EB | .075+++ | .072+++ | 0.771+++ | 173.803+++ | 1.21+++ | 0.37+++ |
| | 20 EB | .078+++ | .071+++ | 0.732+++ | 153.788++ | 1.06+++ | 0.36+++ |
| 5 | ML | .194 | .187 | 20.789 | 21.953 | 43.24 | 4.72 |
| | 1 EB | .167+++ | .163+++ | 24.500--- | 21.713 | 64.52--- | 5.53--- |
| | 5 EB | .152+++ | .144+++ | 18.534+++ | 7.447+++ | 31.00+++ | 3.29+++ |
| | 10 EB | .138+++ | .136+++ | 11.878+++ | 4.876+++ | 16.34+++ | 2.32+++ |
| | 15 EB | .135+++ | .134+++ | 9.133+++ | 4.151+++ | 11.73+++ | 1.86+++ |
| | 20 EB | .135+++ | .136+++ | 7.169+++ | 4.137+++ | 10.77+++ | 1.85+++ |

Table 3.3 (Continued)

| PS | $\alpha_3$ | $P_1$ | $P_2$ | $O_1$ | $O_2$ | OR | RR |
|---|---|---|---|---|---|---|---|
| | | Estimated Parameter | | | | | |
| 6 | ML | .178 | .173 | 6.312 | 384.129 | 2.59 | 0.82 |
| 1 | EB | .160+++ | .155+++ | 6.988-- | 1820.567--- | 3.65--- | 0.89--- |
| 5 | EB | .142+++ | .145+++ | 4.877 | 225.302+++ | 2.12+++ | 0.68+++ |
| 10 | EB | .136+++ | .132+++ | 4.011 | 45.532+++ | 1.30+++ | 0.51+++ |
| 15 | EB | .133+++ | .131+++ | 3.055 | 36.427+++ | 1.07+++ | 0.47+++ |
| 20 | EB | .133+++ | .130+++ | 2.414 | 34.160+++ | 0.86+++ | 0.44+++ |
| 7 | ML | .189 | .191 | 16.988 | 15.407 | 144.48 | 8.52 |
| 1 | EB | .164+++ | .165+++ | 20.322--- | 18.367--- | 171.79--- | 8.35 |
| 5 | EB | .149+++ | .147+++ | 15.490+++ | 12.157+++ | 91.13+++ | 3.59+++ |
| 10 | EB | .137+++ | .138+++ | 10.101+++ | 9.238+++ | 55.23+++ | 2.59+++ |
| 15 | EB | .134+++ | .135+++ | 7.975+++ | 6.965+++ | 38.68+++ | 2.25+++ |
| 20 | EB | .135+++ | .137+++ | 6.688+++ | 6.521+++ | 32.82+++ | 2.26+++ |
| 8 | ML | .165 | .174 | 2.801 | 464.809 | 4.94 | 1.05 |
| 1 | EB | .147+++ | .144+++ | 3.206--- | 576.095--- | 4.85 | 0.87+++ |
| 5 | EB | .130+++ | .135+++ | 2.089+++ | 467.024 | 3.33+++ | 0.71+++ |
| 10 | EB | .126+++ | .131+++ | 1.531+++ | 388.352+++ | 2.40+++ | 0.65+++ |
| 15 | EB | .129+++ | .128+++ | 1.367+++ | 336.005+++ | 1.88+++ | 0.61+++ |
| 20 | EB | .135+++ | .124+++ | 1.275+++ | 276.218+++ | 1.63+++ | 0.60+++ |
| 9 | ML | .100 | .090 | 3.460 | 15.174 | 25.61 | 3.46 |
| 1 | EB | .093+++ | .084+++ | 3.909--- | 12.597 | 32.28--- | 3.79--- |
| 5 | EB | .086+++ | .080+++ | 3.250+++ | 6.997++ | 19.98+++ | 2.80+++ |
| 10 | EB | .081+++ | .080+++ | 2.616+++ | 4.756++ | 13.26+++ | 2.21+++ |
| 15 | EB | .079+++ | .083+++ | 2.344+++ | 5.045++ | 10.56+++ | 1.88+++ |
| 20 | EB | .078+++ | .085+++ | 2.215+++ | 5.902+++ | 13.55+++ | 2.41+++ |
| 10 | ML | .034 | .026 | 1.064 | 8.790 | 17.61 | 2.87 |
| 1 | EB | .032+++ | .025+++ | 1.205--- | 9.585 | 22.02--- | 3.15--- |
| 5 | EB | .030+++ | .025+++ | 0.989+++ | 5.349++ | 15.45 | 2.53++ |
| 10 | EB | .028+++ | .025 | 0.836+++ | 6.900++ | 13.25++ | 2.37++ |
| 15 | EB | .027+++ | .027 | 0.775+++ | 7.097+ | 13.58+++ | 2.37++ |
| | ML | .034 | .026 | 1.073 | 8.997 | 18.00 | 2.93 |
| 20 | EB | .027+++ | .027 | 0.765+++ | 7.178+ | 14.03++ | 2.51++ |

Table 3.3 (Continued)

| PS | $\alpha_3$ | Estimated Parameter $P_1$ | $P_2$ | $O_1$ | $O_2$ | OR | RR |
|---|---|---|---|---|---|---|---|
| 11 | ML | .025 | .026 | 9.750 | 5.283 | 6.41 | 0.83 |
| 1 | EB | .023+++ | .025+++ | 10.543--- | 5.683 | 8.26--- | 0.97--- |
| 5 | EB | .023++ | .025 | 9.259+++ | 3.862+++ | 5.71 | 0.87 |
| 10 | EB | .024 | .026 | 8.559+++ | 3.386++ | 5.53 | 0.84 |
| 15 | EB | .026 | .027- | 8.785++ | 3.494+ | 6.29 | 0.91 |
| 20 | EB | .027--- | .027-- | 8.951++ | 4.068+ | 6.66 | 1.02-- |
| 12 | ML | .046 | .049 | 50.826 | 38.447 | 32.02 | 2.76 |
| 1 | EB | .042+++ | .045+++ | 58.031--- | 43.688 | 46.85--- | 3.94--- |
| 5 | EB | .043+ | .046+++ | 47.128+++ | 29.749 | 32.37 | 3.38- |
| 10 | EB | .046 | .049 | 40.438+++ | 32.365 | 32.79 | 3.27- |
| 15 | EB | .049- | .051 | 37.794+ | 29.787 | 33.30 | 3.24 |
| 20 | EB | .050-- | .051 | 42.228++ | 36.422 | 33.99 | 4.03- |
| 13 | ML | .195 | .105 | 38.356 | 6.658 | 36.95 | 2.81 |
| 1 | EB | .166+++ | .096+++ | 46.792--- | 6.340+ | 53.55--- | 3.04--- |
| 5 | EB | .150+++ | .089+++ | 33.840++ | 4.329+++ | 29.87+++ | 2.09+++ |
| 10 | EB | .138+++ | .084+++ | 19.591+++ | 3.203+++ | 15.69+++ | 1.57+++ |
| 15 | EB | .134+++ | .083+++ | 13.101+++ | 2.753+++ | 10.61+++ | 1.35+++ |
| 20 | EB | .136+++ | .084+++ | 9.174+++ | 2.774+++ | 9.05+++ | 1.35+++ |
| 14 | ML | .050 | .026 | 54.678 | 10.026 | 41.53 | 1.24 |
| 1 | EB | .045+++ | .025+++ | 64.094--- | 10.201 | 56.90-- | 1.39--- |
| 5 | EB | .046+++ | .025 | 49.909+++ | 5.580++ | 31.09+ | 1.09 |
| 10 | EB | .049 | .027 | 43.785+++ | 5.538++ | 34.63+ | 1.20 |
| 15 | EB | .053 | .027- | 41.583++ | 5.853++ | 32.40+ | 1.24 |
| 20 | EB | .054- | .027 | 45.221+++ | 6.775++ | 39.20 | 1.40 |
| 15 | ML | .028 | .027 | 0.736 | 2.336 | 3.47 | 0.58 |
| 1 | EB | .026+++ | .026+++ | 0.850--- | 2.391 | 4.58--- | 0.64--- |
| 5 | EB | .024+++ | .025+++ | 0.729 | 1.727+ | 3.25++ | 0.51+++ |
| 10 | EB | .022+++ | .025+++ | 0.603+++ | 1.603+++ | 2.49+++ | 0.44+++ |
| 15 | EB | .021+++ | .025+ | 0.532+++ | 1.548+++ | 2.00+++ | 0.40+++ |
|  | ML | .027 | .026 | 0.704 | 2.296 | 3.27 | 0.56 |
| 20 | EB | .019+++ | .026 | 0.460+++ | 1.560+ | 2.01+++ | 0.42+++ |

Table 3.3 (Continued)

| PS | $\alpha_3$ | | Estimated Parameter | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $P_1$ | $P_2$ | $O_1$ | $O_2$ | OR | RR |
| 16 | | ML | .085 | .084 | 180.326 | 1.116 | 941.15 | 46.89 |
| | 1 | EB | .078+++ | .076+++ | 203.621--- | 1.001+++ | 1261.62--- | 54.59--- |
| | 5 | EB | .073+++ | .071+++ | 168.591+++ | 0.782+++ | 758.95+++ | 40.81+++ |
| | 10 | EB | .071+++ | .068+++ | 122.219++ | 0.670+++ | 440.43+++ | 29.10+++ |
| | 15 | EB | .071+++ | .070+++ | 119.916++ | 0.673+++ | 460.26+++ | 28.36+++ |
| | | ML | .087 | .086 | 206.176 | 1.190 | 1115.41 | 52.28 |
| | 20 | EB | .071+++ | .068+++ | 177.888 | 0.653+++ | 1261.68 | 32.06+++ |
| 17 | | ML | .085 | .086 | 125.245 | 1.279 | 5306.49 | 87.16 |
| | 1 | EB | .078+++ | .082+++ | 143.799--- | 1.401--- | 8874.49- | 200.27 |
| | 5 | EB | .075+++ | .074+++ | 131.005--- | 1.041+++ | 3721.25++ | 27.02+ |
| | 10 | EB | .072+++ | .072+++ | 114.310+++ | 0.863+++ | 2448.90+++ | 14.07+ |
| | 15 | EB | .071+++ | .073+++ | 105.777+++ | 0.717+++ | 1839.14+++ | 13.86+ |
| | 20 | EB | .070+++ | .078+++ | 95.566+++ | 0.752+++ | 1566.69+++ | 15.77+ |
| 18 | | ML | .020 | .023 | 60.923 | 0.528 | 1485.55 | 12.50 |
| | 1 | EB | .019+++ | .022+++ | 69.642--- | 0.468+++ | 1930.51- | 14.70--- |
| | 5 | EB | .020 | .023 | 60.090 | 0.399+++ | 1360.69+ | 12.48 |
| | 10 | EB | .021 | .024 | 53.581 | 0.417++ | 1189.46 | 12.53 |
| | 15 | EB | .020 | .024 | 55.139 | 0.434+ | 1391.44 | 12.80 |
| | | ML | .020 | .023 | 53.228 | 0.551 | 660.07 | 9.98 |
| | 20 | EB | .021 | .024 | 48.459 | 0.457+ | 616.23 | 10.14 |
| 19 | | ML | .021 | .018 | 136.227 | 0.060 | 25138.16 | 124.60 |
| | 1 | EB | .019+++ | .017+++ | 163.887--- | 0.063--- | 71696.87--- | 584.75--- |
| | | ML | .021 | .018 | 136.227 | 0.060 | 81176.30 | 222.06 |
| | 5 | EB | .018+++ | .016+++ | 157.301--- | 0.053+++ | 104687.18 | 290.09 |
| | 10 | EB | .018+++ | .014+++ | 153.689--- | 0.046+++ | 51228.58 | 90.80++ |
| | 15 | EB | .018+++ | .014+++ | 148.770--- | 0.043+++ | 23225.55 | 56.03++ |
| | | ML | .021 | .018 | 136.901 | 0.059 | 80929.83 | 216.49 |
| | 20 | EB | .018+++ | .014+++ | 140.670 | 0.041+++ | 14480.73 | 40.50++ |

