ORNL-4700
UC-32 – Mathematics and Computers

BAYES-SUGGESTED SOLUTIONS IN BINOMIAL ESTIMATION

(Thesis)

T. A. DeRouen
T. J. Mitchell

OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.95

ORNL-4700

Contract No. W-7405-eng-26

MATHEMATICS DIVISION

BAYES-SUGGESTED SOLUTIONS IN BINOMIAL ESTIMATION

T. A. DeRouen and T. J. Mitchell

Submitted as a dissertation to the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Statistics

AUGUST 1971

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U. S. ATOMIC ENERGY COMMISSION

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Mrs. Mary Kau for her assistance with computer programs requiring the use of the Calcomp plotter, and to Mrs. Elizabeth Christian and Mrs. Betty Mashburn for typing the manuscript. One of the authors also expresses his gratitude to Oak Ridge Associated Universities for the financial aid received through the Oak Ridge Graduate Fellowship Program at Oak Ridge National Laboratory, sponsored by the U. S. Atomic Energy Commission under contract with Union Carbide Corporation.

ABSTRACT

This dissertation is concerned with the estimation of three functions of the binomial probability of "success": estimation of a linear combination of independent binomial probabilities; fixed precision estimation of the binomial probability; and estimation of the logit transformation of the binomial probability. A Bayesian viewpoint is adopted temporarily to "suggest" a wide class of admissible estimators for each problem. Designated the class C of SBP estimators, it is the class of Bayes estimators derived from Symmetric Beta Priors (the class of conjugate priors for the binomial), and often includes the maximum likelihood estimator as a special case.

Once a class of estimators for each problem is suggested by the Bayesian viewpoint, three criteria are used to obtain the "optimum" estimators in that class. Two of these criteria are classical in nature: minimax risk and minimax weighted risk. The third criterion, the solution of which is the estimator corresponding to the "least favorable" prior in the class of priors considered, is subjective in nature and would appeal more to Bayesians.

For each of the three problems, a class C of SBP estimators is suggested, and the optimum estimators from this class are obtained. In addition, for a special case in the estimation of a linear combination of binomials, an estimator is found that is minimax among all estimators, as well as minimax among SBP estimators.

CHAPTER I. INTRODUCTION

The experimenter frequently encounters the situation where he observes the same experiment a fixed number (n) of times, and each observation can be classified in one of two ways, which we shall arbitrarily term "success" and "failure." After collecting n such observations, he may be interested in using the information from this sample to make inferences about the population from which the sample was drawn. For example, he may want to estimate the proportion of "successes" in his experiment if he were able to subject the entire population to the experimental conditions. The sample of size n represents a collection of n independent and identically distributed Bernoulli random variables, each having probability distribution

$$\left.\begin{aligned} Pr(\text{success}) &= \theta , \\ Pr(\text{failure}) &= 1 - \theta \end{aligned}\right\} \quad 0 \le \theta \le 1 \ .$$

The sum of Bernoulli random variables is itself a random variable, say X, where X represents the number of successes observed in n Bernoulli trials. X has probability distribution that is binomial with parameters n and θ, i.e.,

$$Pr(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

for $0 \le \theta \le 1$, and $x = 0,1,\ldots,n$. The interest of the experimenter in estimating the probability of success in the entire population can now be formulated as the problem of statistical estimation of the parameter θ in the binomial distribution.

This thesis is a study of three problems of estimation involving binomial probabilities. The first is the estimation of a linear combination of independent binomial probabilities, when loss is measured by squared error. A special case of this problem, estimation of the difference between two binomial probabilities, was discussed (for equal sample sizes) in a paper by Hodges and Lehmann [14].

The second problem involves fixed precision estimation of θ. The problem of finding optimal fixed precision estimators of the binomial parameter θ was suggested as early as 1957 by Steinhaus [20], and one estimator was discussed as a special case in a paper by Naddeo [18].

The third problem is the estimation of the logit of θ, $\ell n(\theta/(1-\theta))$, when the loss is squared error. The logit transformation is often used in the analysis of multidimensional contingency tables (see Woolf [23]) and in the estimation of the parameters of the logistic function in quantal bioassay problems (see Berkson [3]).

1.1 The General Approach to the Three Estimation Problems

The general problem of statistical point estimation may be formulated from a decision-theoretic point of view as follows. The population parameter of interest, θ, is unknown and is to be estimated. Although the parameter itself cannot be observed, a random variable X whose probability distribution, $p(x|\theta)$, is a function of θ is available. Based on an observation of X, an estimate $\hat{\theta}(x)$ is chosen for θ. To evaluate an estimate, we must measure its accuracy in estimating θ with a loss function, $L(\hat{\theta}(x),\theta)$. This loss function is defined such that, for any estimate $\hat{\theta}(x)$, and any value of θ, the loss is nonnegative, and the loss is equal to 0 if the estimate is correct.

It is possible to specify a rule that determines the estimate $\hat{\theta}(x)$ to be made for any observation of X. Denoting this rule by $\hat{\theta}$, $\hat{\theta} = \{\hat{\theta}(x)\}$ for all x defined by X. Whereas $\hat{\theta}(x)$ is called an estimate of θ, the rule $\hat{\theta}$ is called an estimator of θ. The worth of an estimator is measured by a risk function, $R(\hat{\theta},\theta)$, which is equal to the average loss incurred when the estimator $\hat{\theta} = \{\hat{\theta}(x)\}$ is used to estimate θ, i.e., $R(\hat{\theta},\theta) = \mathcal{E}_X[L(\hat{\theta}(x),\theta)]$.

The evaluation of an estimator is usually made by considering some measure of the overall performance of its risk function. For example, an estimator may be considered optimal if its risk function has maximum value over Θ that is smaller than that for any other estimator. Such an estimator is designated a minimax estimator. If θ, the parameter being estimated, is considered a random variable with known density function $\xi(\theta)$, called the prior, then the worth of an estimator may be measured by its expected risk. If Θ is the space of values θ may assume, the expected risk (or "global" risk) for an estimator $\hat{\theta}$ is defined by

$$\mathcal{E}_\theta[R(\hat{\theta},\theta)] = \int_\Theta R(\hat{\theta},\theta)\ \xi(\theta)\ d\theta \quad .$$

For this situation, the optimum estimator is the one that minimizes the expected risk, designated the Bayes estimator, while the minimum expected risk is designated the Bayes risk.

In this thesis, two general measures of optimality for estimators will be emphasized. The first is the criterion of minimax risk. Introduced by Wald [21], the estimator that minimizes the maximum risk (often called simply the minimax estimator) is widely used in the decision-theoretic approach to estimation for situations in which it is desirable to be protected against the worst that can happen.

The second criterion is based on the integrated weighted risk:

$$\int_{\Theta} R(\hat{\theta}, \theta)\ \omega(\theta)\ d\theta \quad ,$$

where $\omega(\theta)$ is a weighting function such that $\int \omega(\theta)\ d\theta = 1$. This integrated weighted risk shall be called simply the weighted risk. Consider statistical estimation as a game played between the Statistician and an Adversary who has at his disposal a set of functions Ω from which he may choose the weighting function $\omega(\theta)$. Let the Statistician then choose the estimator that minimizes the maximum possible weighted risk for all $\omega(\theta) \in \Omega$. In other words, he chooses the estimator $\hat{\theta}$ to minimize

$$\max_{\omega(\theta)} \int_{\Theta} R(\hat{\theta}, \theta)\ \omega(\theta)\ d\theta \quad .$$

This shall be called the minimax weighted risk estimator.

The general concept of minimax weighted risk includes two other criteria as special cases. If Ω, the set of weighting functions available to the Adversary, includes all possible weighting functions, then full weight can be placed at the value of θ where the maximum risk occurs. Thus, the minimax weighted risk is equal to the minimax risk, and the minimax weighted risk estimator is the minimax estimator. On the other hand, if Ω contains only one function $\omega(\theta)$, then the estimator possessing minimax weighted risk is the Bayes estimator corresponding to the prior $\omega(\theta)$.

In general, the three problems discussed in this thesis involve the estimation of functions of probabilities from independent binomial distributions. When estimating a binomial probability, it is usually desirable to have estimates whose values are symmetric, i.e., $\hat{\theta}(x) = 1 - \hat{\theta}(n - x)$. When this property holds, the estimates for θ remain unchanged when the roles of "success" and "failure" in the binomial experiment are interchanged. An example of an estimator that does not possess this symmetry property is $\tilde{\theta} = (x + 1)/(n + 1)$. If, in a biological experiment performed to estimate the survival rate of mice in a polluted environment, "success" is defined to be survival and 8 out of 10 mice survive, the estimated survival rate is $\tilde{\theta} = 9/11$. If "success" is redefined in the same experiment to be death, then the estimated survival rate is $1 - \tilde{\theta} = 1 - 3/11$ or $8/11$. To avoid such discrepancies, only estimators with this symmetry property will be considered.

All loss functions used in this thesis also happen to have the property of symmetry in that $L(\hat{\theta}(x), \theta) = L(1 - \hat{\theta}(x), 1 - \theta)$. This, in conjunction with the symmetry of the estimators just mentioned, can be used to show that the corresponding risk functions, $R(\hat{\theta}, \theta)$, are symmetric. Consider the risk at θ:

$$R(\hat{\theta}, \theta) = \sum_{x=0}^{n} \binom{n}{x} \theta^x (1 - \theta)^{n-x} L(\hat{\theta}(x), \theta) \quad .$$

The risk at $1 - \theta$ is

$$R(\hat{\theta}, 1 - \theta) = \sum_{x=0}^{n} \binom{n}{x} (1 - \theta)^x \theta^{n-x} L(\hat{\theta}(x), 1 - \theta) \quad .$$

From the symmetry property of the estimator (i.e., $\hat{\theta}(x) = 1 - \hat{\theta}(n - x)$),

$$R(\hat{\theta}, 1 - \theta) = \sum_{x=0}^{n} \binom{n}{x} (1 - \theta)^x \theta^{n-x} L(1 - \hat{\theta}(n - x), 1 - \theta) \quad .$$

But the symmetry of the loss function indicates that

$$L(1 - \hat{\theta}(n - x), 1 - \theta) = L(\hat{\theta}(n - x), \theta) \quad ,$$

so that

$$R(\hat{\theta}, 1 - \theta) = \sum_{x=0}^{n} \binom{n}{x} (1 - \theta)^x \theta^{n-x} L(\hat{\theta}(n - x), \theta) \quad .$$

If $y = n - x$, then the above expression may be written as

$$R(\hat{\theta}, 1 - \theta) = \sum_{y=0}^{n} \binom{n}{y} \theta^y (1 - \theta)^{n-y} L(\hat{\theta}(y), \theta)$$

$$= R(\hat{\theta}, \theta)$$

and the risk function is symmetric about $\theta = 1/2$.

When comparing estimators with symmetric risk functions, using the criterion of minimax weighted risk, it is possible to restrict the class of weighting functions to the class of symmetric weighting functions. This is true because, for any weighted risk produced by a nonsymmetric weighting function, the same weighted risk may be produced by a symmetric weighting function. To demonstrate this, let WR denote the weighted risk produced by a nonsymmetric weighting function, $\omega(\theta)$. Then

$$WR = \int_0^1 R(\hat{\theta}, \theta)\, \omega(\theta)\, d\theta \quad .$$

Let $\omega^*(\theta) = \frac{\omega(\theta) + \omega(1 - \theta)}{2}$. Since $\omega^*(\theta) = \omega^*(1 - \theta)$, $\omega^*(\theta)$ is a symmetric weighting function. Let WR^* denote the weighted risk produced by $\omega^*(\theta)$. Then

$$WR^* = \int_0^1 R(\hat{\theta}, \theta)\ \omega^*(\theta)\ d\theta$$

$$= \int_0^1 R(\hat{\theta}, \theta) \left[\frac{\omega(\theta) + \omega(1 - \theta)}{2}\right] d\theta$$

$$= \tfrac{1}{2}\int_0^1 R(\hat{\theta}, \theta)\ \omega(\theta)\ d\theta + \tfrac{1}{2}\int_0^1 R(\hat{\theta}, \theta)\ \omega(1 - \theta)\ d\theta \quad .$$

From the symmetry of the risk function, $R(\hat{\theta}, \theta) = R(\hat{\theta}, 1 - \theta)$, so that the second term in the above expression may be written

$$\tfrac{1}{2}\int_0^1 R(\hat{\theta}, 1 - \theta)\ \omega(1 - \theta)\ d\theta$$

$$= \tfrac{1}{2}\int_1^0 R(\hat{\theta}, 1 - \theta)\ \omega(1 - \theta)\ d(1 - \theta)$$

$$= \tfrac{1}{2}\int_0^1 R(\hat{\theta}, \theta)\ \omega(\theta)\ d\theta \quad .$$

Therefore

$$WR^* = \int_0^1 R(\hat{\theta}, \theta)\ \omega(\theta)\ d\theta = WR \quad .$$

The particular class Ω of symmetric weighting functions $\omega(\theta)$ to be considered in this thesis is the class of symmetric beta functions defined on Θ:

$$\Omega = \left\{\omega(\theta) \,|\, \omega(\theta) = \frac{\theta^b (1 - \theta)^b}{B(b + 1, b + 1)}\right\}, \quad b > -1 \quad ,$$

where $B(b + 1, b + 1) = \int_0^1 \theta^b(1 - \theta)^b \, d\theta$. This class of symmetric weighting functions was chosen because it is a fairly flexible class of one-parameter functions. For example, the members of Ω range from functions that approach a discrete distribution with weight 1/2 at $\theta = 0$ and 1 (for b near -1), to a function that weighs all intervals in Θ uniformly ($b = 0$), to functions that place full weight at $\theta = 1/2$ (for very large b). The flexibility of this class of functions, together with the ease of computations achieved by using a beta weighting function indexed by only one parameter, led to this choice for Ω. In Chapter II, where Θ is k-dimensional, this concept is generalized to a class of independent symmetric beta weighting functions of the form

$$\Omega = \left\{ \omega(\underline{\theta}) \,|\, \omega(\underline{\theta}) = \prod_{i=1}^{k} \frac{\theta_i^{b_i} (1 - \theta_i)^{b_i}}{B(b_i + 1, b_i + 1)} \right\}, \quad b_i > -1 \ .$$

In searching for optimum solutions to the three problems to be discussed in this thesis, consideration will be restricted to a particular class of estimators. This class will consist of estimators that can be derived as Bayes estimators corresponding to certain prior distributions on Θ. Although a Bayesian viewpoint is needed to conceive of θ having a prior distribution, Bayesian theory will be used here only to <u>suggest</u> a class of estimators from which optimum estimators may be selected. Therefore, the estimators under consideration will be referred to as "Bayes-suggested" estimators. They will be derived from the class Ξ of symmetric beta prior distributions on Θ, $\xi(\theta)$, where

$$\xi(\theta) = \frac{\theta^{a}(1-\theta)^{a}}{B(a+1, a+1)}, \qquad a > -1 .$$

Since all Bayes estimators are admissible, the class of Bayes estimators from symmetric beta priors constitute a class of admissible estimators.