Table 3.3 (Continued)

| PS | $\alpha_3$ | | Estimated Parameter | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $P_1$ | $P_2$ | $O_1$ | $O_2$ | OR | RR |
| 20 | | ML | .042 | .042 | 310.106 | 1.570 | 2100.39 | 118.83 |
| | 1 | EB | .039+++ | .040+++ | 348.726--- | 1.275+++ | 2802.17--- | 139.89--- |
| | 5 | EB | .039+++ | .040+++ | 292.898++ | 0.888++ | 1774.66+++ | 116.05 |
| | 10 | EB | .041 | .041 | 265.324++ | 0.963+ | 1785.95+++ | 118.69 |
| | 15 | EB | .041 | .042 | 275.088++ | 1.039+ | 1948.64 | 119.95 |
| | | ML | .042 | .043 | 314.899 | 1.600 | 2042.73 | 120.33 |
| | 20 | EB | .041 | .043 | 282.278+ | 1.067+ | 1925.80 | 120.09 |
| 21 | | ML | .040 | .035 | 194.673 | 0.124 | 48240.29 | 361.39 |
| | 1 | EB | .035+++ | .034+++ | 235.699-- | 0.133--- | 8488672.90 | 212061.18 |
| | | ML | .040 | .035 | 194.673 | 0.124 | 73397.38 | 584.32 |
| | 5 | EB | .034+++ | .030+++ | 222.336-- | 0.110+++ | 66299.56 | 367.76+++ |
| | 10 | EB | .034+++ | .028+++ | 217.043-- | 0.091+++ | 42254.21++ | 166.78+++ |
| | 15 | EB | .034+++ | .027+++ | 214.487- | 0.085+++ | 21063.44+++ | 82.81+++ |
| | 20 | EB | .034+++ | .027+++ | 193.076 | 0.081+++ | 13856.25+++ | 71.40+++ |
| 22 | | ML | .031 | .030 | 0.964 | 0.864 | 3.97 | 0.39 |
| | 1 | EB | .028+++ | .028+++ | 1.095--- | 1.002--- | 4.56--- | 0.39 |
| | 5 | EB | .027+++ | .026+++ | 0.971 | 0.855 | 3.91 | 0.35+++ |
| | 10 | EB | .025+++ | .024+++ | 0.814+++ | 0.756+++ | 3.24+++ | 0.31+++ |
| | 15 | EB | .024+++ | .023+++ | 0.725+++ | 0.658+++ | 2.76+++ | 0.29+++ |
| | | ML | .031 | .030 | 0.944 | 0.862 | 4.06 | 0.40 |
| | 20 | EB | .023+++ | .022+++ | 0.659+++ | 0.597+++ | 2.47+++ | 0.27+++ |

Parameter sets 19 and 21 each have one set of tables out of the 200 runs in which EB estimates are not found for the case $\alpha_3 = 1$. The difficulty is that a zero count for $x_2$ occurs in such a way that the estimate of the hyperparameter $\alpha_2$ is less than one-half. This in turn leads to an undefined beta integral as the estimate of OR (and RR). Several parameter sets have sets of tables when $\alpha_3 = 20$ where the complex method of Box fails to converge. The number of times this occurs for each parameter set is given in Table 3.4

Table 3.4 NUMBER OF TABLE SETS WHERE COMPLEX METHOD OF BOX FAILS TO CONVERGE ($\alpha_3 = 20$)

| parameter set | number of failures |
|---|---|
| 10 | 5 |
| 15 | 16 |
| 16 | 53 |
| 18 | 14 |
| 19 | 1 |
| 20 | 6 |
| 22 | 8 |

The sets in which either of the above problems occurred are eliminated and the TMSE is recalculated for both the ML and EB procedures.

The results of the TMSE comparison generally indicate that for all six functions of $P_1$ and $P_2$ under consideration and values of $\alpha_3$ greater than one, the EB procedure significantly reduces the TMSE.

When estimating OR, the TMSE for ML is never significantly lower than the TMSE for EB based on $\alpha_3 > 1$. When estimating OR with $\alpha_3 > 1$, the TMSE for EB is never significantly higher than for ML. Based on the simulation results the ML procedure usually had significantly lower TMSE than the EB procedure when estimating:

1.) $O_1$, $O_2$, OR, and RR with $\alpha_3 = 1$.

2.) $P_1$ and $P_2$ with a large value of $\alpha_3$ when the data is in the center of the unit square with a positive correlation between $P_1$ and $P_2$ and the number of tables is small.

3.) $O_1$ where the data is in the lower right corner of the unit square with a negative correlation between $P_1$ and $P_2$. The EB procedure becomes poorer as the number of tables decreases from 21 to 6.

4.) $O_2$ where the data is in the upper left corner of the unit square with a negative correlation between $P_1$ and $P_2$. Only the parameter set with $g = 21$ tables was simulated under the above conditions, but the author expects that for $g < 21$, the EB procedure would do relatively worse when estimating $O_1$.

5.) RR where the data is in the center of the unit square with a positive correlation between $P_1$ and $P_2$ and the number of tables is small ($g = 6$, say).

Based on the TMSE comparisons, it is not clear what the best choice of $\alpha_3$ is for a given set of data points. As a general rule, however, EB for $\alpha_3 = 5$ does well when estimating $P_1$ or $P_2$ and EB for $\alpha_3 = 10$ does well when estimating $O_1$, $O_2$, OR, or RR.

### 3.5.2 Check on Simulation

Using the fact that the ML estimators of $P_1$ and $P_2$ are unbiased, a hypothesis test is performed as a check on the simulation study. For each parameter point a test of the ML estimate equalling the actual parameter value for $P_1$ and $P_2$ is performed at the .05 level (two-tailed). 6.018 percent of the 648 hypothesis tests are rejected, which is not significantly different from 5 percent (p-value = .234).

### 3.5.3 Bias of the EB Estimators

Based on the simulation results, Figure 3.2 (a - d) shows typical plots of the direction and magnitude of the bias using the EB procedure. The empirical bias obtained from parameter set 1 is shown for all six estimated parameters. The bias plots of other parameter sets are similar to the results in parameter set 1.

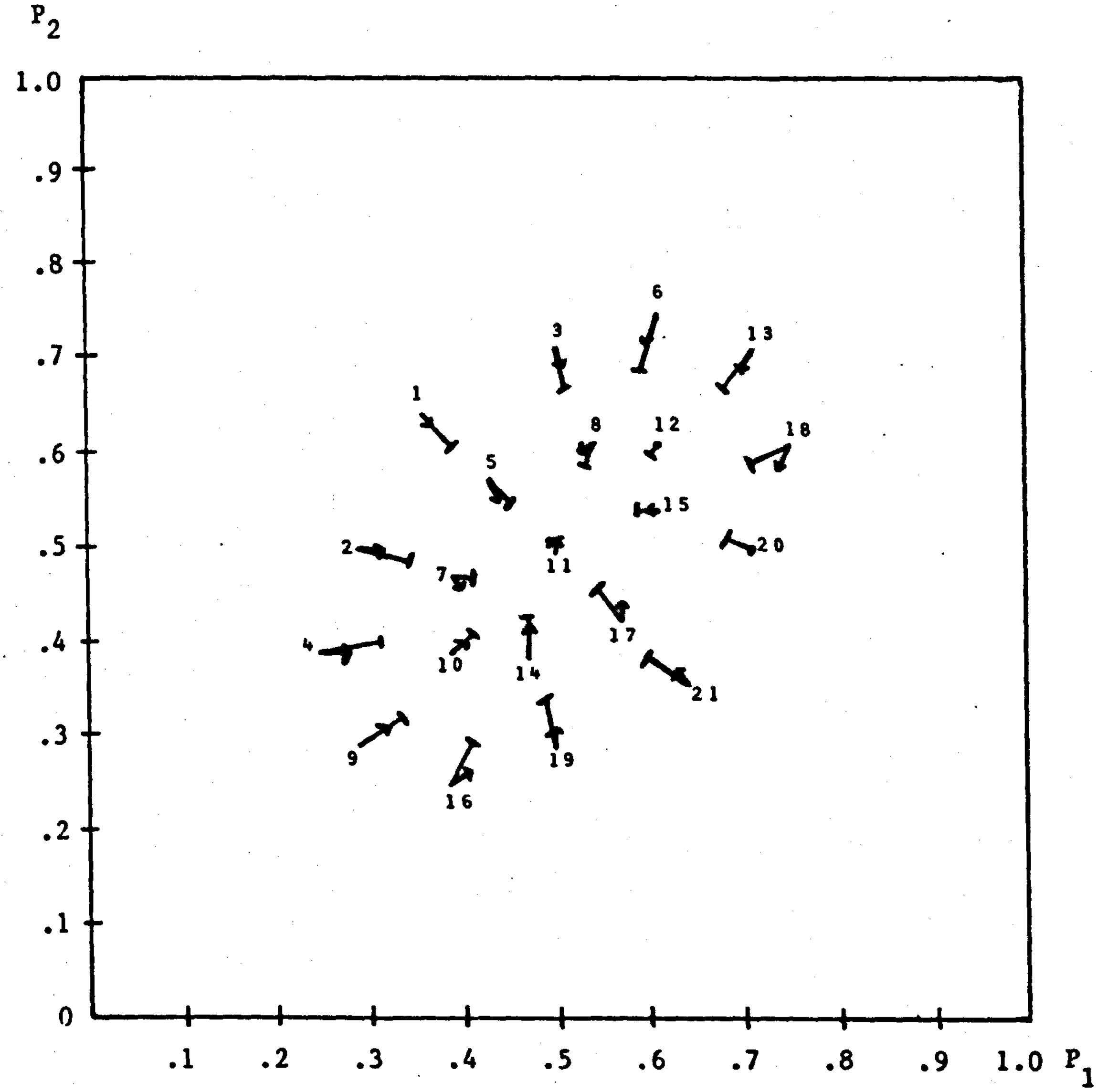



Figure 3.2 (a) Estimated bias of the EB procedure for $P_1$ and $P_2$. The endpoints ^ and - are associated with the priors with $\alpha_3 = 5$ and $\alpha_3 = 20$, respectively. The mean of the EB estimates are the endpoints of vectors emanating from the actual parameter value.

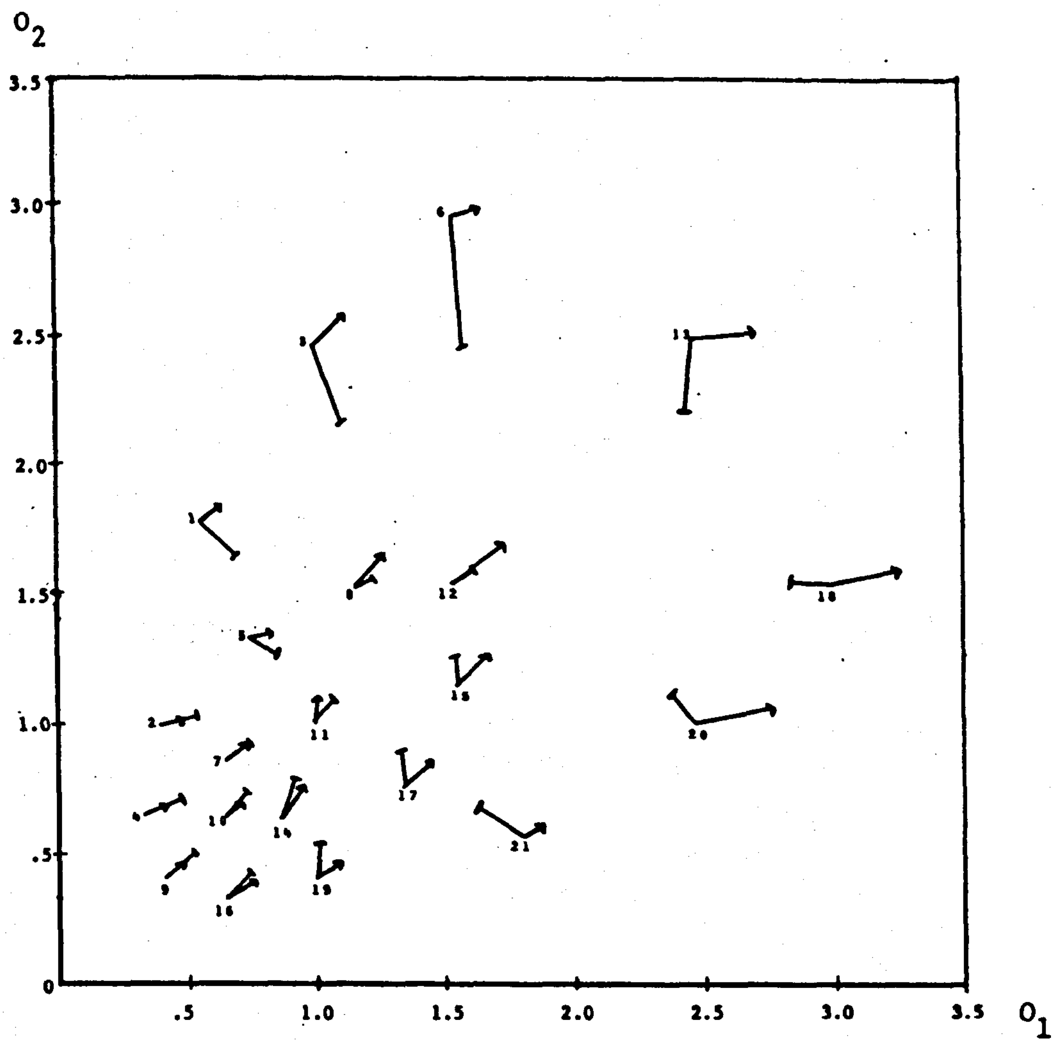



Figure 3.2 (b) Estimated bias of the EB procedure for $O_1$ and $O_2$. The endpoints ^ and - are associated with the priors with $\alpha_3 = 5$ and $\alpha_3 = 20$, respectively. The mean of the EB estimates are the endpoints of vectors emanating from the actual parameter value.

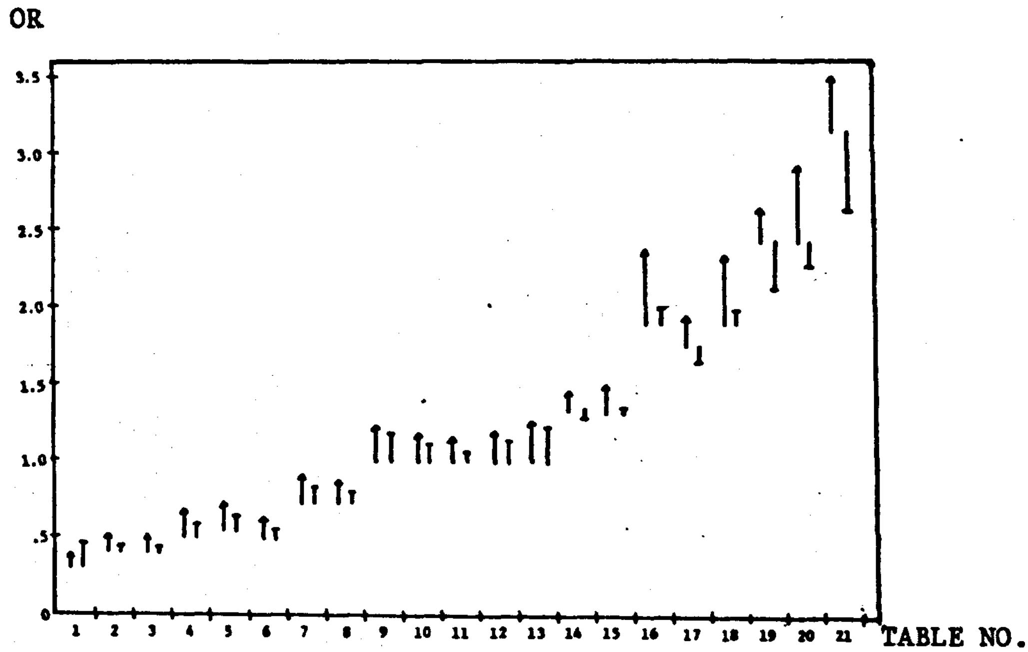



Figure 3.2 (c) Estimated bias of the EB procedure for OR.
The endpoints ^ and - are associated with the priors with $\alpha_3 = 5$ and $\alpha_3 = 20$, respectively. The mean of the EB estimates are the endpoints of vectors emanating from the actual parameter value.

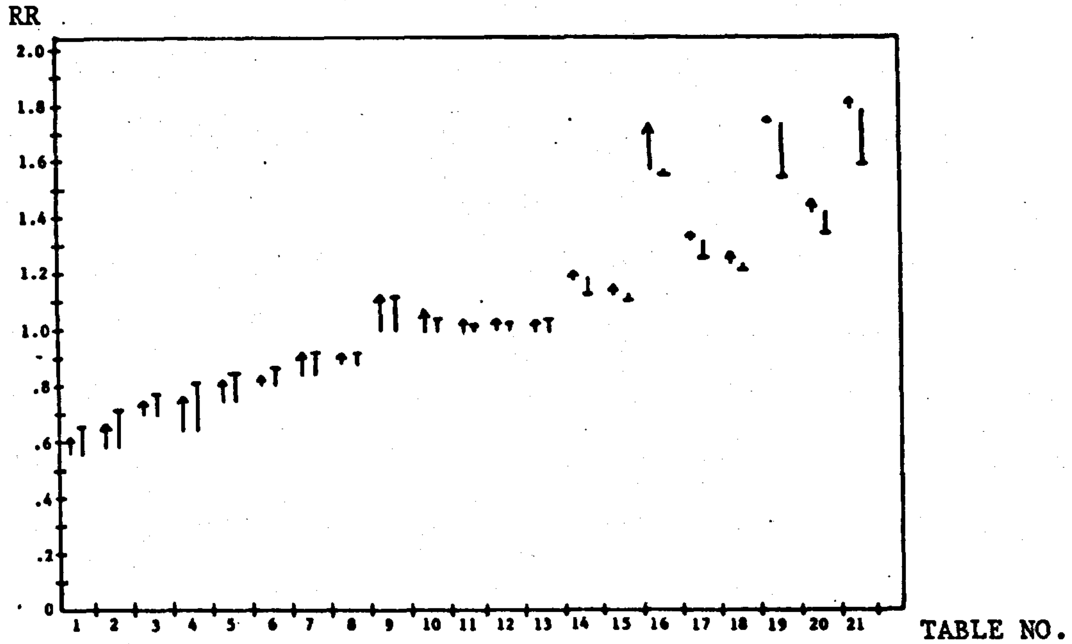



Figure 3.2 (d) Estimated bias of the EB procedure for RR. The endpoints ^ and - are associated with the priors with $\alpha_3 = 5$ and $\alpha_3 = 20$, respectively. The mean of the EB estimates are the endpoints of vectors emanating from the actual parameter value.

Figure 3.2 (a - d) Estimated bias of the EB procedure for θ (parameter set 1).

The EB estimates appear to shrink towards the center of the parameter points. In the plots presented in this section, the shrinkage towards a central value is more apparent when $\alpha_3 = 20$.

Although both values of $\alpha_3$ reduce MSE significantly, they have noticeable differences in the way they affect the bias. The reason for these differences might be that given a particular data set with a certain correlation, the value chosen for $\alpha_3$ may restrict the likelihood function. Thus, for different values of $\alpha_3$, the correlation in the estimated prior may not represent the correlation present in the data because of the restrictions $\alpha_3$ places on the estimated hyperparameters.

Even though the overall TMSE is significantly reduced by using EB instead of ML, some individual tables can have MSE's which are significantly larger for the EB procedure. Table 3.5 displays parameter points from parameter set 1 with $\alpha_3 = 20$. These are the parameter points that produced the most individual table significant differences between MSE's for the ML procedure and the EB procedure. Each of the MSE differences displayed in Table 3.5 are significant at the .0001 level.

Table 3.5 PARAMETER POINTS IN PARAMETER SET 1 AT WHICH THE MAXIMUM DIFFERENCE BETWEEN EMPIRICAL MSE'S (EB VS ML) BASED ON THE Z-TEST IS ATTAINED

Refer to Figure 3.2 where the parameter points are identified by number.

| estimated parameter | MSE (EB) < MSE (ML) | MSE (EB) > MSE (ML) |
|---|---|---|
| $P_1$ | 11 | 4 |
| $P_2$ | 11 | 19 |
| $O_1$ | 6 | 4 |
| $O_2$ | 8 | 19 |
| OR | 10 | 2 |
| RR | 10 | 2 |

It should be noted that the parameter point at which EB is most advantageous has an associated parameter value that is in the center of the 21 parameter values. Also, the parameter point where ML is most advantageous tends to have an associated parameter value that is in the tail of the distribution of the 21 parameter values. For example, if we consider OR in Figure 3.2, we find that parameter point 10 is near the center of the set of 21 OR values while parameter point 2 is near the lower end of the 21 OR values.

### 3.5.4 Confidence Interval Comparison

#### 3.5.4.1 Accuracy

Accuracy results based on the confidence interval procedures discussed in 2.4.4 are presented in this section.

The choice of values at which the accuracy is computed for the different estimated parameters is as follows:

1.) $P_1$, $P_2$ : actual value is exponentiated by degrees of .20, .50, .80, 1.00, 1.25, 2.00, and 5.00;

2.) $O_1$, $O_2$, OR, RR : actual value is multiplied by factors of .1, .2, .5, 1.0, 2.0, 5.0, and 10.0.

Only the results for parameter set 1 are presented since similar results are found for the other parameter sets.

Table 3.6 contains the average accuracy over all 21 tables in parameter set 1 for each estimated parameter.

Table 3.6 AVERAGE ACCURACY (%) OF NOMINAL 95% CONFIDENCE INTERVALS FOR PARAMETER SET 1 ($\alpha_3 = 10$)

Figures are based on 200 simulated sets of tables. $\theta$ is the generic symbol for one of the six parameters.

| estimated parameter | | $\theta^{.2}$ | $\theta^{.5}$ | $\theta^{.8}$ | $\theta$ | $\theta^{1.25}$ | $\theta^{2}$ | $\theta^{5}$ |
|---|---|---|---|---|---|---|---|---|
| $P_1$ | ML | 99.9 | 80.5 | 13.6 | 3.7 | 17.9 | 93.4 | 100.0 |
| | EB | 100.0 | 81.6 | 11.0 | 2.7 | 16.9 | 94.0 | 100.0 |
| $P_2$ | ML | 99.8 | 81.6 | 14.3 | 3.6 | 16.1 | 93.4 | 100.0 |
| | EB | 100.0 | 84.9 | 12.9 | 2.5 | 14.6 | 92.8 | 100.0 |
| | | $.1\theta$ | $.2\theta$ | $.5\theta$ | $\theta$ | $2\theta$ | $5\theta$ | $10\theta$ |
| $O_1$ | ML | 100.0 | 99.6 | 61.0 | 3.7 | 58.9 | 99.2 | 100.0 |
| | EB | 100.0 | 100.0 | 68.7 | 2.5 | 52.6 | 99.7 | 100.0 |
| $O_2$ | ML | 100.0 | 99.7 | 60.5 | 3.6 | 59.2 | 99.5 | 100.0 |
| | EB | 100.0 | 100.0 | 66.2 | 2.5 | 55.1 | 100.0 | 100.0 |
| OR | ML | 99.8 | 95.1 | 30.5 | 3.1 | 29.5 | 94.0 | 99.8 |
| | EB | 100.0 | 98.6 | 37.3 | 1.7 | 20.6 | 93.9 | 100.0 |
| RR | ML | 100.0 | 99.9 | 87.5 | 4.2 | 86.0 | 100.0 | 100.0 |
| | EB | 100.0 | 100.0 | 91.9 | 2.6 | 85.4 | 100.0 | 100.0 |

In general, the average accuracy comparisons are as follows.