This particular class of priors, Ξ, was chosen for several reasons. First of all, from a Bayesian viewpoint, the beta prior is the most commonly used prior for problems involving the binomial (see, for example, Lindley [17] and Good [8]). It meets the Raiffa and Schlaifer criteria of tractability, richness, and ease of interpretation ([19], p. 44), making it the "natural conjugate" for the binomial.

Moreover, the symmetry of the prior ensures, in the problems to be considered in this thesis, that the estimates have the symmetry property stipulated above, namely, $\hat{\theta}(x) = 1 - \hat{\theta}(n - x)$. This can be shown as follows. If x is observed, the posterior density of θ is

$$\xi(\theta|x) = \frac{\xi(\theta)p(x|\theta)}{\int_0^1 \xi(v)\ p(x|v)\ dv}$$

where $p(x|\theta) = \binom{n}{x}\ \theta^{x}(1-\theta)^{n-x}$. If $n - x$ is observed, the posterior density of θ is

$$\xi(\theta|n-x) = \frac{\xi(\theta)\ p(n-x|\theta)}{\int_0^1 \xi(v)\ p(n-x|v)\ dv} .$$

$$\xi(1-\theta|n-x) = \frac{\xi(1-\theta)\ p(n-x|1-\theta)}{\int_0^1 \xi(v)\ p(n-x|v)\ dv} = \frac{\xi(1-\theta)\ p(n-x|1-\theta)}{\int_1^0 \xi(1-u)\ p(n-x|1-u)\ d(1-u)} .$$

Using (i) $\xi(\theta) = \xi(1 - \theta)$ (from the symmetry of the prior), and (ii) $p(x|\theta) = p(n - x|1 - \theta)$ (a property of the binomial distribution),

$$\xi(1 - \theta|n - x) = \frac{\xi(\theta)\ p(x|\theta)}{\int_0^1 \xi(u)\ p(x|u)\ du}$$

$$= \xi(\theta|x) \quad .$$

It will be demonstrated later (Section 3.1.1) that the Bayes estimator is made up of the set of estimates that minimize the posterior expected loss for $x = 0,1,\ldots,n$. Let $\hat{\theta}(x)$ be the Bayes estimate when x is observed, i.e.,

$$\int_0^1 \xi(\theta|x)\ L(\hat{\theta}(x),\theta)\ d\theta \le \int_0^1 \xi(\theta|x)\ L(z,\theta)\ d\theta \quad \text{for all } z \ .$$

Using the fact that $\xi(\theta|x) = \xi(1 - \theta|n - x)$ developed above and the symmetry property of the risk function $(L(\hat{\theta}(x),\theta) = L(1 - \hat{\theta}(x),1 - \theta))$, the above inequality may be written

$$\int_0^1 \xi(1-\theta|n-x)\ L(1-\hat{\theta}(x),1-\theta)\ d\theta \le \int_0^1 \xi(1-\theta|n-x)\ L(1-z,1-\theta)\ d\theta \quad \text{for all } z$$

$$\Longrightarrow \int_0^1 \xi(u|n-x)\ L(1-\hat{\theta}(x),u)\ du \le \int_0^1 \xi(u|n-x)\ L(w,u)\ du \quad \text{for all } w$$

$$\Longrightarrow \hat{\theta}(n - x) = 1 - \hat{\theta}(x) \quad .$$

Finally, note that each member of the class of Bayes-suggested estimators derived from symmetric beta priors is optimum in the sense that it minimizes the weighted risk for some weighting function $\omega(\theta)$ in Ω, which has already been chosen to consist of all symmetric beta weighting functions (see Section 1.2).

In Chapter II, where Θ is k-dimensional, the class Ξ of priors will consist of the set of k-dimensional independent symmetric beta priors, $\xi(\underline{\theta})$, of the form

$$\xi(\underline{\theta}) = \prod_{i=1}^{k} \frac{\theta_i^{a_i}(1-\theta_i)^{a_i}}{B(a_i+1, a_i+1)} , \qquad a_i > -1 .$$

The class of estimators considered in this thesis will be denoted as the class C of admissible SBP (Symmetric Beta Prior) estimators. The optimum SBP estimators will be offered as solutions to the three problems discussed. Unless otherwise specified, the "optimum" estimators derived here will be those estimators that are best in the class C. For example, the term "minimax weighted risk" estimator will be used to refer to the estimator in C that minimizes the maximum weighted risk, where the class of weighting functions is the class of symmetric beta functions. However, referring to the minimax estimator in C, the term "C-minimax" will often be employed to distinguish it from the "universal minimax" estimator.

In addition to the C-minimax estimators and minimax weighted risk estimators, a third type of estimator will be derived, namely, the Bayes estimator associated with the "least favorable prior." The least favorable prior is that prior distribution in Ξ which maximizes the Bayes risk for all priors in Ξ. From a Bayesian point of view, it represents the prior distribution against which the Statistician will be least successful in minimizing the expected risk, even if he knows the prior and selects his estimator accordingly.

Under certain conditions, the Bayes estimator corresponding to the least favorable prior will be the minimax estimator, so the least favorable prior can be used in some situations to derive minimax estimators (see Wald [22]). Unfortunately, these conditions do not hold for the problems considered here, mainly because of the restrictions on Ξ, so the derivation of least favorable priors and their corresponding Bayes estimators may be only of academic interest.

The order in which the three estimation problems will be discussed is as follows. In Chapter II, the problem of estimating a linear combination of binomial probabilities is considered. Chapter III concerns the search for optimal fixed precision estimators of θ. Chapter IV deals with the estimation of the logit (i.e., $\ell n \frac{\theta}{1 - \theta}$). In each chapter, estimators in C are found that (i) minimize the maximum risk; (ii) minimize the maximum weighted risk; and (iii) correspond to the least favorable prior.

1.2 Notation

The following notation will be employed throughout this thesis.

X is a random variable with outcomes x in a sample space χ.

$p(x|\theta)$ is the conditional probability distribution of X. In this thesis, the conditional distribution of X is binomial with parameters n and θ, so that $p(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$, for all $x \in \chi$.

χ is the sample space of a binomial distribution with parameters n and θ; i.e., $\chi = \{0,1,2,...,n\}$.

n is the sample size for the binomial distribution. It will be considered a fixed quantity.

θ is the parameter to be estimated, the probability of "success" in the binomial distribution. In some cases, more than one binomial distribution may be involved in the estimation problem, in which case the parameter to be estimated will be $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)'$, so that $\underline{\theta}$ is a vector whose ith element is the probability of "success" in the ith binomial distribution.

Θ is the parameter space for $\underline{\theta}$. If the estimation problem is one-dimensional, then the space Θ is just the interval $[0,1]$. If the estimation problem is k-dimensional, i.e., $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)'$, then the space Θ is the k-dimensional cube $0 < \theta_i < 1$, $i = 1,2,\ldots,k$.

T is the estimation space. The dimension of T is the same as the dimension of Θ.

$\hat{\theta}(x)$ is a point in T, an individual estimate of θ, which is a function of the outcome x.

$\hat{\theta}$ is a rule that specifies $\hat{\theta}(x)$ for every x in χ, designated an estimator of θ.

$L(\cdot,\cdot)$ is the loss function defined on $T \times \Theta$. Used for estimation problems, $L(\hat{\theta}(x),\theta)$ is equal to the loss incurred when X has distribution based on θ, outcome x is observed, and the estimate $\hat{\theta}(x)$ is chosen.

$R(\cdot,\cdot)$ is the risk (expected loss) function. $R(\hat{\theta},\theta)$ is the expected loss when estimator $\hat{\theta}$ is used and X has conditional probability distribution $p(x|\theta)$. The expectation is taken over the conditional distribution of X, so

$$R(\hat{\theta},\theta) = \sum_{x=0}^{n} L(\hat{\theta}(x),\theta) \binom{n}{x} \theta^x (1-\theta)^{n-x} .$$

Ω is the class of symmetric beta weighting functions of the form

$$\omega(\underline{\theta}) = \prod_{i=1}^{k} \frac{\theta_i^{b_i}(1-\theta_i)^{b_i}}{B(b_i+1, b_i+1)}, \quad b_i > -1, \quad k \geq 1,$$

which may be used to weight the risk functions.

Ξ is the class of symmetric beta density functions, defined on Θ, of the form

$$\xi(\underline{\theta}) = \prod_{i=1}^{k} \frac{\theta_i^{a_i}(1-\theta_i)^{a_i}}{B(a_i+1, a_i+1)}, \quad a_i > -1, \quad k \geq 1 .$$

This class of density functions is used to suggest a class C of SBP estimators (Bayes estimators from symmetric beta priors) from which optimal estimators are obtained.

$\hat{\theta}_\xi(x)$ will be used to designate the Bayes estimate of θ based on outcome x, for the prior distribution $\xi(\theta)$. Similarly, the Bayes estimator from $\xi(\theta)$ will be denoted by $\hat{\theta}_\xi = \{\hat{\theta}_\xi(x),\ x = 0,1,\ldots,n\}$. Although it will be demonstrated in Chapter III that the Bayes estimator for a prior is made up of those estimates that minimize the posterior expected loss, the Bayes estimator is defined to be the one that minimizes the weighted or average risk, when the risk function is weighted by the prior $\xi(\theta)$, i.e.,

$$\int_\Theta R(\hat{\theta}_\xi, \theta)\ \xi(\theta)\ d\theta = \inf_{\hat{\theta}} \int_\Theta R(\hat{\theta}, \theta)\ \xi(\theta)\ d\theta .$$

Thus, each SBP estimator is optimum (it has minimum weighted risk) for a weighting function $\omega(\theta) \in \Omega$, if $\omega(\theta) = \xi(\theta)$. Each SBP estimator is also an admissible estimator, since all Bayes estimators are admissible. Because all estimators in the class C under consideration are Bayes estimators, the subscript ξ will often be omitted.

$\widetilde{R}(\xi,\omega)$ is the expected or weighted risk for an SBP estimator based on the prior $\xi(\theta)$, when its risk function is weighted by the weighting function $\omega(\theta)$:

$$\widetilde{R}(\xi,\omega) = \int_{\Theta} R(\hat{\theta}_{\xi},\theta)\ \omega(\theta)\ d\theta \quad .$$

Therefore, $\widetilde{R}(\xi,\xi)$ is the Bayes risk for the prior $\xi(\theta)$.

CHAPTER II. ESTIMATION OF A LINEAR COMBINATION OF BINOMIAL PROBABILITIES

Let X_i, $i = 1,2,\ldots,k$, be independent random variables, each having a binomial distribution with parameters n_i and θ_i, and let γ be an arbitrary linear combination of the binomial probabilities of "success," i.e.,

$$\gamma = \sum_{i=1}^{k} \alpha_i \theta_i = \underline{\alpha}' \underline{\theta} \tag{2.1}$$

where the α_i's are specified constants. Estimation of γ is the topic for this chapter.

Examples of situations in which estimation of a linear combination of binomial probabilities is of interest include the following.

(1) If a comparison of the proportions of response for one treatment, T, versus a control, C, is desired, then estimation of $\gamma = \theta_T - \theta_C$, the difference between the probabilities of response for the two groups, is one method that may be used. In this case, $\alpha_1 = 1$, and $\alpha_2 = -1$.

(2) It is often of interest to compare the average proportion of response for several treatment groups with a control group. If t equals the number of treatment groups, then this comparison may be made by estimating

$$\gamma = \frac{1}{t} \sum_{i=1}^{t} \theta_i - \theta_C ,$$

so that $\alpha_1 = \alpha_2 = \ldots = \alpha_t = \frac{1}{t}$, and $\alpha_{t+1} = -1$.

(3) In a 2^n factorial experimental design with proportions as observations, a main effect for a factor is defined to be the sum of the

probabilities of "success" at the upper level of the factor minus the sum of the probabilities of "success" at the lower level of the factor, all divided by 2^{n-1}. Thus, estimation of a main effect involves the estimation of a linear combination of 2^n binomial probabilities, half with coefficient $1/2^{n-1}$, and the other half with coefficient $-1/2^{n-1}$.

This chapter is a study of the estimation of γ using the squared error loss function; i.e., $L(\hat{\gamma},\gamma) = (\hat{\gamma}-\gamma)^2$, where $\hat{\gamma}$ is an estimate of γ. In this study, consideration is restricted to the class of Bayes estimates derived from symmetric beta prior distributions on Θ (referred to as the class C of SBP estimates). The chapter is divided into three sections, dealing with (i) the search for minimax estimates of γ among the class C of SBP estimates; (ii) the derivation of the SBP estimates which minimize the maximum weighted risk, when the risk function is weighted by any member of the class of symmetric beta distributions; and (iii) the derivation of the set of Bayes estimates corresponding to the "least favorable" symmetric beta prior.

To obtain an expression for the SBP estimates, first consider the "prior" distribution over the parameter space Θ, which is the unit hypercube $\{0 \le \theta_1 \le 1,\ 0 \le \theta_2 \le 1,\ \ldots,\ 0 \le \theta_k \le 1\}$. Define the prior to be

$$\xi(\underline{\theta}) = \prod_{i=1}^{k} \xi_i(\theta_i) \qquad (2.2)$$

where $\xi_i(\theta_i)$ is a symmetric beta function of the form

$$\xi_i(\theta_i) = \frac{\theta_i^{a_i} (1 - \theta_i)^{a_i}}{B(a_i+1, a_i+1)} \quad , \quad a_i > -1 \quad . \qquad (2.3)$$

Therefore, the posterior distribution of θ_i is given by

$$\xi_i(\theta_i|x_i) = \frac{\theta_i^{\alpha_i+x_i}(1-\theta)^{\alpha_i+n_i-x_i}}{B(\alpha_i+x_i+1,\alpha_i+n_i-x_i+1)} \quad . \tag{2.4}$$

If $\underline{\alpha}'\underline{\theta}$ is to be estimated using the squared error loss function, then the familiar properties of that loss function (see, for example, Lehmann [16], Chapter IV, p. 31) say that the Bayes estimate of $\underline{\alpha}'\underline{\theta}$ is $\mathcal{E}(\underline{\alpha}'\underline{\theta}|\underline{x})$, the posterior expectation of $\underline{\alpha}'\underline{\theta}$. Now,

$$\mathcal{E}(\theta_i|x_i) = \int_0^1 \theta_i\, \xi_i(\theta_i|x_i)\, d\theta_i \tag{2.5}$$

$$= \frac{\int_0^1 \theta_i^{\alpha_i+x_i+1}(1-\theta_i)^{\alpha_i+n_i-x_i}\, d\theta_i}{B(\alpha_i+x_i+1,\alpha_i+n_i-x_i+1)} \tag{2.6}$$

$$= \frac{B(\alpha_i+x_i+2,\alpha_i+n_i-x_i+1)}{B(\alpha_i+x_i+1,\alpha_i+n_i-x_i+1)}$$

$$= \frac{\alpha_i+x_i+1}{2\alpha_i+n_i+2} \quad , \tag{2.7}$$

a well-known result (see [8], p. 17, for its equivalent).

Thus, $\mathcal{E}(\underline{\alpha}'\underline{\theta}|\underline{x}) = \sum_i \alpha_i\hat{\theta}_i$, where $\hat{\theta}_i = \dfrac{x_i + c_i}{n_i + 2c_i}$ and $c_i = \alpha_i + 1$.

The SBP estimate of $\gamma = \sum_i \alpha_i\theta_i$ is therefore of the form

$$\hat{\gamma} = \sum_i \alpha_i \left(\frac{x_i + c_i}{n_i + 2c_i}\right), \quad c_i > 0 \ , \ i = 1,2,\ldots,k \quad . \tag{2.8}$$

This chapter is concerned with finding the set of constants, $\{c_1, c_2, \ldots, c_k\}$ in (2.8) which have optimum properties for the three estimation criteria given in Chapter I.

2.1 Estimates Which Minimize the Maximum Risk

The search for estimates of $\gamma = \sum \alpha_i \theta_i$ which minimize the maximum of the risk function was conducted by analyzing the form of the risk function for the class C of estimates given in (2.8).

Estimators that are minimax in the class C (designated C-minimax) were obtained in two special cases:

Case 1. All sample sizes are the same ($n_1 = n_2 = \ldots = n_k$), and $\alpha_i = \pm 1$, for $i = 1,2,\ldots,k$. (These conditions pertain in particular to factorial experimental designs.) It is shown that, in this case, the C-minimax estimators are also universal-minimax estimators.

Case 2. Estimation of the difference between probabilities for two groups: $\gamma = \theta_1 - \theta_2$, n_1 not necessarily equal to n_2.

Although estimates which minimize the maximum risk are obtained for only these two special cases, it is convenient to present first a discussion of the general behavior of the risk function for estimators of the form (2.8).

2.1.1 The Form of the Risk Function

If $\gamma = \sum \alpha_i \theta_i$, and $\hat{\gamma} = \sum \alpha_i \left(\frac{x_i + c_i}{n_i + 2c_i} \right)$, then

$$\mathcal{E}_x(\hat{\gamma}) = \sum \alpha_i \left(\frac{n_i \theta_i + c_i}{n_i + 2c_i} \right) ,$$

since $\mathcal{E}_x(x_i) = n_i \theta_i$. The bias of $\hat{\gamma}$ is defined to be $\mathcal{E}_x(\hat{\gamma}) - \gamma$, so

$$\text{Bias}_x(\hat{\gamma}) = \sum\left[\frac{\alpha_i n_i \theta_i + \alpha_i c_i}{n_i + 2c_i} - \frac{\alpha_i \theta_i (n_i + 2c_i)}{n_i + 2c_i}\right]$$

$$= \sum\left[\frac{c_i \alpha_i (1 - 2\theta_i)}{n_i + 2c_i}\right] .$$

The variance of $\hat{\gamma}$ is given by

$$\text{Var}_x(\hat{\gamma}) = \sum \frac{\alpha_i^2}{(n_i + 2c_i)^2} n_i \theta_i (1 - \theta_i) ,$$

since $\text{Var}_x(x_i) = n_i \theta_i (1 - \theta_i)$. Let $z_i = \theta_i - 1/2$. Then $1 - 2\theta_i = -2z_i$ and

$$\text{Bias}_x^2(\hat{\gamma}) = \left[\sum \frac{2c_i \alpha_i z_i}{n_i + 2c_i}\right]^2 .$$

Also, $\theta_i(1 - \theta_i) = (\frac{1}{2} + z_i)(\frac{1}{2} - z_i) = \frac{1}{4} - z_i^2$, so that

$$\text{Var}_x(\hat{\gamma}) = \sum \frac{\alpha_i^2}{(n_i + 2c_i)^2} n_i (\tfrac{1}{4} - z_i^2) .$$

The risk function, when the loss function is squared error, may be expressed as

$$R(\hat{\gamma}, \underline{z}) = \mathcal{E}_x(\hat{\gamma} - \gamma)^2 \qquad \text{(mean squared error)}$$

$$= \text{Bias}_x^2(\hat{\gamma}) + \text{Var}_x(\hat{\gamma})$$

$$= \left[\sum \frac{2c_i \alpha_i z_i}{n_i + 2c_i}\right]^2 + \sum \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} (\tfrac{1}{4} - z_i^2) . \qquad (2.9)$$

Now let

$$y_i = \frac{\sqrt{n_i}\ \alpha_i\ z_i}{n_i + 2c_i} \tag{2.10}$$

and

$$v_i = \frac{2c_i}{\sqrt{n_i}} \ . \tag{2.11}$$

Then $\sum_i \frac{2c_i \alpha_i z_i}{n_i + 2c_i} = \sum_i y_i v_i$, and $\sum_i \frac{n_i \alpha_i^2 z_i^2}{(n_i + 2c_i)^2} = \sum_i y_i^2$, so that

$$R(\hat{\gamma},\underline{y}) = \underline{y}'\ v\ \underline{v}'\ \underline{y} - \underline{y}'\ \underline{y} + \tfrac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} \ , \tag{2.12}$$

where $\underline{y}' = (y_1, y_2, \ldots, y_k)$ and $\underline{v}' = (v_1, v_2, \ldots, v_k)$. Alternatively,

$$R(\hat{\gamma},\underline{y}) = \underline{y}'\ A\underline{y} + \tfrac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} \ , \tag{2.13}$$

where

$$A = \underline{v}\ \underline{v}' - I_k \tag{2.14}$$

and I_k is the unit matrix of order k.

With the risk function expressed as a quadratic form plus a constant term, as in (2.13), it is now possible to analyze the behavior of this risk function for given $(c_1, c_2, \ldots, c_k)$. This is facilitated by expressing the quadratic form, $\underline{y}'\ A\underline{y}$, in (2.13), in terms of canonical variables $Y_1, Y_2, \ldots, Y_k$. This transformation permits the simplified expression

$$\underline{y}'\ A\underline{y} = \sum_{i=1}^{k} \lambda_i\ Y_i^2 \ , \tag{2.15}$$

where the λ_i are the eigenvalues of the matrix A. This is accomplished by performing an orthogonal transformation on $\underline{y}$, $\underline{y} = P\underline{Y}$, such that $\underline{y}' A\underline{y}$ is equal to $\underline{Y}'P'AP\underline{Y}$, a quadratic form in the canonical variables $(Y_1, Y_2, \ldots, Y_k)$. The matrix of this quadratic form, $P'AP$, is a diagonal matrix with diagonal elements $(\lambda_1, \lambda_2, \ldots, \lambda_k)$. Such a transformation may be accomplished by forming the orthogonal matrix $P = [\underline{p}_1, \underline{p}_2, \cdots, \underline{p}_k]$, where $\underline{p}_i$ is a linear invariant vector (eigenvector) corresponding to the eigenvalue λ_i. Under this transformation, $Y_i = \underline{p}_i' \underline{y}$, for $i = 1,2,\ldots,k$.

The axis for the ith canonical variable, Y_i, is given by $\underline{y} = \mu\, \underline{p}_i$ (where μ is any real number). This may be seen by finding the values of the canonical variables when $\underline{y} = \mu\, \underline{p}_i$:

$$Y_i = \underline{p}_i' \underline{y} = \mu\, \underline{p}_i' \underline{p}_i = \mu \quad ;$$

$$Y_j = \underline{p}_j' \underline{y} = \mu\, \underline{p}_j' \underline{p}_u = 0 \quad , \quad \text{for } j \neq i \quad ;$$

both results due to the orthogonality of P.

In order to find the eigenvalues of the matrix A, it is necessary to solve the determinantal equation

$$|A - \lambda I| = 0 \quad . \tag{2.16}$$

Recall from (2.14) that $A = \underline{v}\,\underline{v}' - I$. Therefore,

$$|A - \lambda I| = \begin{vmatrix} v_1^2-1-\lambda & v_1 v_2 & \cdots & v_1 v_k \\ v_1 v_2 & v_2^2-1-\lambda & \cdots & v_2 v_k \\ \vdots & & \ddots & \\ v_1 v_k & v_2 v_k & \cdots & v_k^2-1-\lambda \end{vmatrix} .$$

Now factor out v_i from the ith row and ith column, $i = 1,2,\ldots,k$, to obtain

$$|A - \lambda I| = v_1^2 v_2^2 \cdots v_k^2 \begin{vmatrix} 1 - \frac{1+\lambda}{v_1^2} & 1 & \cdots & 1 \\ 1 & 1 - \frac{1+\lambda}{v_2^2} & \cdots & 1 \\ \vdots & & \ddots & \\ 1 & 1 & & 1 - \frac{1+\lambda}{v_k^2} \end{vmatrix} \qquad (2.17)$$

The last term in (2.17) is the determinant of a matrix having the same form as the matrix B in (2.60) of Appendix 2A. Therefore, this determinant may be evaluated using (2.61), replacing β_i with $\frac{1+\lambda}{v_i^2}$. This yields the result

$$|A - \lambda I| = v_1^2 v_2^2 \cdots v_k^2 \, (-1)^{k-1} \frac{(1+\lambda)^k}{v_1^2 v_k^2 \cdots v_k^2} \left[\frac{v_1^2}{1+\lambda} + \frac{v_2^2}{1+\lambda} + \ldots + \frac{v_k^2}{1+\lambda} - 1 \right]$$

$$= (-1)^{k-1} (1+\lambda)^{k-1} [v_1^2 + v_2^2 + \ldots + v_k^2 - 1 - \lambda] \quad . \qquad (2.18)$$

Expression (2.18) implies that, of the k roots of the equation $|A - \lambda I| = 0$, $k - 1$ of them are given by $\lambda = -1$, and the other root is $\lambda = \sum_{i=1}^{k} v_i^2 - 1$. The quadratic form $\underline{y}' A\underline{y}$ may now be written in canonical form:

$$\underline{y}' A\underline{y} = \left(\sum_i v_i^2 - 1 \right) Y_1^2 - \sum_{j=2}^{k} Y_j^2 \quad . \qquad (2.19)$$

Since Y_1 is the only canonical variable whose coefficient in (2.19) is not always negative, the Y_1 axis will be of considerable interest.