1.) The average accuracy of the EB confidence intervals is lower than the ML-based intervals for all estimated parameters at the actual parameter value. Both the EB and ML based intervals apparently have a confidence level which is greater than 95%.

2.) The average accuracy of the EB confidence intervals is higher than the ML-based intervals for $O_1$, $O_2$, OR, and RR at parameter values less than the actual parameter value.

3.) The average accuracy of the EB confidence intervals is lower than the ML-based intervals for all estimated parameters at parameter values greater than the actual parameter value.

The author suspects that the EB intervals may be slightly biased, but did not include a $\theta$ only slightly larger than the true $\theta$ in the simulation experiment. Therefore, a demonstration of bias is not available yet.

Figure 3.3 is accuracy curves for individual tables. The two tables chosen for each estimated parameter are the tables which had the most significant changes in MSE discussed in the section on bias (3.5.2) and listed in Table 3.5. The ln or ln-ln scaling for the $\theta$ values was used so that there would be equal spacing about the actual parameter value.

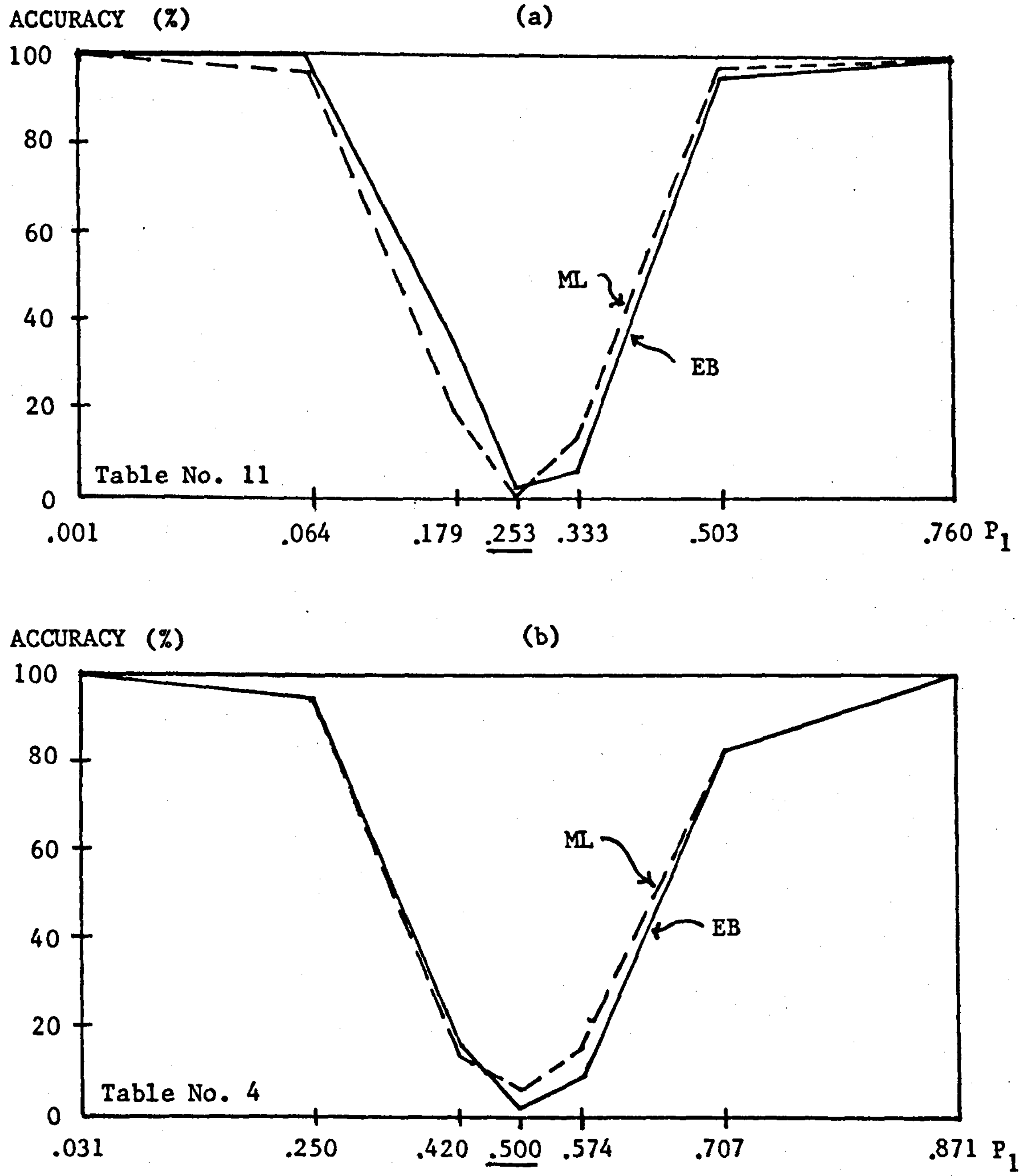
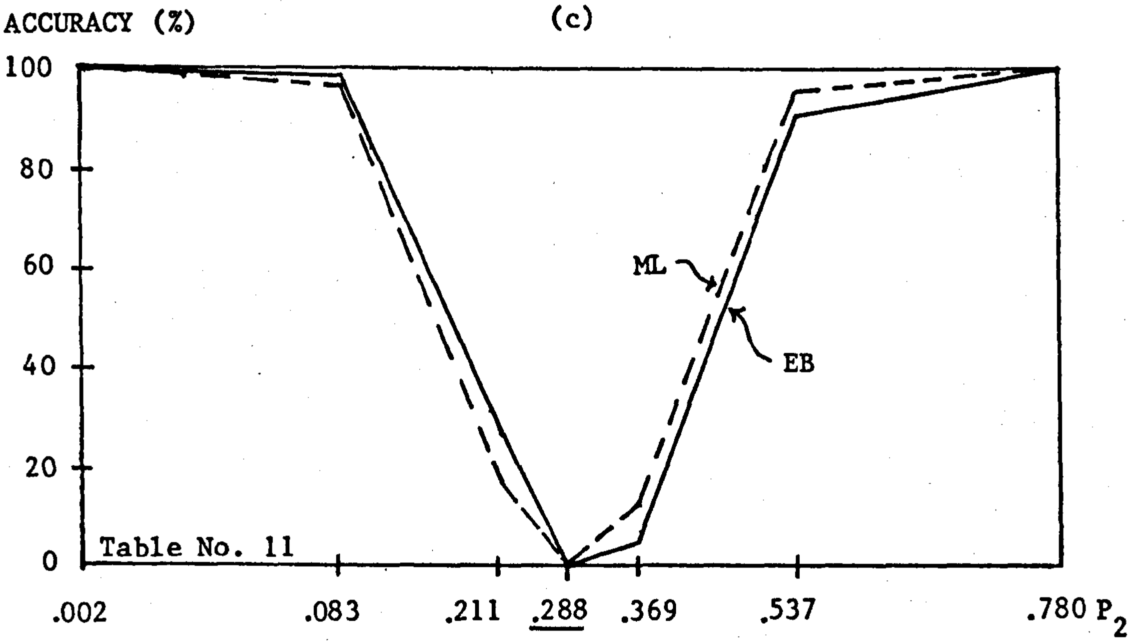



Figure 3.3 (a - 1) Accuracy curves associated with the tables (parameter points) given in Table 3.5. The actual parameter value is underlined. The abscissa for $P_1$, $P_2$ are on a ln-ln scale. The abscissa for $O_1$, $O_2$, OR, and RR are on a ln scale.



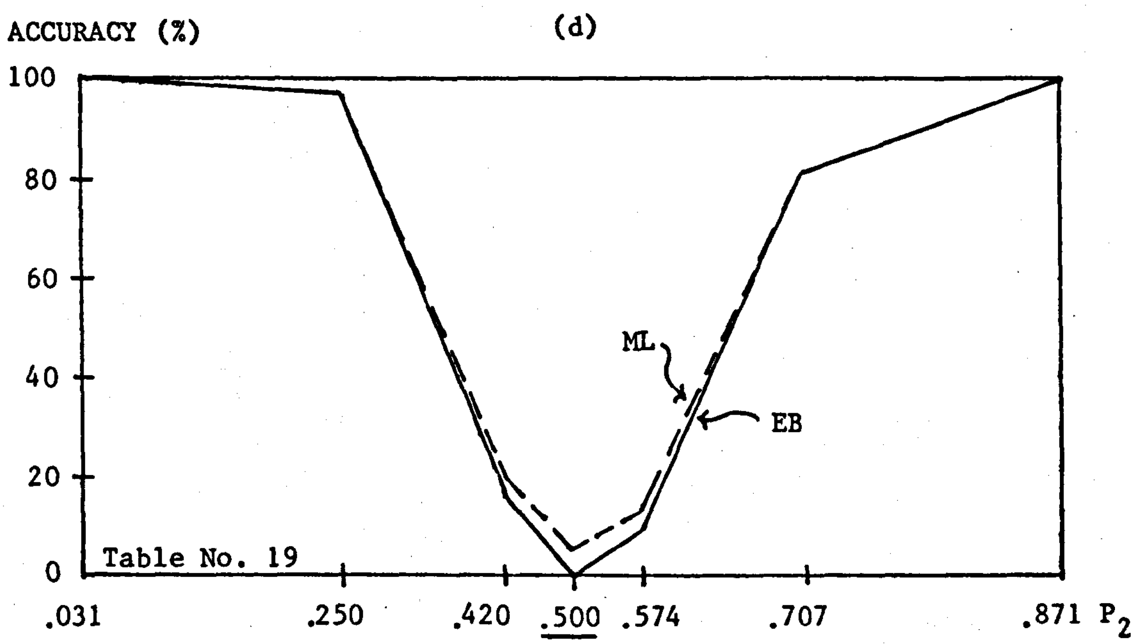
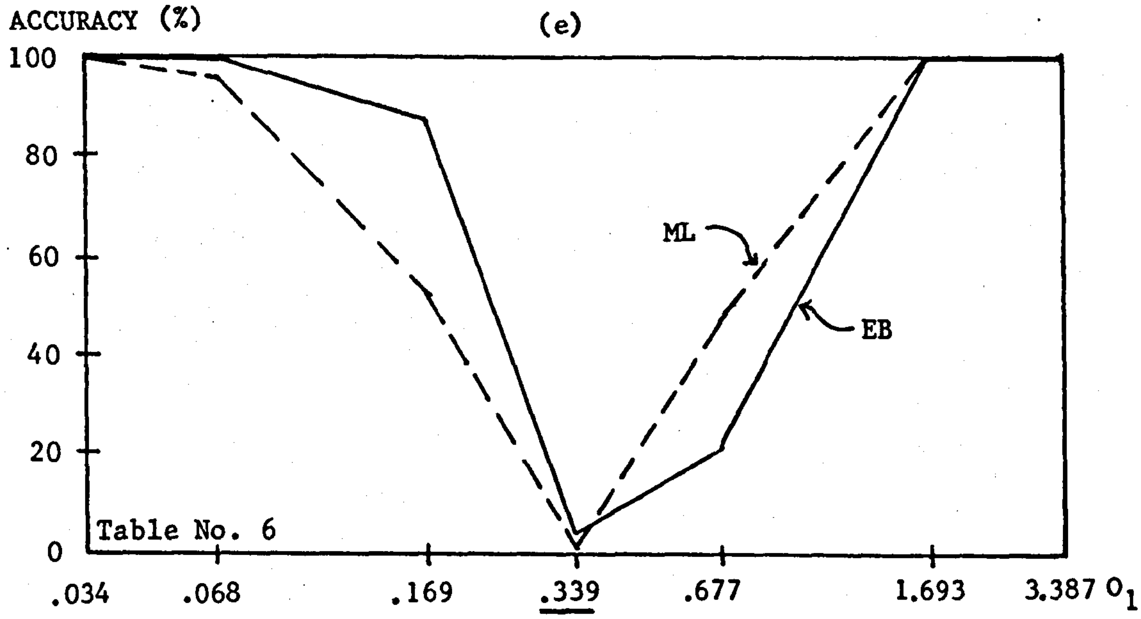