This axis may be obtained by finding the eigenvector corresponding to $\lambda = \underline{v}'\,\underline{v} - 1$. This eigenvector is the solution for $\underline{x}$ in the equation

$$[A - (\underline{v}'\underline{v} - 1)\, I]\,\underline{x} = \underline{0} \;. \tag{2.20}$$

Since $A = \underline{v}\,\underline{v}' - I$, (2.20) may be written

$$[\underline{v}\,\underline{v}' - (\underline{v}'\underline{v})\, I]\,\underline{x} = \underline{0} \;. \tag{2.21}$$

This equation has solution

$$\underline{x} = \mu\,\underline{v} \qquad (\mu \in \text{Re}) \;, \tag{2.22}$$

since replacing $\underline{x}$ by $\mu\,\underline{v}$ in (2.21) yields the identity

$$\mu\,\underline{v}(\underline{v}'\underline{v}) - \mu(\underline{v}'\underline{v})\,\underline{v} = \underline{0} \;.$$

The axis corresponding to Y_1 is given by

$$\underline{y} = \mu\,\underline{v} \;.$$

Recall from (2.10) and (2.11) that $y_i = \dfrac{\sqrt{n_i}\,\alpha_i\,z_i}{n_i + 2c_i}$, and $v_i = \dfrac{2c_i}{\sqrt{n_i}}$, so that, along the Y_1 axis, $y_i = \mu\,v_i$, or

$$z_i = \frac{2c_i(n_i + 2c_i)}{n_i\alpha_i}\,\mu \;, \qquad \text{for } i = 1,2,\ldots,k \;. \tag{2.23}$$

The risk (2.13) may now be written

$$R(\hat{\gamma},\underline{Y}) = \tfrac{1}{4}\sum \frac{n_i\alpha_i^2}{(n_i + 2c_i)^2} + \left(\sum v_i^2 - 1\right) Y_1^2 - \sum_{j=2}^{k} Y_j^2 \;. \tag{2.24}$$

Substitution of $\dfrac{2c_i}{\sqrt{n_i}}$ for v_i yields

$$R(\hat{\gamma},\underline{Y}) = \tfrac{1}{4}\sum \frac{n_i\alpha_i^2}{(n_i + 2c_i)^2} + \left(\sum \frac{4c_i^2}{n_i} - 1\right)Y_1^2 - \sum_{j=2}^{k} Y_j^2 \quad . \qquad (2.25)$$

This expression can now be used to study the behavior of the risk function, in particular the location of its maximum for fixed $(c_1, c_2, \ldots, c_k)$.

Case 1: $\sum_i \frac{4c_i^2}{n_i} < 1$. In this case, all eigenvalues of the canonical form are negative, and the maximum of (2.25) occurs when $(Y_1, Y_2, \ldots, Y_k) = (0,0,\ldots,0)$. (This corresponds to the point $\theta_1 = \theta_2 = \ldots = \theta_k = .5$.) The risk at this point is $\frac{1}{4}\sum_i \frac{n_i\alpha_i^2}{(n_i + 2c_i)^2}$. A special example of this case is the choice of all $c_i = 0$, which corresponds to the maximum likelihood estimate of γ.

Case 2: $\sum_i \frac{4c_i^2}{n_i} > 1$. If $Y_2, Y_3, \ldots, Y_k$ are all held constant at zero, then the risk function (2.25) increases as Y_1 moves in either direction away from the canonical origin. Therefore, the maximum of the risk function occurs on the boundary of the parameter space. The exact location of the maximum does not appear easy to determine.

Case 3: $\sum_i \frac{4c_i^2}{n_i} = 1$. In this case, the risk function (2.25) attains its maximum value at all points on the Y_1 axis. Along that axis, the risk function has a constant value of $\frac{1}{4}\sum_i \frac{n_i\alpha_i^2}{(n_i + 2c_i)^2}$.

2.1.2 Two Special Cases

Using the class of SBP estimates, minimax estimates for two special cases will now be derived. In the first case (Section 2.1.2.1), all sample sizes are equal and the absolute values of the α_i's are equal. It is shown that the class of SBP estimates contains a universal minimax estimator. The results are applicable in particular to the design and analysis of factorial experiments for binomial responses. In the second case (Section 2.1.2.2), estimation of the difference between two binomial parameters is considered without requiring equal sample sizes. The estimates which are obtained, after a somewhat involved proof, are C-minimax; it is not known whether or not they are universal minimax.

2.1.2.1 <u>Minimax Estimator for γ When $n_i = n$, $\alpha_i = \pm 1$, for $i = 1,2,\ldots,k$</u>. According to Theorem 2.1 of Hodges and Lehmann [14], estimates which have risk that is maximum and constant over some subspace Φ of Θ, and which can be expressed as Bayes estimates from some prior distribution on Φ, are the estimates that minimize the maximum value of the risk function. This theorem will now be used to demonstrate the following theorem.

<u>Theorem 2.1</u>. If $n_1 = n_2 = \cdots = n_k = n$, and $\alpha_i = \pm 1$ for $i = 1,2,\ldots,k$, then the SBP estimates

$$\hat{\gamma} = \sum \alpha_i \left(\frac{x_i + c_i}{n_i + 2c_i} \right) , \tag{2.26}$$

based on the set of constants $\left\{ c_i = \sqrt{\frac{n}{4k}} \text{ , } i = 1,2,\ldots,k \right\}$, are universal minimax.

Proof: If $c_i = \sqrt{\frac{n}{4k}}$, then $\sum \frac{4c_i^2}{n} = 1$. From Section 2.1.1 (Case 3), the risk function is maximum and constant along the eigenvector corresponding to the eigenvalue 0. The value of the maximum risk is

$$\frac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} = \frac{k}{4\left(\sqrt{n} + \frac{1}{\sqrt{k}}\right)^2} \tag{2.27}$$

and from (2.23) it achieves this maximum all along the line

$$z_i = \frac{\sqrt{\frac{n}{k}} + \frac{1}{k}}{\alpha_i} \cdot \mu \quad (\mu \in Re) , \quad \text{for } i = 1,2,\ldots,k . \tag{2.28}$$

But (2.28) may be equivalently described as the subspace $\Phi = \{\alpha_1 z_1 = \alpha_2 z_2 = \ldots = \alpha_k z_k\}$. In order to apply the Hodges-Lehmann theorem, it remains only to find a prior distribution on Φ whose Bayes estimates of γ are given by (2.26).

In Φ, since $\alpha_1 z_1 = \alpha_j z_j$ for $j = 2,\ldots,k$, and since $z_i = \theta_i - \frac{1}{2}$, then

$$\alpha_1 \theta_1 - \frac{\alpha_1}{2} = \alpha_j \theta_j - \frac{\alpha_j}{2}$$

$$\Longrightarrow \theta_j = \frac{\alpha_1}{\alpha_j} \theta_1 - \frac{1}{2}\left(\frac{\alpha_1}{\alpha_j} - 1\right) \qquad \text{for } j = 2,\ldots,k .$$

Suppose there are η positive α_i's. Label the θ_i's such that $\{\theta_1, \theta_2, \ldots, \theta_\eta\}$ correspond to the positive α_i's. Then

$$\gamma = \sum \alpha_i \theta_i = \sum_{i=1}^{\eta} \theta_i + \sum_{i=\eta+1}^{k} (-1) \theta_i ,$$

and

$$\Phi = \{\underline{\theta} | \theta_1 = \theta_2 = \ldots = \theta_\eta = 1 - \theta_{\eta+1} = \ldots = 1 - \theta_k\} . \tag{2.29}$$

A particular point in the subspace Φ can be written

$$\varphi = \{\theta_1, \theta_1, \ldots, \theta_1, 1 - \theta_1, \ldots, 1 - \theta_1\} .$$

Assume some symmetric beta prior density for θ_1:

$$\xi(\theta_1) = \frac{\theta_1^{a}(1-\theta_1)^{a}}{B(a+1,a+1)} . \tag{2.30}$$

This defines the probability density over Φ, since, for given θ_1, Φ is described by

$$\theta_2 = \theta_3 = \dots = \theta_\eta = \theta_1 ,$$

$$\theta_{\eta+1} = \dots = \theta_k = 1 - \theta_1 .$$

The conditional distributions of the random variables X_i, $i = 1,2,\dots,k$, over the subspace Φ, are given by

$$f(x_i|\theta_i) = \begin{cases} \binom{n}{x_i}\theta_1^{x_i}(1-\theta_1)^{n-x_i} & \text{for } 1 \le i \le \eta \\ \binom{n}{x_i}\theta_1^{n-x_i}(1-\theta_1)^{x_i} & \text{for } \eta+1 \le i \le k . \end{cases}$$

Therefore,

$$f(\underline{x}|\varphi) = \left[\prod_{i=1}^{k}\binom{n}{x_i}\right]\theta_1^{\sum_{i=1}^{\eta}x_i+\sum_{i=\eta+1}^{k}(n-x_i)}(1-\theta_1)^{\sum_{i=1}^{\eta}(n-x_i)+\sum_{i=\eta+1}^{k}x_i} .$$

Since $\xi(\varphi)$, the prior probability density for φ is defined by the prior density for θ_1, $\xi(\theta_1)$ in (2.30), then the posterior distribution of φ is given by

$$\xi(\varphi|\underline{x}) = \frac{\xi(\varphi)\ f(x|\varphi)}{\int_{\Phi} \xi(\varphi)\ f(\underline{x}|\varphi)\ d\varphi}$$

$$= \frac{\theta_1^{\,a+\sum_{i=1}^{\eta} x_i+\sum_{i=\eta+1}^{k} (n-x_i)}\ (1-\theta_1)^{\,a+\sum_{i=1}^{\eta} (n-x_i)+\sum_{i=\eta+1}^{k} x_i}}{B\left(\sum_{i=1}^{\eta} x_i+\sum_{i=\eta+1}^{k} (n-x_i)+a+1,\ \sum_{i=1}^{\eta} (n-x_i)+\sum_{i=\eta+1}^{k} x_i+a+1\right)} .$$

The Bayes estimate for γ is the posterior mean of γ, since the loss function is squared error. But

$$\gamma = \sum_{i=1}^{\eta} \theta_i + \sum_{i=\eta+1}^{k} (-\theta_i)$$

$$= k\ \theta_1 - k + \eta ,$$

so that $\widetilde{\gamma}$, the Bayes estimate of γ is

$$\widetilde{\gamma} = \mathcal{E}_{(\varphi|\underline{x})}(\gamma) = \mathcal{E}_{(\varphi|\underline{x})}(k\ \theta_1 - k + \eta)$$

$$= k\ \mathcal{E}_{(\varphi|\underline{x})}(\theta_1) - k + \eta ,$$

where

$$\mathcal{E}_{(\varphi|\underline{x})}(\theta_1) = \frac{\sum_{i=1}^{\eta} x_i + \sum_{i=\eta+1}^{k} (n-x_i) + a + 1}{kn + 2a + 2} .$$

Therefore,

$$\tilde{\gamma} = \frac{k\left[\sum_{i=1}^{\eta} x_i + \sum_{i=\eta+1}^{k} (n-x_i) + \alpha + 1\right]}{kn + 2\alpha + 2} - k + \eta$$

$$= \frac{\left[\sum_{i=1}^{\eta} x_i - \sum_{i=\eta+1}^{k} x_i\right] + \frac{(\alpha+1)(2\eta-k)}{k}}{n + \frac{2(\alpha+1)}{k}} \quad . \qquad (2.31)$$

From (2.26), the SBP estimate for γ when $\left\{c_i = \sqrt{\frac{n}{4k}}\ ,\ i = 1,2,\ldots,k\right\}$, is

$$\hat{\gamma} = \sum_{i=1}^{\eta}\left[\frac{x_i + \sqrt{\frac{n}{4k}}}{n + 2\sqrt{\frac{n}{4k}}}\right] - \sum_{i=\eta+1}^{k}\left[\frac{x_i + \sqrt{\frac{n}{4k}}}{n + 2\sqrt{\frac{n}{4k}}}\right]$$

$$= \frac{\left[\sum_{i=1}^{\eta} x_i - \sum_{i=\eta+1}^{k} x_i\right] + \frac{1}{2}\sqrt{\frac{n}{k}}\ (2\eta - k)}{n + \sqrt{\frac{n}{k}}} \quad . \qquad (2.32)$$

The SBP estimate for γ in (2.32) is equal to the Bayes estimate of γ based on the prior $\xi(\varphi)$ in (2.30) when $\alpha = \frac{\sqrt{nk}}{2} - 1$. This can be demonstrated by replacing α by $\frac{\sqrt{nk}}{2} - 1$ in (2.31), yielding the same expression as (2.32).

Since $\hat{\gamma}$ in (2.26), with $\left\{c_i = \sqrt{\frac{n}{4k}}\ ,\ i = 1,\ldots,k\right\}$, has risk function that is maximum and constant over the subspace Φ in (2.29), and can be expressed as the Bayes estimate of γ for prior $\xi(\varphi)$ in (2.30) when $\alpha = \frac{\sqrt{nk}}{2} - 1$, then $\hat{\gamma}$ satisfies the Hodges-Lehmann criteria for a universal minimax estimate.

The minimax estimate for the difference of two binomials, when $n_1 = n_2$, was given by Hodges and Lehmann [14]. It should be pointed out, since the difference of two binomials is a special case of Theorem 2.1, that the SBP estimate in (2.26) defined by Theorem 2.1 for $k = 2$ is identical to the Hodges-Lehmann estimate.

2.1.2.2. C-Minimax Estimator for $\gamma = \theta_1 - \theta_2$, $n_1 \neq n_2$. A minimax estimator for the difference between two binomial parameters was found by Hodges and Lehmann [14] for case $n_1 = n_2$. In this section, the difference $\gamma = \theta_1 - \theta_2$ is considered for $n_1 \neq n_2$. Attention is limited to the class of SBP estimators $\hat{\gamma}$, given in (2.8), and the estimator in C that minimizes the maximum risk is obtained. Although this estimator is C-minimax, it is not necessarily universal minimax.

Development of the C-Minimax Estimator

The SBP estimators of $\gamma = \theta_1 - \theta_2$ are of the form

$$\hat{\gamma} = \frac{x_1 + c_1}{n_1 + 2c_1} - \frac{x_2 + c_2}{n_2 + 2c_2} \quad ; \quad c_1, c_2 > 0 . \qquad (2.33)$$

The search for the SBP estimator that minimizes the maximum risk involves finding the set of constants $\{c_1, c_2\}$ that produce this result. Any estimator of the type in (2.33) may be considered as a point (c_1, c_2) in the positive quadrant of two-dimensional C-space, represented in Figure 2.1. The optimum estimator will be obtained by finding the point in C-space that produces the minimax risk.

The form of the risk function for the SBP estimators was discussed in Section 2.1.1. Frequent references will be made to expressions developed in Section 2.1.1, especially the risk function as a constant plus a quadratic form in (2.13), and the canonical representation of the quadratic form in (2.19).

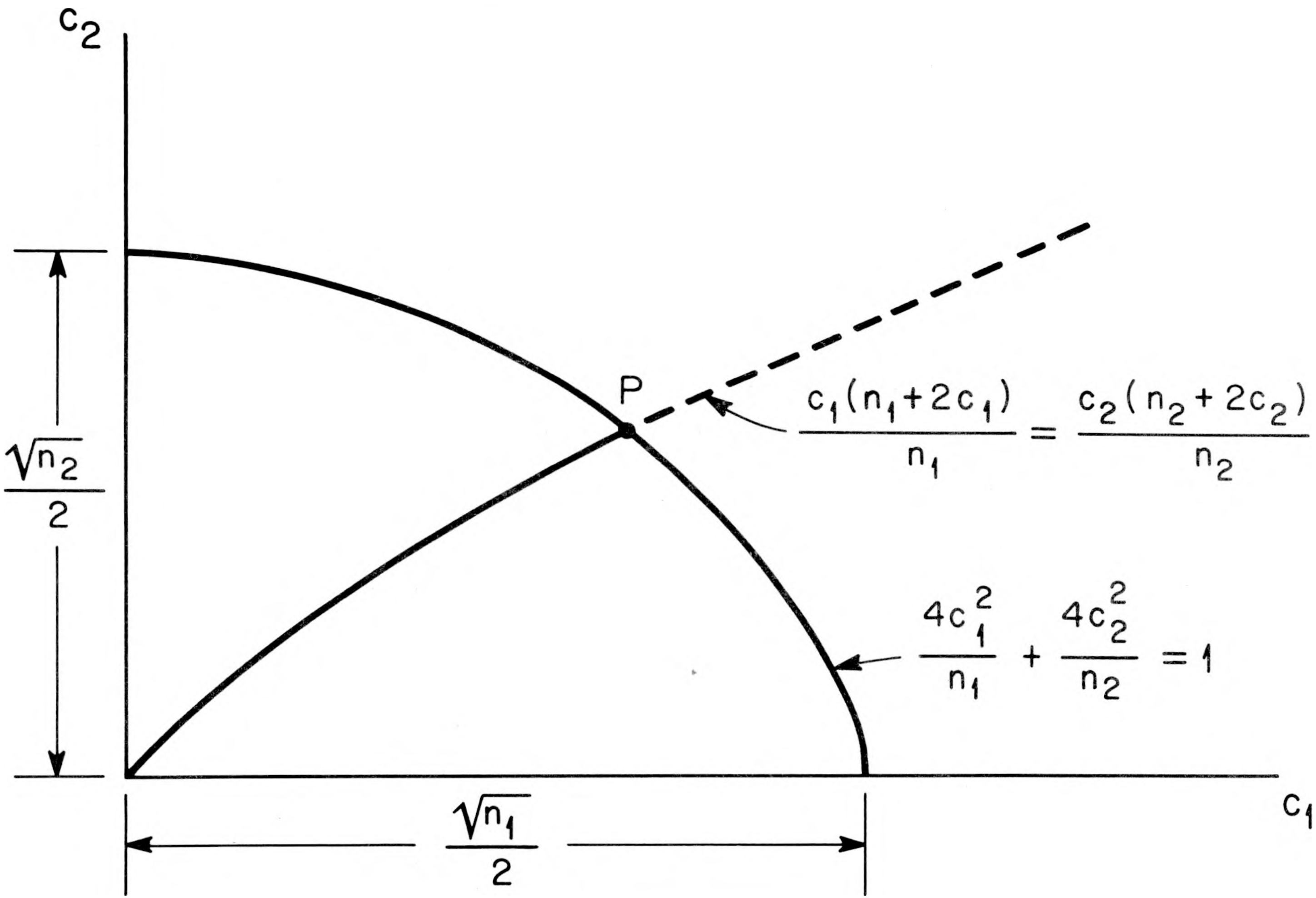

Figure 2.1. Partition of Two-Dimensional C-Space

In Figure 2.1, the ellipse $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} = 1$ corresponds to the condition that the eigenvalue, $\lambda = \sum_{i=1}^{2} \frac{4c_i^2}{n_i} - 1$, is zero. The importance of this eigenvalue, as a coefficient in the canonical representation of the risk, was discussed in Section 2.1.1.

Theorem 2.2. Let $R^*(c_1, c_2)$ denote the maximum risk over Θ for the estimator in (2.33) that corresponds to the point (c_1, c_2). Then there exists a point (c_1', c_2') on the ellipse $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} = 1$ such that $R^*(c_1', c_2') \leq R^*(c_1, c_2)$.

The proof of this theorem is rather long and tedious, and only a summary of it will appear in this section. A detailed version is given on pp. 44-55.

Theorem 2.2 is proved by partitioning the (c_1, c_2) space into several regions. For each of these regions, it is then shown that, starting at any point (c_1, c_2) in that region, the maximum risk may be continuously reduced by following a path which eventually leads to a point on the ellipse $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} = 1$.

Case 1: (c_1, c_2) lies on the ellipse, and the theorem is true at once.

Case 2: (c_1, c_2) lies within the ellipse, i.e., $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} < 1$. All eigenvalues of the matrix A are negative so that, in the canonical representation of the risk function in (2.25), the coefficients of all of the canonical variables are negative. The risk assumes its maximum value at the center of the parameter space Θ (i.e., when $\theta_1 = \theta_2 = \frac{1}{2}$), and this maximum risk is

$$R^*(c_1,c_2) = \frac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] .$$

$R^*(c_1,c_2)$ can now be continuously reduced by increasing c_1 and/or c_2 until the ellipse is encountered.

Case 3: (c_1,c_2) lies outside the ellipse on the dashed curve in Figure 2.1. For (c_1,c_2) to be on the dashed line, the slope of the Y_1 axis, with respect to the y_1 axis, must be equal to the slope of the line extending from the center A, of the parameter space, to the corner D, as in Figure 2.2.

The maximum risk for this case occurs at corners D and F in Figure 2.2, where the risk is

$$R^*(c_1,c_2) = \left[\frac{c_1}{n_1 + 2c_1} + \frac{c_2}{n_2 + 2c_2}\right]^2 .$$

$R^*(c_1,c_2)$ can be reduced by decreasing c_1 and c_2 simultaneously, moving along the dashed line in Figure 2.1 to point P. Let (c_1',c_2') equal the coordinates of P, and Theorem 2.2 is satisfied.

Case 4: (c_1,c_2) lies outside the ellipse to the right of the dashed curve. For any point (c_1,c_2) satisfying this case, the slope of the Y_2 axis is less than the slope of the line AD, as indicated in Figure 2.3.

The maximum risk for (c_1,c_2) can be shown to occur either at point D, or in the interval CD in the parameter space pictured in Figure 2.3. (Equivalently the same maximum risk also occurs either at point F, or in the interval EF, because of the symmetry of the risk function about each canonical axis.) The location of the maximum risk depends on the

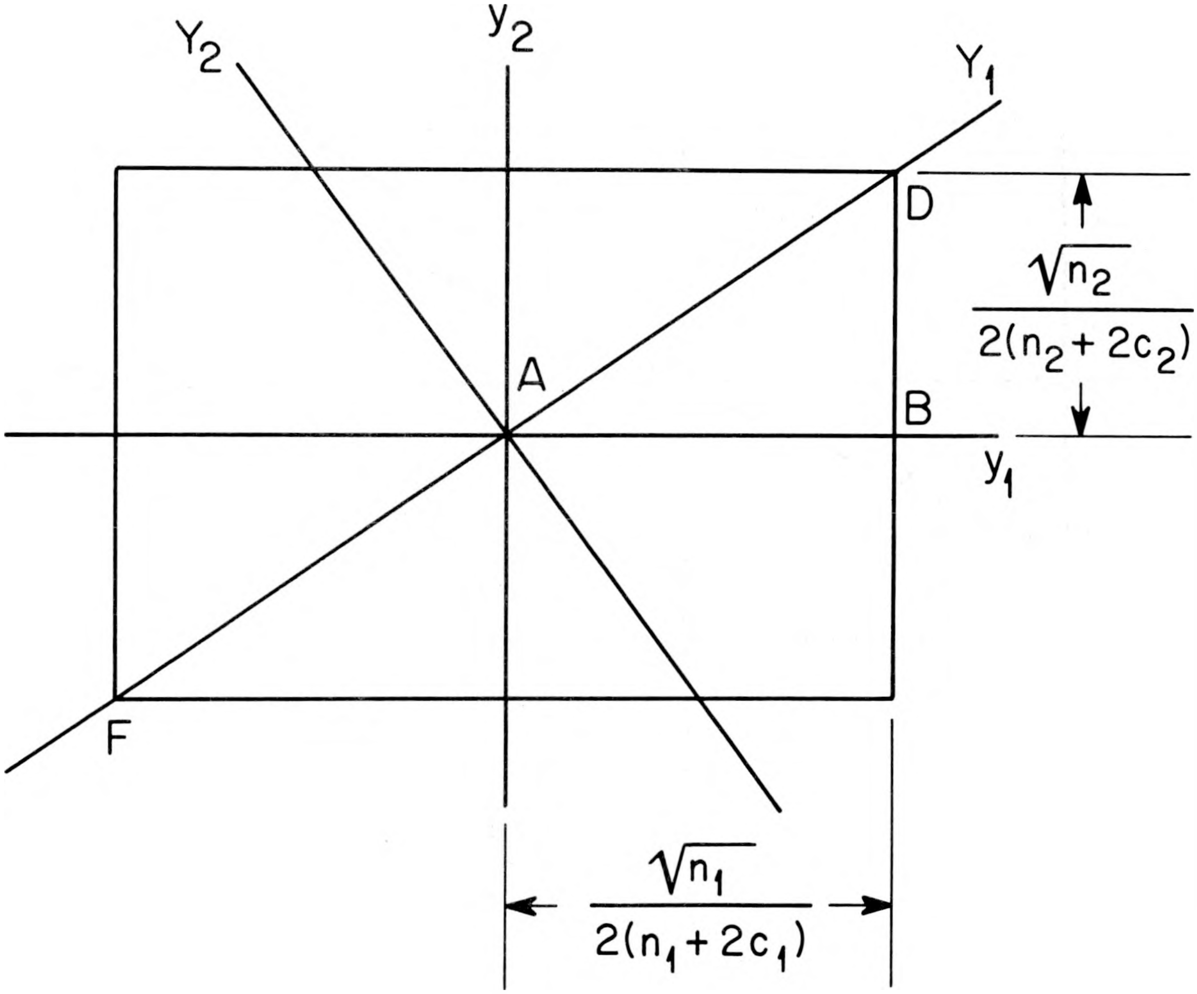

Figure 2.2. Canonical Axes: Case 3

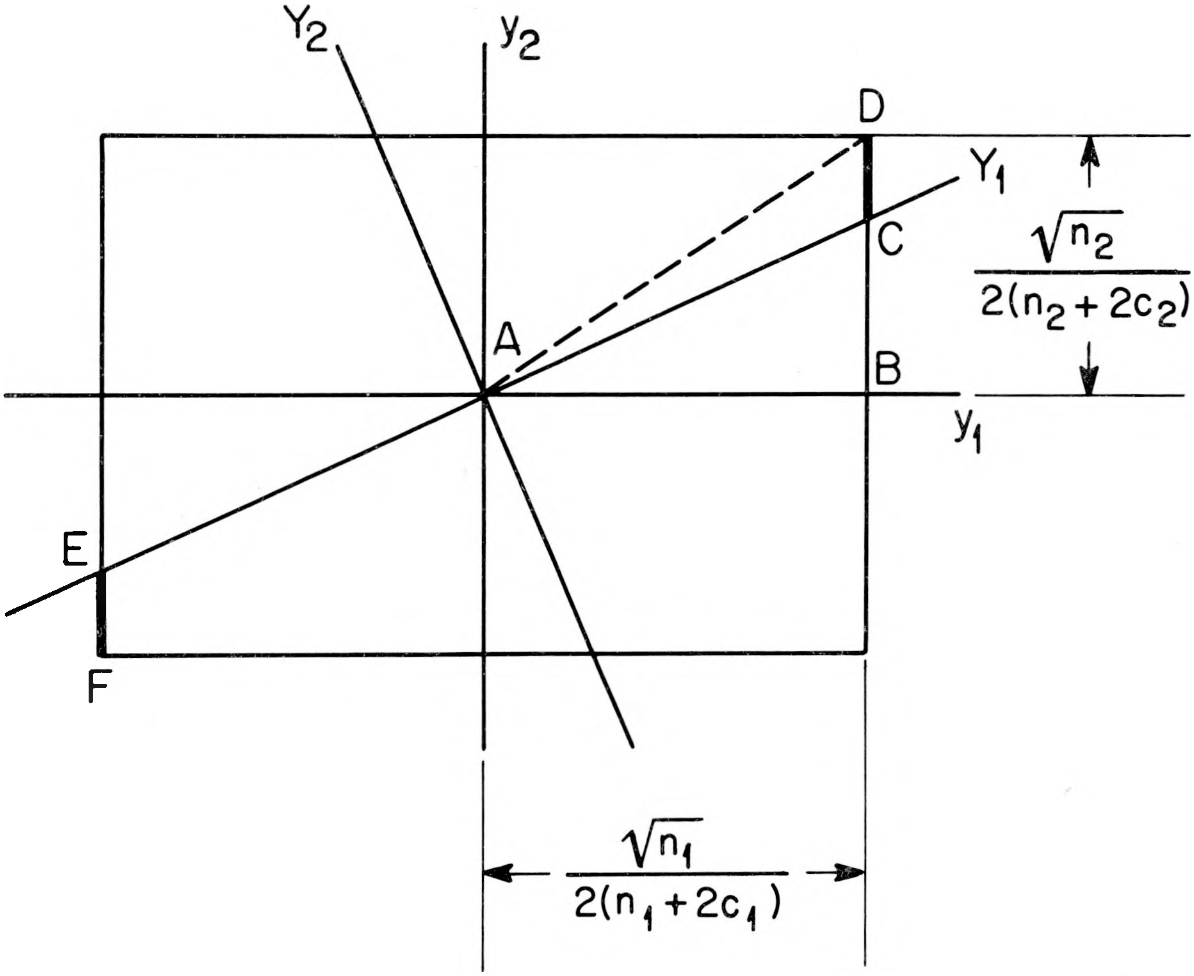

Figure 2.3. Canonical Axes: Case 4

sign of $\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_2}\right|_D$, i.e., the derivative of the risk, with respect to y_2 evaluated at D.