Figure 3.3 (continued)



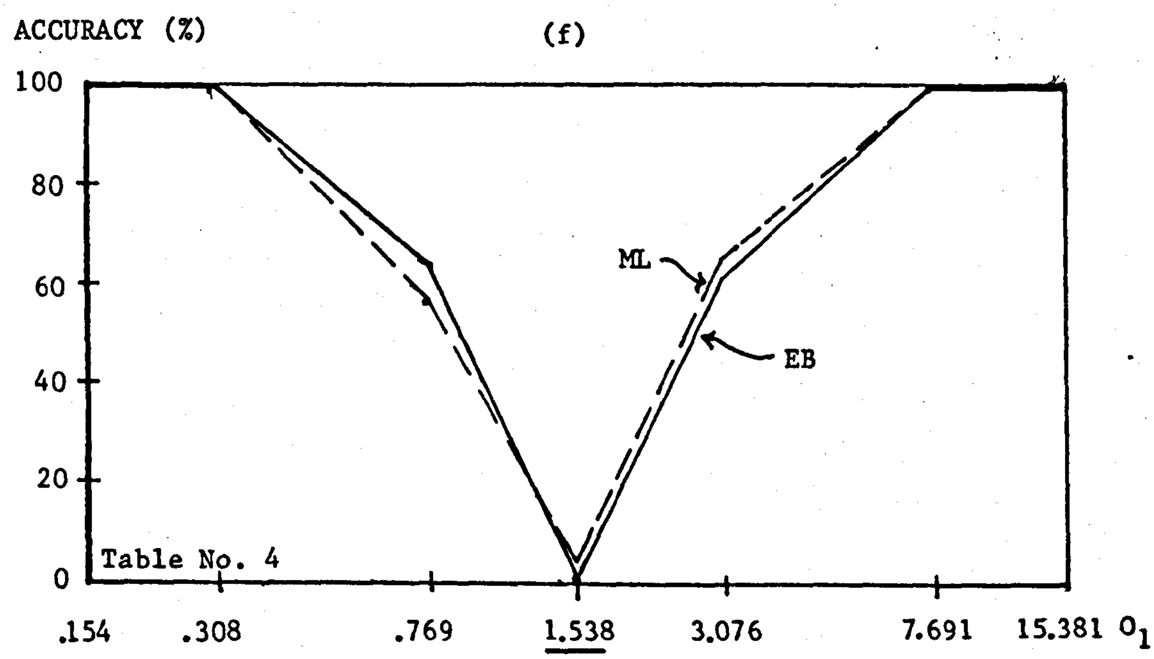
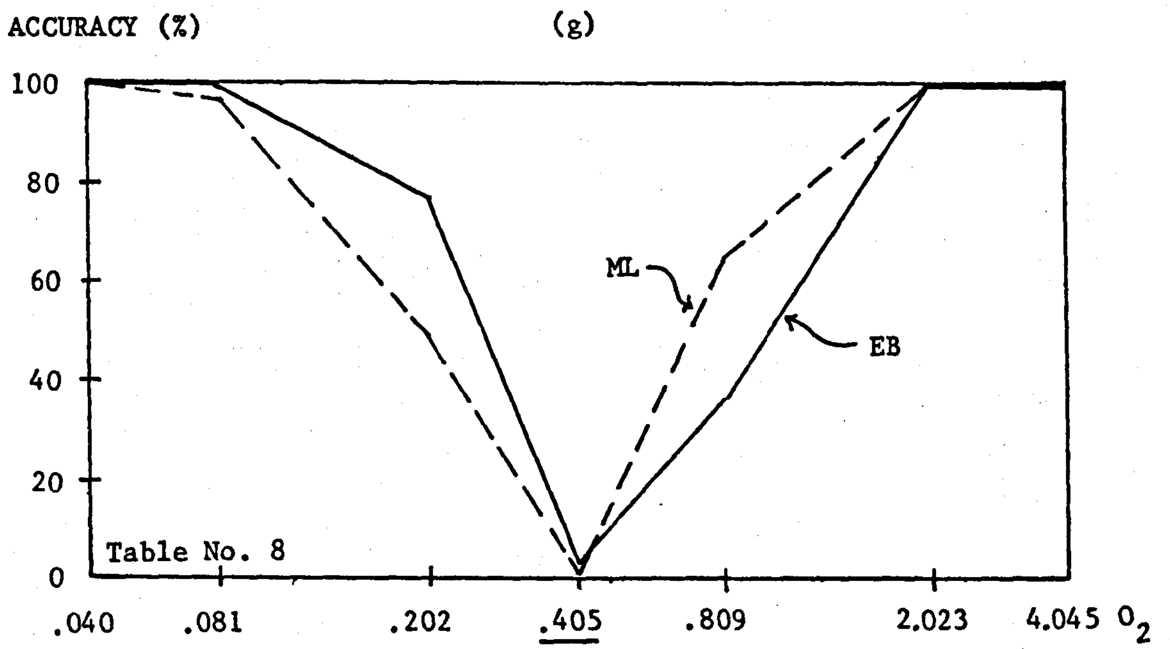




Figure 3.3 (continued)



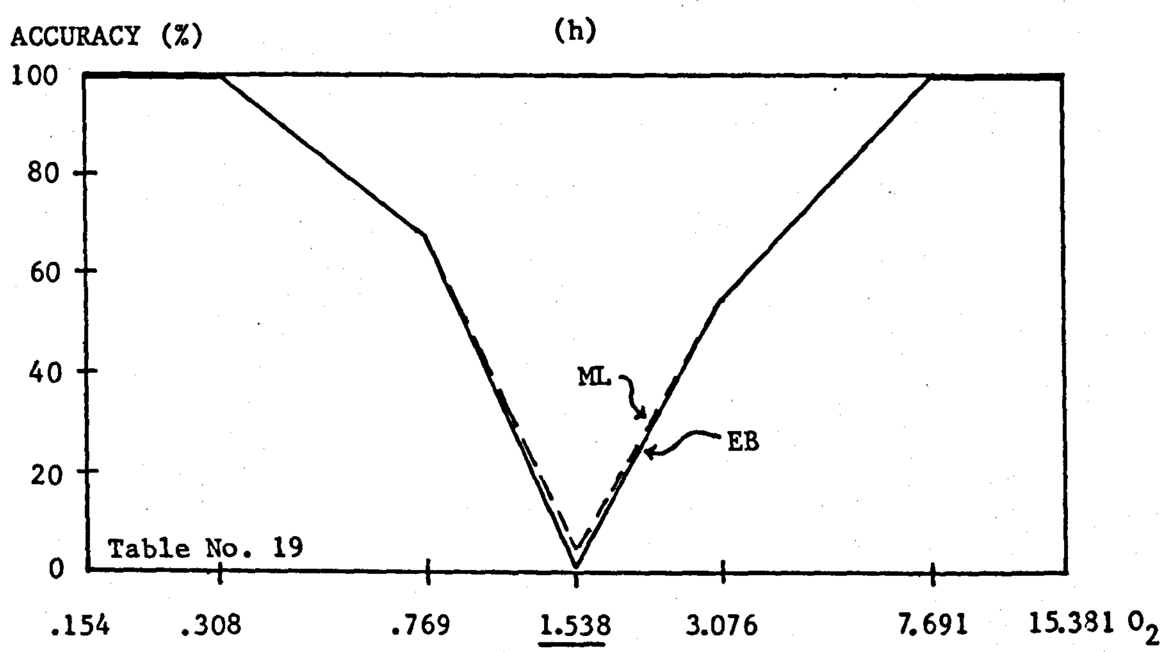
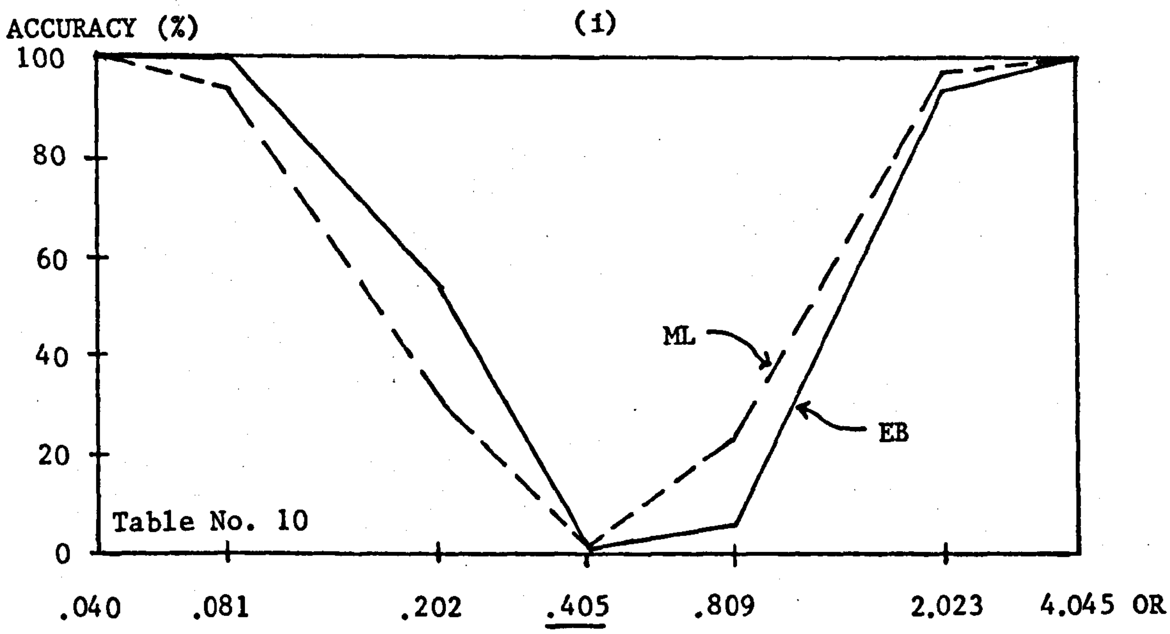




Figure 3.3 (continued)