Case 4a: If (c_1,c_2) is such that $\left.\frac{\partial R}{\partial y_2}\right|_D$ is negative, the maximum risk occurs between C and D and has value

$$R^*(c_1,c_2) = \frac{1}{4}\left[\frac{n_2}{(n_2 + 2c_2)^2} - \frac{4c_1^2}{(n_1 + 2c_1)^2\left(\frac{4c_2^2}{n_2} - 1\right)}\right].$$

But $\left.\frac{\partial R}{\partial y_2}\right|_D < 0 \Rightarrow \frac{4c_2^2}{n_2} - 1 < 0$, making the second term in the above expression for $R^*(c_1,c_2)$ positive. Thus, the maximum risk can be reduced by decreasing c_1 until the ellipse in Figure 2.1 is encountered, in which case the theorem holds, or until the dashed line is encountered. In the latter case, conditions for Case 3 are again satisfied, and the theorem is true.

Case 4b: If (c_1,c_2) is such that $\left.\frac{\partial R}{\partial y_2}\right|_D$ is positive or zero, the maximum risk occurs at D, and has value

$$R^*(c_1,c_2) = \left[\frac{c_1}{n_1 + 2c_1} + \frac{c_2}{n_2 + 2c_2}\right]^2 .$$

This maximum risk may be continously reduced by decreasing c_1 in C-space until

(a) the ellipse is encountered;

(b) the dashed line is encountered, Case 3 holds, and the theorem is true;

(c) $\left.\frac{\partial R}{\partial y_2}\right|_D$ becomes negative, Case 4a holds, and the theorem is true.

Case 5: (c_1, c_2) lies outside the ellipse to the left of the dashed curve in Figure 2.1. All points in this case satisfy the condition that the slope of the Y_1 axis is greater than the slope of line AD, as demonstrated in Figure 2.4. As in Case 4, the maximum risk is known to occur at corners D and F, or in the shaded intervals CD and EF. The location of the maximum is dependent on the sign of $\left.\frac{\partial R(\hat{\gamma}, \underline{y})}{\partial y_1}\right|_D$.

Case 5a: $\left.\frac{\partial R(\hat{\gamma}, \underline{y})}{\partial y_1}\right|_D < 0$. The maximum risk occurs between C and D and has value

$$R^*(c_1, c_2) = \frac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} - \frac{4c_2^2}{(n_2 + 2c_2)^2\left(\frac{4c_1^2}{n_1} - 1\right)}\right] .$$

But $\left.\frac{\partial R}{\partial y_1}\right|_D < 0 \Rightarrow \frac{4c_1^2}{n_1} - 1 < 0$, so the second term in the expression for $R^*(c_1, c_2)$ is positive. The maximum risk, therefore, can be continuously reduced by decreasing c_2 until the ellipse or dashed line is encountered. If the ellipse, the theorem follows immediately. If the dashed line, Case 3 holds, and the theorem is true.

Case 5b: $\left.\frac{\partial R(\gamma, \underline{y})}{\partial y_1}\right|_D \geq 0$. The maximum risk occurs at the corner D of the parameter space, and is equal to

$$R^*(c_1, c_2) = \left[\frac{c_1}{n_1 + 2c_1} + \frac{c_2}{n_2 + 2c_2}\right]^2 .$$

This maximum risk may be continuously reduced by decreasing c_2 until

(a) the ellipse is encountered, Case 1 holds, and the theorem is true;

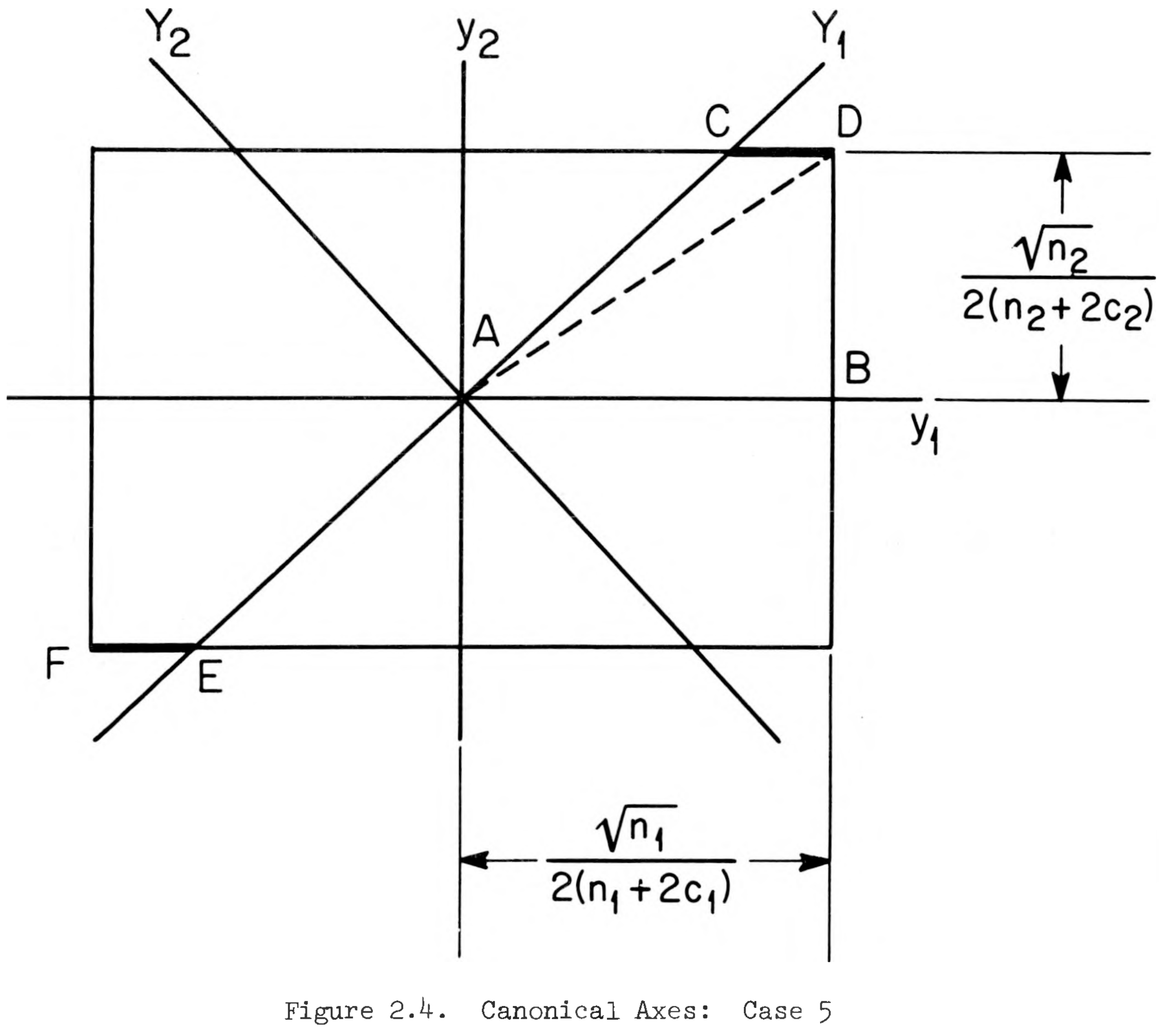

Figure 2.4. Canonical Axes: Case 5

(b) the dashed line is encountered, Case 3 holds, and the theorem is true;

(c) $\left.\frac{\partial R}{\partial y_1}\right|_D$ becomes negative, Case 5a holds, and the theorem is true.

Cases 1 through 5 include all points (c_1, c_2) such that $c_1 \geq 0$, $c_2 \geq 0$. Since it has been demonstrated that Theorem 2.2 is true for all five cases, the theorem is proved.

From Theorem 2.2, it is known that the point (c_1^*, c_2^*) having minimax risk is on the ellipse $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} = 1$. The maximum risk for all points on this ellipse is given by the expression

$$R^*(c_1, c_2) = \tfrac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] . \qquad (2.34)$$

To find the point (c_1^*, c_2^*) having minimax risk, it is necessary to minimize $R^*(c_1, c_2)$ with respect to c_1 and c_2, subject to the restriction that (c_1, c_2) is on the ellipse. This will be done using a Lagrange multiplier.

Let

$$F = \tfrac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] - \lambda\left[\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} - 1\right] .$$

Then

$$\frac{\partial F}{\partial c_1} = \tfrac{1}{4}\left[\frac{-4n_1}{(n_1 + 2c_1)^3}\right] - \frac{8\lambda c_1}{n_1} ;$$

$$\frac{\partial F}{\partial c_2} = \tfrac{1}{4}\left[\frac{-4n_2}{(n_2 + 2c_2)^3}\right] - \frac{8\lambda c_2}{n_2} ; \quad \text{and}$$

$$\frac{\partial F}{\partial \lambda} = \frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} - 1 .$$

Now, $\frac{\partial F}{\partial c_1}$, $\frac{\partial F}{\partial c_2}$, and $\frac{\partial F}{\partial \lambda}$ must be set equal to zero, and the three resulting equations solved simultaneously. But

$$\frac{\partial F}{\partial c_1} = 0 \Rightarrow \lambda = \frac{n_1}{8c_1}\left[\frac{-n_1}{(n_1 + 2c_1)^3}\right] ,$$

and

$$\frac{\partial F}{\partial c_2} = 0 \Rightarrow \lambda = \frac{n_2}{8c_2}\left[\frac{-n_2}{(n_2 + 2c_2)^3}\right] ,$$

so that the unknown λ may be omitted by combining these first two equations to give:

$$\frac{n_1^2}{c_1(n_1 + 2c_1)^3} = \frac{n_2^2}{c_2(n_2 + 2c_2)^3}$$

or

$$\frac{c_1(n_1 + 2c_1)^3}{n_1^2} = \frac{c_2(n_2 + 2c_2)^3}{n_2^2} .$$

This equation must be solved simultaneously with the equation

$$\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} = 1$$

in order to obtain the point (c_1^*, c_2^*) having minimax risk.

These equations have been solved numerically for many values of n_1 and n_2, and the results are given in Table I. For each pair (n_1, n_2),

Table I

C-Minimax Estimators of $\gamma = \theta_1 - \theta_2$

The C-minimax estimator of γ is of the form

$$\hat{\gamma} = \frac{x_1 + c_1^*}{n_1 + 2c_1^*} - \frac{x_2 + c_2^*}{n_2 + 2c_2^*} .$$

For each pair (n_1, n_2) the following entries are given (in order):

(i) c_1^*

(ii) c_2^*

(iii) the C-minimax risk corresponding to c_1^*, c_2^*.

The C-minimax estimators for those values of (n_1, n_2) in Table I with entries omitted may be obtained from the cell for (n_2, n_1), in which case the entries are (in order):

(i) c_2^*

(ii) c_1^*

(iii) the C-minimax risk corresponding to c_1^*, c_2^*.

For example, for $n_1 = 1$, $n_2 = 2$, the entries given in the table are

$c_1^* = .40345$

$c_2^* = .41767$

C-minimax risk = .13877 .

For $n_1 = 2$, $n_2 = 1$, the C-minimax estimator is obtained from the cell for $n_1 = 1$, $n_2 = 2$ by interchanging the roles of c_1^* and c_2^*:

$c_2^* = .40345$

$c_1^* = .41767$

C-minimax risk = .13877 .

Table I (continued)

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10
1	.35355 .35355 .17157	.40345 .41767 .13877	.43231 .43511 .12198	.45101 .43167 .11138	.46377 .41782 .10399	.47273 .39891 .09851	.47916 .37795 .09428	.48384 .35662 .09091	.48730 .33589 .08817	.48989 .31626 .08589
2		.50000 .50000 .11111	.55103 .54273 .09689	.58637 .55887 .08785	.61214 .55965 .08147	.63147 .55111 .07669	.64625 .53689 .07297	.65770 .51933 .06997	.66667 .50000 .06750	.67375 .47993 .06543
3			.61237 .61237 .08404	.66243 .64414 .07585	.70022 .65788 .07007	.72959 .65983 .06571	.75284 .65387 .06229	.77147 .64255 .05954	.78655 .62769 .05725	.79885 .61057 .05533
4				.70711 .70711 .06823	.75582 .73207 .06283	.79456 .74365 .05877	.82593 .74580 .05557	.85165 .74121 .05299	.87293 .73177 .05084	.89066 .71887 .04903
5					.79057 .79057 .05772	.83792 .81084 .05387	.87692 .82062 .05084	.90943 .82262 .48382	.93676 .81881 .04634	.95992 .81063 .04461
6						.86603 .86603 .05018	.91209 .88285 .04727	.95099 .89114 .04492	.98414 .89286 .04296	1.0125 .88950 .04131
7							.93541 .93541 .04447	.98028 .94961 .04220	1.0189 .95665 .04032	1.0523 .95804 .03872
8								1.0000 1.0000 .04000	1.0437 1.0121 .03817	1.0820 1.0181 .03662
9									1.0607 1.0607 .03638	1.1034 1.0711 .03487
10										1.1180 1.1180 .03339

n_1 \ n_2	10	20	30	40	50	60	70	80	90	100
10	1.1180 1.1180 .03339	1.3889 1.0685 .02551	1.4910 .91114 .02224	1.5342 .76487 .02044	1.5543 .64845 .01929	1.5646 .55840 .01852	1.5703 .48831 .01795	1.5737 .43285 .01752	1.5758 .38817 .01718	1.5772 .35155 .01691
20		1.5811 1.5811 .01864	1.8520 1.5347 .01575	2.0014 1.4101 .01413	2.0867 1.2703 .01308	2.1372 1.1390 .01235	2.1681 1.0236 .01181	2.1877 .92456 .01140	2.2007 .84024 .01107	2.2095 .76834 .01081
30			1.9365 1.9365 .01307	2.1949 1.8911 .01156	2.3628 1.7874 .01057	2.4734 1.6628 .00988	2.5472 1.5363 .00937	2.5975 1.4168 .00897	2.6324 1.3079 .00865	2.6572 1.2102 .00840
40				2.2361 2.2361 .01011	2.4816 2.1914 .00917	2.6564 2.1013 .00851	2.7815 1.9900 .00801	2.8718 1.8722 .00763	2.9376 1.7558 .00732	2.9861 1.6455 .00707
50					2.5000 2.5000 .00826	2.7337 2.4561 .00762	2.9103 2.3754 .00714	3.0440 2.2748 .00677	3.1455 2.1657 .00647	3.2232 2.0548 .00630
60						2.7386 2.7386 .00699	2.9619 2.6954 .00653	3.1377 2.6216 .00617	3.2762 2.5297 .00588	3.3856 2.4283 .00564
70							2.9580 2.9580 .00607	3.1721 2.9156 .00572	3.3460 2.8471 .00543	3.4871 2.7621 .00520
80								3.1623 3.1623 .00537	3.3682 3.1205 .00509	3.5394 3.0563 .00486
90									3.3541 3.3541 .00481	3.5527 3.3130 .00459
100										3.5355 3.5355 .00436

the point (c_1^*, c_2^*) corresponding to the C-minimax estimator of the form (2.33) is given, along with the value of the minimax risk (obtained by evaluating (2.34) at (c_1^*, c_2^*)).

Proof of Theorem 2.2

Case 1: (c_1, c_2) lies on the ellipse. Let $(c_1', c_2') = (c_1, c_2)$ and the theorem is satisfied.

Case 2: (c_1, c_2) lies within the ellipse, i.e., $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} < 1$. When this is true, then the eigenvalue of A given by

$$\lambda = \sum \frac{4c_i^2}{n_i} - 1 \tag{2.35}$$

and all other eigenvalues of A are negative. Recall that the eigenvalues of A are the coefficients of the canonical variables when the quadratic form $\underline{y}'A\underline{y}$ is expressed in canonical form. From Section 2.1.1 (Case 1) it is known that, when all the eigenvalues of A are negative, the risk function attains its maximum at the center of the parameter space, $(\theta_1, \theta_2) = (\frac{1}{2}, \frac{1}{2})$ or $(z_1, z_2) = (0,0)$. From (2.9), the risk at the center is given by

$$R^*(c_1, c_2) = \frac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] . \tag{2.36}$$

$R^*(c_1, c_2)$ in (2.36) can be reduced by increasing c_1 and/or c_2 until the ellipse is encountered at some point. Denoting this point by (c_1', c_2'), the theorem is satisfied.

For the remaining three cases, (c_1, c_2) lies outside the ellipse; i.e., $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} > 1$, and the eigenvalue in (2.35) is positive. From

Section 2.1.1 (Case 2), it is known that, when the eigenvalue in (2.35) is positive, the risk increases in both directions along the Y_1 axis. Along the Y_1 axis,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mu \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mu \begin{pmatrix} \frac{2c_1}{\sqrt{n_1}} \\ \frac{2c_2}{\sqrt{n_2}} \end{pmatrix} ,$$

so that the slope of this axis, with respect to the y_1 axis, is

$$\frac{v_2}{v_1} = \frac{c_2}{\sqrt{n_2}} \frac{\sqrt{n_1}}{c_1} . \tag{2.37}$$

In Figures 2.2, 2.3, and 2.4, note that AD is the line extending from the center to the corner of the parameter space. The slope of this line is equal to the ratio BD/AB. Using the relationship $y_i = \frac{\sqrt{n_i}\ \alpha_i\ z_i}{n_i + 2c_i}$, and the z-coordinates of the points A, B, and D, the lengths of these line segments were found to be

$$BD = \frac{\sqrt{n_2}}{2(n_2 + 2c_2)} , \tag{2.38}$$

$$AB = \frac{\sqrt{n_1}}{2(n_1 + 2c_1)} . \tag{2.39}$$

The slope of AD is therefore equal to

$$\frac{\sqrt{n_2}\ (n_1 + 2c_1)}{(n_2 + 2c_2)\ \sqrt{n_1}} . \tag{2.40}$$

When the slope of the Y_1 axis is equal to the slope of the line AD, expression (2.37) equals expression (2.40) and the following equation holds:

$$\frac{c_1(n_1 + 2c_1)}{n_1} = \frac{c_2(n_2 + 2c_2)}{n_2} \quad . \tag{2.41}$$

This equation is represented by the dashed curve in Figure 2.1.

When the slope of the Y_1 axis is less than the slope of the line AD, the inequality obtained by replacing the equal sign in (2.41) with a "greater than" sign holds. This corresponds to points lying to the right of the dashed curve in Figure 2.1. Similarly, when the slope of the Y_1 axis is greater than the slope of the line AD, the equal sign in (2.41) must be replaced by a "less than" sign, and the resulting inequality is satisfied by all points to the left of the dashed line in Figure 2.1.

The last three cases depend on the relationship between the slope of the Y_1 axis and the slope of the line AD.

<u>Case 3</u>: $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} > 1$ and $\frac{c_1(n_1 + 2c_1)}{n_1} = \frac{c_2(n_2 + 2c_2)}{n_2}$, i.e., (c_1,c_2) lies outside the ellipse on the dashed line in Figure 2.1, and the slope of the Y_1 axis is equal to the slope of AD, as indicated in Figure 2.2.

When the eigenvalue in (2.35) is positive and all other eigenvalues are negative, the canonical form of the risk function in (2.25) indicates that the risk increases with a decrease in $|Y_2|$. Thus, for any point in the parameter space in Figure 2.2 not on the Y_1 axis, a larger risk may be obtained by moving along a perpendicular dropped to the Y_1 axis. Therefore, the maximum risk must occur on the Y_1 axis.

Since the eigenvalue in (2.35) is positive, a move in either direction along the Y_1 axis away from the origin will result in an increase in the risk. The maximum risk, then, must occur at one of the intersections of the Y_1 axis with the boundaries of the parameter space. Since the slope of the Y_1 axis is equal to the slope of the line AD, these intersections occur at the corners D and F in Figure 2.2. Due to the symmetry of the parameter space and risk function, the risks at D and F are the same. From (2.9), the risk at D, where $(z_1,z_2) = (\frac{1}{2},-\frac{1}{2})$, is

$$R^*(c_1,c_2) = \left[\frac{c_1}{n_1 + 2c_1} + \frac{c_2}{n_2 + 2c_2}\right]^2 . \tag{2.42}$$

The curve $\frac{c_1(n_1 + 2c_1)}{n_1} = \frac{c_2(n_2 + 2c_2)}{n_2}$ is monotone increasing, so $R^*(c_1,c_2)$ can be reduced by decreasing c_1 and c_2 simultaneously, moving along the dashed line in Figure 2.1 to point P. Let (c_1',c_2') equal the coordinates of the point P, and Theorem 2.2 is satisfied.

<u>Case 4</u>: $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} > 1$ and $\frac{c_1(n_1 + 2c_1)}{n_1} > \frac{c_2(n_2 + 2c_2)}{n_2}$; i.e., (c_1,c_2) lies outside the ellipse to the right of the dashed curve in Figure 2.1, and the slope of the Y_1 axis is less than the slope of the line AD, as indicated in Figure 2.3. For this case, the maximum risk occurs somewhere on the boundaries $y_1 = \pm \frac{\sqrt{n_1}}{2(n_1 + 2c_1)}$, at the points of intersection of the Y_1 axis with the boundaries, or between the points of intersection and certain corners of the parameter space. These regions are indicated by the darkened lines CD and EF in Figure 2.3. It is known that the maximum must occur in these regions because, for any point located elsewhere in the parameter space, a larger risk may be

found by moving along a perpendicular dropped to the Y_1 axis, or by moving along, or parallel to, the Y_1 axis toward the boundary. Again, the behavior of the risk is explained by the canonical representation of the risk function in (2.25) when the coefficient of Y_1^2 is positive and the coefficient of Y_2^2 is negative.

Because of the symmetry of the risk about the canonical axis, every point in the interval CD has the same risk as its image in EF. Thus, the search for the maximum risk can be restricted to the interval CD.

The exact location of the maximum risk along line CD depends on the sign of the derivative of the risk, with respect to y_2, at the point of intersection of the Y_1 axis with the boundary of the parameter space. The point of intersection, for this case, is point C in Figure 2.3.

The risk, from equation (2.13), is given by

$$R(\hat{\gamma},\underline{y}) = \tfrac{1}{4}\sum \frac{n_i\,\alpha_i^2}{(n_i + 2c_i)^2} + \underline{y}'\,A\underline{y}$$

$$= \tfrac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] + y_1^2(v_1^2 - 1)$$

$$+ y_2^2(v_2^2 - 1) + 2y_1y_2v_1v_2 \quad , \qquad (2.43)$$

where $v_i = 2c_i/\sqrt{n_i}$. Therefore, along the right-hand boundary of the parameter space,

$$\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_2} = 2y_2(v_2^2 - 1) + 2y_1 v_1 v_2 \Bigg|_{y_1 = \frac{\sqrt{n_1}}{2(n_1+2c_1)}}$$

$$= 2y_2(v_2^2 - 1) + \frac{v_1 v_2 \sqrt{n_1}}{n_1 + 2c_1} \quad . \tag{2.44}$$

At the point of intersection C, $y_1 = \frac{\sqrt{n_1}}{2(n_1 + 2c_1)}$, and $y_2 = \frac{v_2}{v_1} \frac{\sqrt{n_1}}{2(n_1 + 2c_1)}$. Thus,

$$\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_2}\Bigg|_C = \frac{v_2}{v_1} \frac{\sqrt{n_1}}{(n_1 + 2c_1)} (v_2^2 - 1) + \frac{v_1 v_2 \sqrt{n_1}}{(n_1 + 2c_1)}$$

$$= \frac{v_2}{v_1} \frac{\sqrt{n_1}}{(n_1 + 2c_1)} \left[v_2^2 - 1 + v_1^2\right]$$

$$= \frac{v_2}{v_1} \frac{\sqrt{n_1}}{(n_1 + 2c_1)} \left[\sum_{i=1}^{2} v_i^2 - 1\right] \quad . \tag{2.45}$$

For this case, $\sum_{i=1}^{2} v_i^2 = \frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} > 1$, and the derivative in (2.45) is positive. A positive value for the derivative at C indicates that the maximum risk in CD occurs at D, or between C and D, but not at C. The location of the maximum can now be pinpointed by observing the sign of the derivative of the risk, with respect to y_2, at the corner, D. The expression for the derivative at D is found by evaluating (2.44) at $y_2 = \frac{\sqrt{n_2}}{2(n_2 + 2c_2)}$.

$$\left.\frac{\partial R(\hat{\gamma},y)}{\partial y_2}\right|_D = \frac{\sqrt{n_2}}{n_2 + 2c_2}(v_2^2 - 1) + \frac{v_1 v_2 \sqrt{n_1}}{n_1 + 2c_1}$$

$$= \frac{\sqrt{n_2}}{n_2 + 2c_2}\left(\frac{4c_2^2}{n_2} - 1\right) + \frac{4c_1 c_2}{\sqrt{n_2}\,(n_1 + 2c_1)} \quad . \qquad (2.46)$$

The location of the maximum risk is now defined, depending on which of the following two cases hold:

<u>Case 4a</u>: $\left.\frac{\partial R(\gamma,y)}{\partial y_2}\right|_D < 0$. The maximum risk in this case occurs on the boundary $y_1 = \frac{\sqrt{n_1}}{2(n_1 + 2c_1)}$ between the intersection of the Y_1 axis (point C) and the corner (point D).