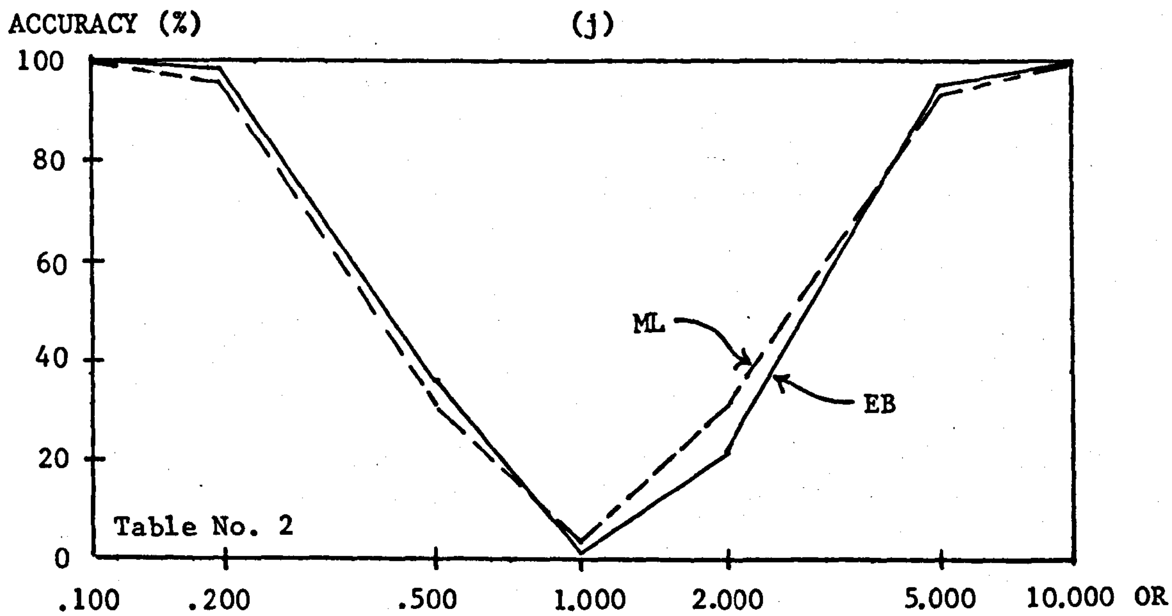
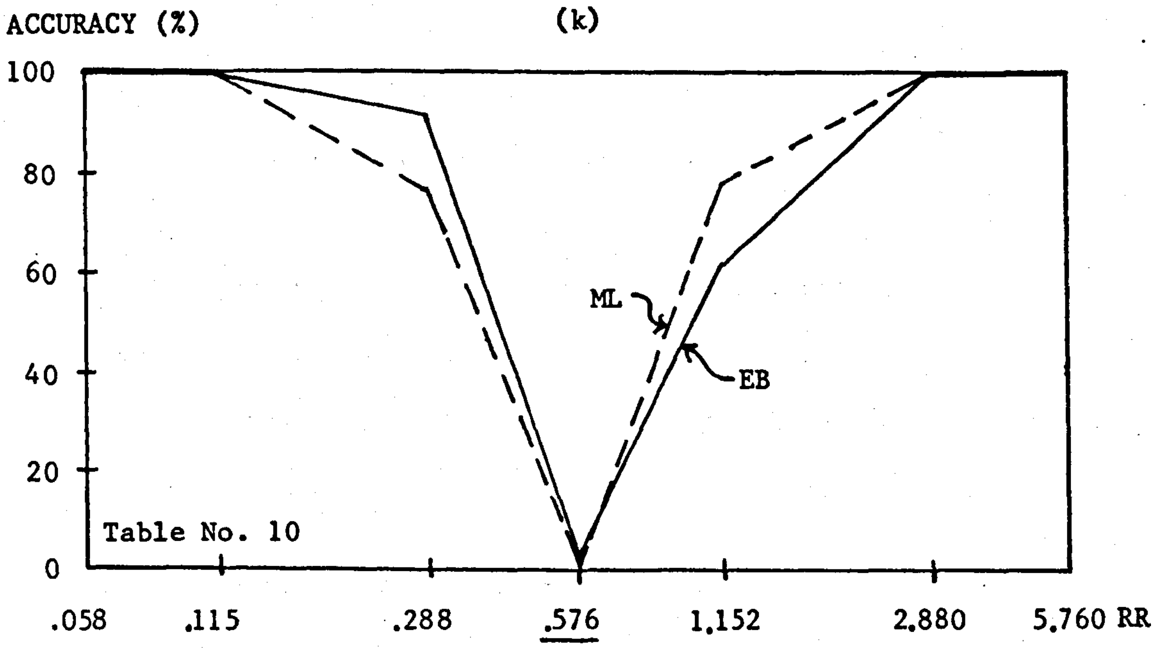




Figure 3.3 (continued)



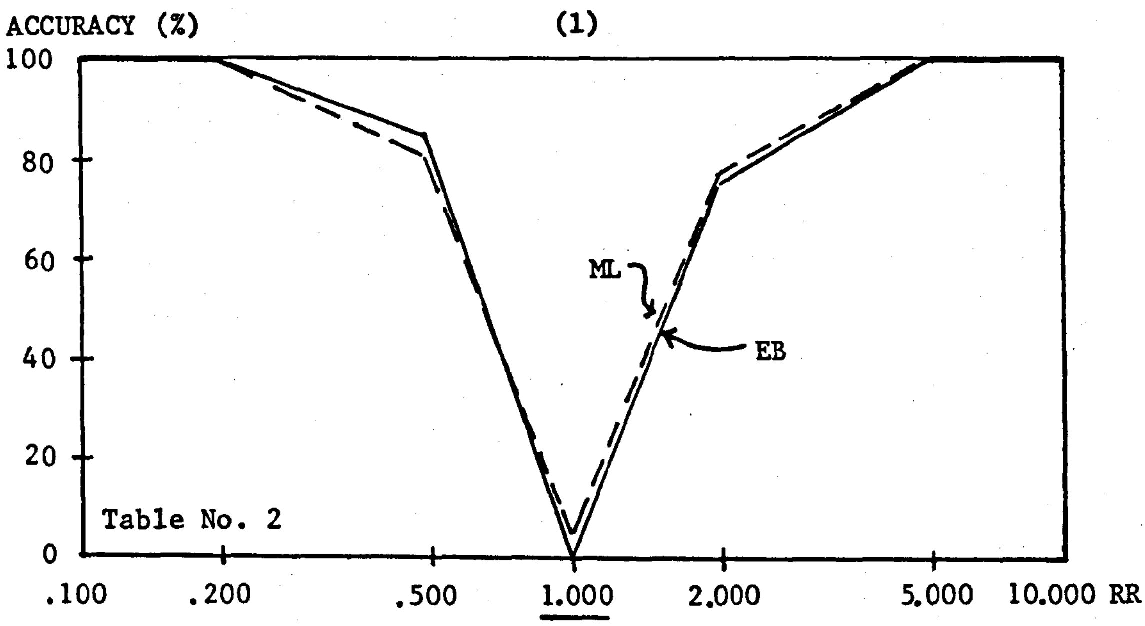




Figure 3.3 (continued)

The accuracy curves for the individual tables associated with a reduction in MSE using the EB procedure are similar to the average accuracy curves. The tables associated with an increase in MSE seem to have lower accuracy at the actual parameter value and have more variability from the ML accuracy curve then the tables associated with a decrease in MSE.

Based on the parameter sets studied, the EB procedure is at the nominal level and is accurate enough for practical purposes.

3.5.4.2 Confidence Interval Width Comparison

The confidence interval width comparison results are given in Table 3.7. For each parameter set, the arithmetic mean width as well as the geometric mean ratio of upper to lower endpoint summed across the g tables is given.

Table 3.7 CONFIDENCE INTERVAL WIDTH COMPARISON ($\alpha_3 = 10$)

An asterisk indicates that a zero is encountered as an endpoint of a confidence interval and therefore the geometric mean is not calculated.

| | Estimated Parameter | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Odds Ratio | | | | Relative Risk | | | |
| | Arithmetic Mean Width | | Geometric Mean Ratio | | Arithmetic Mean Width | | Geometric Mean Ratio | |
| PS | ML | EB | ML | EB | ML | EB | ML | EB |
| 1 | 2.93 | 2.89 | 6.24 | 6.15 | 0.97 | 0.96 | 2.32 | 2.30 |
| 2 | 0.74 | 0.76 | 7.61 | 7.12 | 0.50 | 0.50 | 2.65 | 2.56 |
| 3 | 4.01 | 3.79 | 6.24 | 6.14 | 1.07 | 1.02 | 2.90 | 2.33 |
| 4 | 0.99 | 0.97 | 7.58 | 7.06 | 0.54 | 0.54 | 2.67 | 2.55 |
| 5 | 6.17 | 5.45 | 15.23 | 14.55 | 1.55 | 1.48 | 3.40 | 3.29 |
| 6 | 1.44 | 1.43 | * | 17.46 | 0.76 | 0.76 | 4.15 | 3.75 |
| 7 | 9.46 | 7.54 | 15.32 | 14.46 | 1.81 | 1.55 | 3.48 | 3.32 |
| 8 | 1.84 | 1.71 | * | * | 0.82 | 0.80 | 4.24 | 3.74 |
| 9 | 3.55 | 3.26 | 6.29 | 6.15 | 1.18 | 1.11 | 2.35 | 2.32 |
| 10 | 4.70 | 4.43 | 6.66 | 6.48 | 1.61 | 1.52 | 2.44 | 2.41 |
| 11 | 3.98 | 4.09 | 7.78 | 7.66 | 1.15 | 1.19 | 2.60 | 2.56 |
| 12 | 12.18 | 20.97 | 22.75 | 22.42 | 2.40 | 2.80 | 4.09 | 3.97 |
| 13 | 5.40 | 4.80 | 11.06 | 10.64 | 1.24 | 1.22 | 2.92 | 2.82 |
| 14 | 10.74 | 13.03 | 14.53 | 14.10 | 1.44 | 1.48 | 3.24 | 3.09 |
| 15 | 2.99 | 2.95 | 6.08 | 6.00 | 1.08 | 1.05 | 2.31 | 2.29 |
| 16 | 23.28 | 18.59 | 7.60 | 7.52 | 3.54 | 3.12 | 2.64 | 2.61 |
| 17 | 49.25 | 34.97 | 7.58 | 7.44 | 4.24 | 2.84 | 2.67 | 2.59 |
| 18 | 51.64 | 46.12 | 9.08 | 9.25 | 4.42 | 4.43 | 2.93 | 2.94 |
| 19 | 364.16 | 289.78 | 11.18 | 11.33 | 16.83 | 11.61 | 3.87 | 3.79 |
| 20 | 59.71 | 58.48 | 10.27 | 10.60 | 8.15 | 8.49 | 3.13 | 3.15 |
| 21 | 335.66 | 192.79 | 11.06 | 11.01 | 23.20 | 12.54 | 3.97 | 3.79 |
| 22 | 2.70 | 2.75 | 5.88 | 5.84 | 0.89 | 0.89 | 2.26 | 2.25 |

The EB based confidence intervals generally have both narrower arithmetic mean width and geometric mean ratio. Most cases where the reverse is the case seem to be in parameter sets with a small number of parameter points. For the odds ratio, the parameter sets where the EB based confidence intervals are longer differ for each comparison method. For example, in parameter set 2, the EB based confidence intervals have a longer arithmetic mean width, but shorter geometric mean ratio. This illustrates a difference in the methods of comparing confidence interval widths.

# 4. PROBLEMS AND EXTENSIONS

## 4.1 Problems with the EB Procedure

The results of the simulation study indicate that the EB procedure is a superior method of estimation when compared to the ML procedure. There are, however, some problems and concerns which became evident with the simulation which need to be considered. They are as follows:

1.) the method involves a large number of calculations;

2.) the optimization of the likelihood function using the complex method of Box can fail for large values of $\alpha_3$ ($\geq 20$);

3.) the likelihood function needs to be maximized with respect to $\alpha_3$ as well as the other hyperparameters.

The computer programs and the numerical techniques used in the simulation need to be improved to help eliminate the above problems.

## 4.2 An Independent Prior

This section presents a prior distribution which does not allow for correlation between $P_1$ and $P_2$. Let

$$g(p_1,p_2) = \frac{p_1^{\alpha_1-1}(1-p_1)^{\beta_1-1}\, p_2^{\alpha_2-1}(1-p_2)^{\beta_2-1}}{B(\alpha_1,\beta_1)\, B(\alpha_2,\beta_2)},$$

where $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2 > 0$ and $0 < p_1,p_2 < 1$.

Notice that g is the product of two independent beta distributions.

Using g as a prior, the posterior distribution is

$$P(p_1,p_2|x_1,x_2) = \frac{p_1^{x_1+\alpha_1-1}(1-p_1)^{n_1-x_1+\beta_1-1}\, p_2^{x_2+\alpha_2-1}(1-p_2)^{n_2-x_2+\beta_2-1}}{B(x_1+\alpha_1,n_1-x_1+\beta_1)\, B(x_2+\alpha_2,n_2-x_2+\beta_2)}.$$

Using the same methods as given in Chapter 2, the Bayes estimators based on g are:

$$\mathring{P}_1 = \frac{x_1+\alpha_1}{n_1+\alpha_1+\beta_1},$$

$$\mathring{P}_2 = \frac{x_2+\alpha_2}{n_2+\alpha_2+\beta_2},$$

$$\mathring{O}_1 = \frac{x_1+\alpha_1}{n_1-x_1+\beta_1-1},$$

$$\mathring{O}_2 = \frac{x_2+\alpha_2}{n_2-x_2+\beta_2-1},$$

$$O\mathring{R} = \frac{(x_1+\alpha_1)}{(n_1-x_1+\beta_1-1)} \cdot \frac{(n_2-x_2+\beta_2)}{(x_2+\alpha_2-1)} \quad \text{and}$$

$$R\mathring{R} = \frac{(x_1+\alpha_1)}{(n_1+\alpha_1+\beta_1)} \cdot \frac{(n_2+\alpha_2+\beta_2-1)}{(x_2+\alpha_2-1)}.$$

The empirical Bayes estimators are found by estimating the hyperparameters $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$ using the method of moment (MOM) estimation technique.

The marginal distribution of $x_1$ and $x_2$ is

$$M(x_1,x_2) = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2} B(x_1 + \alpha_1, n_1 - x_1 + \beta_1)\ B(x_2 + \alpha_2, n_2 - x_2 + \beta_2)}{B(\alpha_1,\beta_1)\ B(\alpha_2,\beta_2)}.$$

Since $x_1$ and $x_2$ are independent the MOM estimates of $\alpha_1$ and $\beta_1$ can be found independently of $\alpha_2$ and $\beta_2$ by equating the sample mean $(\bar{x}_1)$ and variance $(s_1^2)$ with $E(x_1)$ and var $(x_1)$, respectively, where

$$E_M(x_1) = \frac{n_1 x_1}{\alpha_1 + \beta_1}, \qquad \mathrm{var}_M(x_1) = \frac{n_1\alpha_1\beta_1(\alpha_1 + \beta_1 + n_1)}{(\alpha_1 + \beta_1 + 1)(\alpha_1 + \beta_1)^2},$$

$$\bar{x}_1 = \sum_{i=1}^{g} \frac{x_{i1}}{g} \qquad , \text{ and } s_1^2 = \sum_{i=1}^{g} \frac{(x_{i1} - \bar{x}_1)^2}{g - 1}$$

and solving for $\alpha_1$ and $\beta_1$. The MOM estimators of $\alpha_1$ and $\beta_1$ are

$$\ddot{\alpha}_1 = \frac{\bar{x}_1(\bar{x}_1(n_1 - \bar{x}_1) - s_1^2)}{(n_1 s_1^2 + \bar{x}_1^2 - \bar{x}_1 n_1)} \qquad \text{and}$$

$$\ddot{\beta}_1 = \frac{(n_2 - \bar{x}_2)(\bar{x}_2(n_2 - \bar{x}_2) - s_1^2)}{(n_2 s_2^2 + \bar{x}_2^2 - \bar{x}_2 n_2)}.$$

The MOM estimates of $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$ are then substituted into the Bayes estimators to obtain the EB estimators.

The simulation study was near completion when g was considered as a possible competitor to $g^-$ and $g^+$. One MSE comparison was performed between the estimates based on g and those based on $g^+$ and found that the estimates based on g were competitive. The author believes that a new simulation is needed with a different set of design configurations that can better evaluate the differences between the two priors.

### 4.3 Posterior Plots and Bayesian Confidence Intervals

A computer program to plot the posterior distributions of $\theta$ was written. Another program to compute the quantiles of the posterior distributions for $P_1$, $P_2$, $O_1$, and $O_2$ was written for use in finding Bayesian confidence intervals, but it is expensive in terms of computer time to use. More efficient programs and/or numerical techniques are needed.

### 4.4 Extensions

#### 4.4.1 A Multivariate Prior

The multivariate prior $g^-$ can be extended to a multivariate prior, $g^*$, in q variables given as

$$g^*(p_1,p_2,\ldots,p_q) = K_0\, p_1^{\alpha_1-1} \cdots p_q^{\alpha_q-1} (1 - \beta_1 p_1 - \cdots - \beta_q p_q)^{\alpha_{q+1}-1}$$

where

$$\alpha_i,\ \beta_j \geq 0, \quad i = 0,\ldots,q+1, \quad j = 0,\ldots,q;$$

and $\sum_{i=1}^{q} \beta_i \leq 1$.

Each pair of $p_i,p_j$ $(i \neq j)$ are negatively correlated, but by replacing $p_i$ by $(1 - p_i)$, positive correlations can be formed. $g^*$ might be used in situations where functions of more than two binomial parameters are to be estimated.

### 4.4.2 An Overall Estimate of θ

The use of Bayesian methods for analysis of random effect models for attribute data has been presented by Dempster, Selwyn, and Weeks (1981), Aitchison and Shen (1980), Haseman and Kupper (1979), and Tarone (1981).

If the underlying values of the risk parameter, θ, are assumed to vary from table to table in a random way, then some of the methods developed in this thesis can be applied to get an estimate of the mean effect. Using $g^+$ (or $g^-$) as the basic model for the variation of the $(P_1,P_2)$ pairs, one can use the data from the individual tables to estimate the parameters of $g^+$ (or $g^-$) as was done in the EB procedure. Using the transformations of Chapter 2 on $g^+$ (or $g^-$), the distribution of θ can be determined. A measure of central tendency for the distribution of θ can then be used as an overall estimate of the risk parameter. A comparison of this approach to the usual Mantel-Haenszel statistic (1959) is needed.

# APPENDICES

## APPENDIX A

### LISTING OF $(P_1,P_2)$ PAIRS FOR EACH CONFIGURATION
### $(C_i,\ i = 1,\ldots,15)$

$C_1$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .359 | .641 |
| 2 | .288 | .500 |
| 3 | .500 | .712 |
| 4 | .253 | .394 |
| 5 | .429 | .571 |
| 6 | .606 | .747 |
| 7 | .394 | .465 |
| 8 | .535 | .606 |
| 9 | .288 | .288 |
| 10 | .394 | .394 |
| 11 | .500 | .500 |
| 12 | .606 | .606 |
| 13 | .712 | .712 |
| 14 | .465 | .394 |
| 15 | .606 | .535 |
| 16 | .394 | .253 |
| 17 | .571 | .429 |
| 18 | .747 | .606 |
| 19 | .500 | .288 |
| 20 | .712 | .500 |
| 21 | .641 | .359 |

$C_2$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .209 | .791 |
| 2 | .138 | .650 |
| 3 | .350 | .862 |
| 4 | .103 | .544 |
| 5 | .279 | .721 |
| 6 | .456 | .897 |
| 7 | .244 | .615 |
| 8 | .385 | .756 |
| 9 | .138 | .438 |
| 10 | .244 | .544 |
| 11 | .350 | .650 |
| 12 | .456 | .756 |
| 13 | .562 | .862 |
| 14 | .315 | .544 |
| 15 | .456 | .685 |
| 16 | .244 | .403 |
| 17 | .421 | .579 |
| 18 | .597 | .756 |
| 19 | .350 | .438 |
| 20 | .562 | .650 |
| 21 | .491 | .509 |

$C_3$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .641 | .641 |
| 2 | .500 | .712 |
| 3 | .712 | .500 |
| 4 | .394 | .747 |
| 5 | .571 | .571 |
| 6 | .747 | .394 |
| 7 | .465 | .606 |
| 8 | .606 | .465 |
| 9 | .288 | .712 |
| 10 | .394 | .606 |
| 11 | .500 | .500 |
| 12 | .606 | .394 |
| 13 | .712 | .288 |
| 14 | .394 | .535 |
| 15 | .535 | .394 |
| 16 | .253 | .606 |
| 17 | .429 | .429 |
| 18 | .606 | .253 |
| 19 | .288 | .500 |
| 20 | .500 | .288 |
| 21 | .359 | .359 |

$C_4$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .491 | .791 |
| 2 | .350 | .862 |
| 3 | .562 | .650 |
| 4 | .244 | .897 |
| 5 | .421 | .721 |
| 6 | .597 | .544 |
| 7 | .315 | .756 |
| 8 | .456 | .615 |
| 9 | .138 | .862 |
| 10 | .244 | .756 |
| 11 | .350 | .650 |
| 12 | .456 | .544 |
| 13 | .562 | .438 |
| 14 | .244 | .685 |
| 15 | .385 | .544 |
| 16 | .103 | .756 |
| 17 | .279 | .579 |
| 18 | .456 | .403 |
| 19 | .138 | .650 |
| 20 | .350 | .438 |
| 21 | .209 | .509 |

$C_5$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .300 | .500 |
| 2 | .350 | .350 |
| 3 | .350 | .650 |
| 4 | .400 | .250 |
| 5 | .400 | .500 |
| 6 | .400 | .750 |
| 7 | .450 | .400 |
| 8 | .450 | .600 |
| 9 | .500 | .200 |
| 10 | .500 | .350 |
| 11 | .500 | .500 |
| 12 | .500 | .650 |
| 13 | .500 | .800 |
| 14 | .550 | .400 |
| 15 | .550 | .600 |
| 16 | .600 | .250 |
| 17 | .600 | .500 |
| 18 | .600 | .750 |
| 19 | .650 | .350 |
| 20 | .650 | .650 |
| 21 | .700 | .500 |

$C_6$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .335 | .500 |
| 2 | .445 | .390 |
| 3 | .500 | .170 |
| 4 | .500 | .830 |
| 5 | .555 | .610 |
| 6 | .665 | .500 |

$C_7$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .330 | .670 |
| 2 | .330 | .443 |
| 3 | .161 | .161 |
| 4 | .839 | .839 |
| 5 | .670 | .557 |
| 6 | .670 | .330 |

$C_8$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .380 | .500 |
| 2 | .500 | .620 |
| 3 | .500 | .740 |
| 4 | .500 | .380 |
| 5 | .500 | .260 |
| 6 | .620 | .500 |

$C_9$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .509 | .491 |
| 2 | .438 | .350 |
| 3 | .650 | .562 |
| 4 | .403 | .244 |
| 5 | .579 | .421 |
| 6 | .756 | .597 |
| 7 | .544 | .315 |
| 8 | .685 | .456 |
| 9 | .438 | .138 |
| 10 | .544 | .244 |
| 11 | .650 | .350 |
| 12 | .756 | .456 |
| 13 | .862 | .562 |
| 14 | .615 | .244 |
| 15 | .756 | .385 |
| 16 | .544 | .103 |
| 17 | .721 | .279 |
| 18 | .897 | .456 |
| 19 | .650 | .138 |
| 20 | .862 | .350 |
| 21 | .791 | .209 |

$C_{10}$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .791 | .491 |
| 2 | .650 | .562 |
| 3 | .862 | .350 |
| 4 | .544 | .597 |
| 5 | .721 | .421 |
| 6 | .897 | .244 |
| 7 | .615 | .456 |
| 8 | .756 | .315 |
| 9 | .438 | .562 |
| 10 | .544 | .456 |
| 11 | .650 | .350 |
| 12 | .756 | .244 |
| 13 | .862 | .138 |
| 14 | .544 | .385 |
| 15 | .685 | .244 |
| 16 | .403 | .456 |
| 17 | .579 | .279 |
| 18 | .756 | .103 |
| 19 | .438 | .350 |
| 20 | .650 | .138 |
| 21 | .509 | .209 |

$C_{11}$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .350 | .150 |
| 2 | .600 | .150 |
| 3 | .600 | .400 |
| 4 | .850 | .150 |
| 5 | .850 | .400 |
| 6 | .850 | .650 |

$C_{12}$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .600 | .300 |
| 2 | .700 | .100 |
| 3 | .700 | .200 |
| 4 | .800 | .300 |
| 5 | .800 | .200 |
| 6 | .900 | .100 |

$C_{13}$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .400 | .200 |
| 2 | .600 | .200 |
| 3 | .600 | .400 |
| 4 | .800 | .200 |
| 5 | .800 | .400 |
| 6 | .800 | .600 |
| 7 | .300 | .100 |
| 8 | .500 | .100 |
| 9 | .700 | .100 |
| 10 | .900 | .300 |
| 11 | .900 | .500 |
| 12 | .900 | .700 |

$C_{14}$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .600 | .300 |
| 2 | .700 | .100 |
| 3 | .700 | .200 |
| 4 | .800 | .300 |
| 5 | .800 | .200 |
| 6 | .900 | .100 |
| 7 | .650 | .350 |
| 8 | .650 | .150 |
| 9 | .750 | .250 |
| 10 | .750 | .150 |
| 11 | .850 | .250 |
| 12 | .850 | .050 |

$C_{15}$

| Table No. | $P_1$ | $P_2$ |
|---|---|---|
| 1 | .350 | .600 |
| 2 | .450 | .400 |
| 3 | .450 | .500 |
| 4 | .550 | .600 |
| 5 | .550 | .500 |
| 6 | .650 | .400 |

## APPENDIX B

## THE COMPLEX METHOD OF BOX

The complex method of Box is a sequential search technique which has proven effective in solving problems with nonlinear objective functions subject to nonlinear inequality constraints. No derivatives are required. The procedure should tend to find the global maximum due to the fact that the initial set of points are randomly scattered throughout the feasible region.

The method finds the maximum of a multivariable, nonlinear function subject to nonlinear inequality constraints:

Maximize $\quad F(X_1, X_2, \ldots, X_N)$

Subject to $\quad G_k \leq X_k \leq H_k, \qquad k = 1, 2, \ldots, M.$

The implicit variables $X_{N+1}, \ldots, X_M$ are dependent functions of the explicit independent variables $X_1, X_2, \ldots, X_N$. The upper and lower constraints $H_k$ and $G_k$ are either constants or functions of the independent variables.

The algorithm proceeds as follows:

1.) An original "complex" of $K \geq N + 1$ points is generated consisting of a feasible starting point and $K - 1$ additional points generated from random numbers and constraints for each of the independent variables:

$$X_{i,j} = G_i + r_{i,j}\,(H_i - G_i),$$
$$i = 1, 2, \ldots, N \text{ and } j = 1, 2, \ldots, K\text{-}1$$

where $r_{i,j}$ are random numbers between 0 and 1.

2.) The selected points must satisfy both the explicit and implicit constraints. If at any time the explicit constraints are violated, the point is moved a small distance $\delta$ inside the violated limit. If an implicit constraint is violated, the point is moved one half of the distance to the centroid of the remaining points

$$X_{i,j}(\text{new}) = (X_{i,j}(\text{old}) + \overline{X}_{i,c})/2$$

$$i = 1, 2, \ldots, N$$

where the coordinates of the centroid of the remaining points, $\overline{X}_{i,c}$, are defined by

$$\overline{X}_{i,c} = \frac{1}{K-1} \sum_{j=1}^{K} X_{i,j} - X_{i,j}(\text{old}) \quad , \quad i = 1, 2, \ldots, N.$$

This process is repeated as necessary until all the implicit constraints are satisfied.

3.) The objective function is evaluated at each point. The point having the lowest function value is replaced by a point which is located at a distance $\alpha$ times as far from the centroid of the remaining points as the distance of the rejected point on the line joining the rejected point and the centroid:

$$X_{i,j}(\text{new}) = \alpha\ (\overline{X}_{i,c} - X_{i,j}(\text{old})) + \overline{X}_{i,c}$$

$$i = 1, 2, \ldots, N.$$

Box (1965) recommends a value of $\alpha = 1.3$.

4.) If a point repeats in giving the lowest function value on consecutive trials, it is moved one half the distance to the centroid of the remaining points.

5.) The new point is checked against the constraints and is adjusted as before if the constraints are violated.

6.) Convergence is assumed when the objective function values at each point are within $\beta$ units for $\gamma$ consecutive iterations. An interation is defined as the calculations required to select a new point which satisfies the constraints and does not repeat in yielding the lowest function value.

A flow sheet illustrating the above procedure is given in Figure B.1.

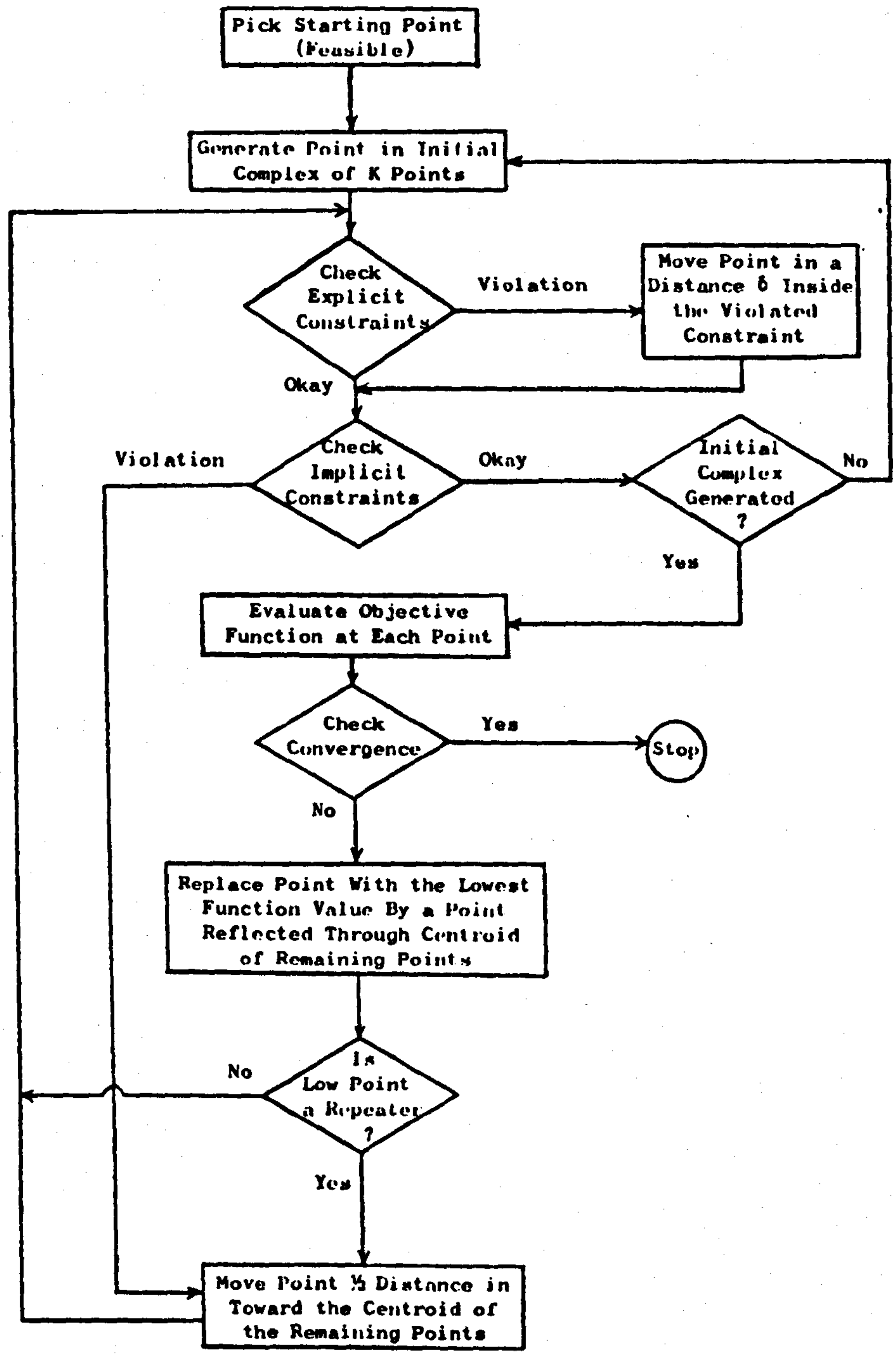



Figure B.1 Logic Diagram for Complex Method of Box

# APPENDIX C

## QUADRATURE RULE BASED ON THE SINC FUNCTION

The quadrature rule based on the sinc function used to evaluate the beta integrals of the thesis is given by

$$\int_0^1 f(x)\,dx \cong h \sum_{k=-M}^{N} a_k\, f(w_k)$$

where

$$a_k = \frac{e^{kh}}{(1 + e^{kh})^2} \qquad \text{and} \qquad w_k = \frac{e^{kh}}{e^{kh} + 1} \quad .$$

The spacing of the $w_k$'s is determined by the choice of h and M,N are positive integers determining the number of points to be evaluated.

# BIBLIOGRAPHY

BIBLIOGRAPHY


Aitchison, J. and Shen, S.M. (1980), "Logistic-Normal Distributions: Some Properties and Uses," Biometrika, 67, no. 2, 261-272.

Beveridge and Schecter (1970), Optimization Theory and Practice, New York: McGraw-Hill, Inc.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975), Discrete Multivariate Analysis: Theory and Practice, Cambridge, Massachusetts: M.I.T. Press.

Box, M.J. (1965), "A New Method of Constrained Optimization and a Comparison with Other Methods," Computer Journal, 8, no. 1, 42-52.

Box, G.E.P. and Tiao, G.C. (1973), Bayesian Inference in Statistical Analysis, Philippines: Addison-Wesley.

DeGroot, Morris H. (1970), Optimal Statistical Decisions, New York: McGraw-Hill, Inc.

Dempster, A.P., Selwyn, M.R., and Weeks, B.J. (1981), "Combining Historical and Randomized Controls for Assessing Trends in Proportions," paper presented at American Statistical Association meetings in Detroit.

Ferguson, Thomas S. (1967), Mathematical Statistics, New York: Academic Press.

Fleiss, J.L. (1979), "Confidence Intervals for the Odds Ratio in Case-Control Studies: The State of the Art," Journal of Chronic Disease, 32, 69-77.

______ (1981), Statistical Methods for Rates and Proportions, John Wiley and Sons, Inc.

Friedman, G.D. (1980), Primer of Epidemiology, New York: McGraw-Hill, Inc.

Gart, J.J. (1962), "On the Combination of Relative Risks," Biometrics, 18, 601-610.

____ (1970), "Point and Interval Estimation of the Common Odds Ratio in the Combination of 2 × 2 Tables with Fixed Marginals," Biometrika, 57, no. 3, 471-475.

Good, I.J. (1980), "Terminology for Various Kinds of Bayesian Methods," JSCS, 11, 309-313.

Graybill, Franklin A. (1976), Theory and Application of the Linear Model, Belmont, California: Wadsworth Publishing Company, Inc.

Haseman, J.K. and Kupper, L.L. (1979), "Analysis of Dichotomous Response Data from Certain Toxilogical Experiments," Biometrics, 35, no. 1, 281-294.

Katz, D., Baptista, J., Azen, S.P., and Pike, M.C. (1978), "Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies," Biometrics, 34, 469-474.

Knuth, D.E. (1969), The Art of Computer Programming - Seminumerical Algorithms, 2, Reading, Massachusetts: Addison-Wesley.

McKinlay, S. (1978), "The Effect of Nonzero Second Order Interaction on Combined Estimators of the Odds Ratio," Biometrika, 65, no. 1, 191-202.

Mantel, N. and Haenszel, W. (1959), "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease," Journal of the National Cancer Institute, 22, 719-748.

Mardia, K.V. (1970), Families of Bivariate Distributions, Griffin's Statistical Monographs and Courses, no. 27, Darien, Connecticut: Hafner Publishing Company.

Ord, J.K. (1972), Families of Frequency Distributions, Griffin's Statistical Monographs and Courses, no. 30, New York: Hafner Publishing Company.

Radhakrishna, S. (1965), "Combination of Results from Several 2 × 2 Contingency Tables," Biometrics, 21, 86.

Rao, C.R. (1973), Linear Statistical Inference and Its Applications, New York: John Wiley and Sons, Inc.

Robbins, H. (1956), "An Empirical Bayes Approach to Statistics," Proceedings of the Third Berkeley Symposium, 1, 157-163.

Snedecor, G.W. and Cochran, W.G. (1980), Statistical Methods, Ames, Iowa: The Iowa State University Press.

Stenger, F. (1981), "Numerical Methods Based on Whittaker Cardinal or Sinc Function," SIAM Review, 23, no. 2.

Tarone, R.E. (1981), "The Use of Historical Control Information in Testing for a Trend in Proportions," Biometrics (to be published).

van Uven, M.J. (1947), "Extension of Pearson's Probability Distributions to Two Variables," Ned. Akad. Wet. Proc., 50, 1252-1264.

Woolf, B. (1955), "On Estimating the Relation Between Blood Group and Disease," Annals of Human Genetics, 19, 251-253.