The location of the maximum risk can be found by observing in (2.43) that, for fixed y_1, the risk is a quadratic function in y_2. Thus, at $y_1 = \frac{\sqrt{n_1}}{2(n_1 + 2c_1)}$,

$$R(\hat{\gamma},y_2) = \frac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] + \frac{1}{4}\frac{n_1}{(n_1 + 2c_1)^2}(v_1^2 - 1)$$

$$+ \frac{\sqrt{n_1}}{n_1 + 2c_1} v_1 v_2 y_2 + (v_2^2 - 1)y_2^2 \quad . \qquad (2.47)$$

In general, if y is a quadratic function of x, i.e.,

$$y = b_0 + b_1 x + b_2 x^2 \quad ,$$

then the stationary point, x_0, is given by

$$x_o = -\frac{b_1}{2b_2} \quad .$$

At that stationary point, the value of y is

$$y_o = b_o + b_1\left(\frac{-b_1}{2b_2}\right) + b_2\left(\frac{b_1^2}{4b_2^2}\right)$$

$$= b_o - \frac{b_1^2}{4b_2} \quad .$$

In (2.47), the risk is a quadratic function with

$$b_o = \tfrac{1}{4}\left[\frac{n_1 v_1^2}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] ,$$

$$b_1 = \frac{\sqrt{n_1}\, v_1 v_2}{n_1 + 2c_1} , \quad \text{and}$$

$$b_2 = (v_2^2 - 1) \quad .$$

The maximum risk occurs at

$$y_2 = \frac{\dfrac{-\sqrt{n_1}\, v_1 v_2}{n_1 + 2c_1}}{2(v_2^2 - 1)}$$

and has value

$$R^*(c_1, c_2) = \tfrac{1}{4}\left[\frac{n_1 v_1^2}{(n_1 + 2c_1)^2} + \frac{n_2}{(n_2 + 2c_2)^2}\right] - \frac{\dfrac{n_1 v_1^2 v_2^2}{(n_1 + 2c_1)^2}}{4(v_2^2 - 1)}$$

$$= \tfrac{1}{4}\left[\frac{n_2}{(n_2 + 2c_2)^2} - \frac{n_1 v_1^2}{(n_1 + 2c_1)^2(v_2^2 - 1)}\right]$$

$$= \tfrac{1}{4}\left[\frac{n_2}{(n_2 + 2c_2)^2} - \frac{4c_1^2}{(n_1 + 2c_1)^2\left(\dfrac{4c_2^2}{n_2} - 1\right)}\right] \quad . \qquad (2.48)$$

Reiterating the restrictions imposed for Case 4a, the following conditions must hold:

Condition 1: $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} > 1$,

Condition 2: $\frac{c_1(n_1 + 2c_1)}{n_1} > \frac{c_2(n_2 + 2c_2)}{n_2}$,

Condition 3: $\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_2}\right|_D < 0$.

Condition 3 and expression (2.46) imply that $\frac{4c_2^2}{n_2} - 1 < 0$. Thus, the second term in (2.48) is positive. Since this term is monotone increasing in c_1, $R^*(c_1,c_2)$ may be reduced by decreasing c_1. Condition 3 will continue to be satisfied (see (2.46)), but either the ellipse or dashed line will eventually be encountered. If the ellipse is encountered, let (c_1',c_2') equal the coordinates of the point of encounter, and Theorem 2.2 is satisfied. If the dashed line is encountered, Case 3 applies, and the theorem is again satisfied.

<u>Case 4b</u>: $\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_2}\right|_D \geq 0$. If (c_1,c_2) satisfies conditions 1 and 2 of Case 4a, but not condition 3, then the maximum risk again occurs at the corner of the parameter space, D. If $\left.\frac{\partial R}{\partial y_2}\right|_D = 0$, the risk has a stationary point, a maximum, at D. If $\left.\frac{\partial R}{\partial y_2}\right|_D > 0$, then, because $\left.\frac{\partial R}{\partial y_2}\right|_C > 0$ and because the risk is a quadratic function of y_2 for fixed y_1, the risk must increase along the boundary from C to D.

The maximum risk at point D is, again,

$$R^*(c_1,c_2) = \left[\frac{c_1}{n_1 + 2c_1} + \frac{c_2}{n_2 + 2c_2}\right]^2 .$$

The maximum risk may be reduced by decreasing c_1 in C-space until one of the following occurs:

(a) the ellipse is encountered, Case 1 holds, and the theorem is satisfied;

(b) the dashed line is encountered, Case 3 holds, and the theorem is satisfied;

(c) Condition 3 is again satisfied, Case 4a holds, and the theorem is satisfied.

<u>Case 5</u>: $\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} > 1$ and $\frac{c_1(n_1 + 2c_1)}{n_1} < \frac{c_2(n_2 + 2c_2)}{n_2}$; i.e., (c_1,c_2) lies outside the ellipse to the left of the dashed curve in Figure 2.1, and the slope of the Y_1 axis is greater than the slope of the line AD, as indicated in Figure 2.4. As in Case 4, the maximum risk is known to occur along the Y_1 axis, or in the shaded intervals CD and EF in Figure 2.4. For a point (c_1,c_2) in any other region, a point of higher risk may be found by moving along a perpendicular dropped to the Y_1 axis, or in a direction parallel to the Y_1 axis, away from the center. Because of symmetry, the search for the maximum risk may again be restricted to the interval CD.

The proof that Theorem 2.2 is true for points (c_1,c_2) satisfying the conditions in Case 5 is almost identical to the proof for Case 4. For this reason, details will be omitted from the proof for Case 5, when the analogy to Case 4 is clear.

The location of the maximum is again dependent on the derivative of the risk at the point of intersection, C. At C, $y_2 = \frac{\sqrt{n_2}}{2(n_2 + 2c_2)}$,

$$y_1 = \frac{v_1}{v_2} \; \frac{\sqrt{n_2}}{2(n_2 + 2c_2)} \text{ , and}$$

$$\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_1}\right|_C = \frac{v_1}{v_2} \; \frac{\sqrt{n_2}}{(n_2 + 2c_2)} \left[v_1^2 + v_2^2 - 1\right] .$$

This derivative is positive under the conditions of this case, so the maximum risk occurs between C and D, or at D, depending on the sign of the derivative of the risk, with respect to y_1, at D. This partial derivative is given by

$$\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_1}\right|_D = \frac{\sqrt{n_1}}{n_1 + 2c_1}\left(\frac{4c_1^2}{n_1} - 1\right) + \frac{4c_1 c_2}{\sqrt{n_1}\,(n_2 + 2c_2)} . \tag{2.49}$$

<u>Case 5a</u>: $\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_1}\right|_D < 0$. The maximum risk occurs at

$$y_1 = \frac{\dfrac{-\sqrt{n_2}\; v_1 v_2}{n_2 + 2c_2}}{2(v_1^2 - 1)}$$

and has value

$$R^*(c_1,c_2) = \frac{1}{4}\left[\frac{n_1}{(n_1 + 2c_1)^2} - \frac{4c_2^2}{(n_2 + 2c_2)^2\left(\dfrac{4c_1^2}{n_1} - 1\right)}\right] . \tag{2.50}$$

The conditions imposed by this case are

Condition 1': $\dfrac{4c_1^2}{n_1} + \dfrac{4c_2^2}{n_2} > 1$;

Condition 2': $\dfrac{c_1(n_1 + 2c_1)}{n_1} < \dfrac{c_2(n_2 + 2c_2)}{n_2}$;

Condition 3': $\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_1}\right|_D < 0$.

From (2.49) if Condition 3' holds (along with Conditions 1' and 2'), then $\frac{4c_1^2}{n_1} - 1 < 0$, and $R^*(c_1,c_2)$ in (2.50) may be decreased by decreasing c_2 until the ellipse or the dashed line is encountered. If the former, Case 1 holds and the theorem is satisfied. If the dashed line is encountered, Case 3 holds and the theorem is satisfied.

Case 5b: $\left.\frac{\partial R(\hat{\gamma},\underline{y})}{\partial y_1}\right|_D \geq 0$. If conditions 1' and 2' hold, but not 3', then the maximum risk occurs at the corner of the parameter space (point D), where the risk is

$$R^*(c_1,c_2) = \left[\frac{c_1}{n_1 + 2c_1} + \frac{c_2}{n_2 + 2c_2}\right]^2 .$$

This maximum risk may be reduced by decreasing c_2 until

(a) the ellipse is encountered, Case 1 holds, and the theorem is true;

(b) the dashed line is encountered, Case 3 holds, and the theorem is true;

(c) Condition 3' is again satisfied, Case 5a holds, and the theorem is true.

It has now been shown that Theorem 2.2 is true if (c_1,c_2) satisfies any of the Cases 1-5. Since there is no point (c_1,c_2), $c_1 \geq 0$, $c_2 \geq 0$, which does not satisfy one of these cases, the theorem is proved.

2.2 Estimates Which Minimize the Maximum Weighted Risk

In this section estimates are derived that minimize the maximum weighted risk, when the risk function is weighted by any member of the class of symmetric beta functions on Θ. Again, the parameter of interest is $\gamma = \sum_{i=1}^{k} \alpha_i \theta_i$, the loss is defined to be squared error, and the estimators of γ considered are those in the class of SBP estimators, given by expression (2.8).

2.2.1 Derivation of the Weighted Risk, $\tilde{R}(\xi,\omega)$

Ω, the class of weighting functions, $\omega(\underline{\theta})$, defined on the k-dimensional space Θ, is restricted to functions of the form

$$\omega(\underline{\theta}) = \prod_{i=1}^{k} \omega_i(\theta_i)$$

where $\omega_i(\theta_i)$, the marginal weighting function of θ_i, is a symmetric beta function of the form

$$\omega_i(\theta_i) = \frac{\theta_i^{b_i}(1-\theta_i)^{b_i}}{B(b_i+1, b_i+1)} \quad , \quad b_i > -1 \; . \tag{2.51}$$

Recall from equation (2.9) that the risk function for an SBP estimator of γ, under squared error loss, may be written

$$R(\hat{\gamma}, \underline{z}) = \left[\sum \frac{2c_i\alpha_i z_i}{n_i + 2c_i}\right]^2 + \sum \frac{n_i\alpha_i^2}{(n_i + 2c_i)^2}\left(\tfrac{1}{4} - z_i^2\right) \; . \tag{2.52}$$

When this risk function is weighted by the weighting function $\omega(\underline{\theta})$, the weighted risk is given by

$$\widetilde{R}(\xi,\omega) = \frac{\int_0^1 \cdots \int_0^1 R(\hat{\gamma},\underline{z}) \prod_{i=1}^{k} \theta_i^{b_i} (1-\theta_i)^{b_i} \, d\theta_i}{\prod_{i=1}^{k} B(b_i+1, b_i+1)} .$$

It is convenient in what follows to view $\omega(\underline{\theta})$ as a probability density function which is in fact the product of the k independent densities given in (2.51). Thus $\widetilde{R}(\xi,\omega)$ can be written as $\mathcal{E}_\omega[R(\hat{\gamma},\underline{z})]$, the expectation of $R(\hat{\gamma},\underline{z})$. Since $\omega_i(\theta_i)$ is a symmetric beta function, $\mathcal{E}_\omega(\theta_i) = \frac{1}{2}$ and $\mathcal{E}_\omega(z_i) = \mathcal{E}_\omega(\theta_i - \frac{1}{2}) = 0$. $\mathcal{E}_\omega(z_i^2) = \mathcal{E}_\omega(\theta_i^2 - \theta_i + \frac{1}{4}) = 1/4(2b_i + 3)$.

The independence of θ_i, θ_j implies that

$$\mathcal{E}_\omega(z_i z_j) = \mathcal{E}(z_i)\,\mathcal{E}(z_j) = 0 ,$$

and

$$\mathcal{E}_\omega \left[\sum \frac{c_i \alpha_i z_i}{n_i + 2c_i}\right]^2 = \sum \frac{c_i^2 \alpha_i^2 \, \mathcal{E}_\omega(z_i^2)}{(n_i + 2c_i)^2} .$$

Thus

$$\mathcal{E}_\omega[R(\hat{\gamma},\underline{z})] = \left\{ \sum \left[\frac{4c_i^2 \alpha_i^2}{(n_i + 2c_i)^2} \cdot \frac{1}{4(2b_i + 3)}\right] + \frac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} - \sum \left[\frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} \cdot \frac{1}{4(2b_i + 3)}\right] \right\}$$

and

$$\widetilde{R}(\xi,\omega) = \mathcal{E}_\omega[R(\hat{\gamma},\underline{z})] = \sum \frac{\alpha_i^2 (4c_i^2 - n_i)}{4(n_i+2c_i)^2(2b_i+3)} + \frac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i+2c_i)^2} . \quad (2.53)$$

Since the prior density, $\xi(\underline{\theta})$, is specified by the vector $\underline{c}$ of constants $c_i = \alpha_i + 1$, and the weighting function $\omega(\underline{\theta})$ is specified by the vector $\underline{b}$ of constants b_i, the weighted risk $\widetilde{R}(\xi,\omega)$ may be indexed by the vectors $\underline{c}$ and $\underline{b}$. Because the weighted risk will be analyzed as a function of $\underline{c}$ and $\underline{b}$, it will often be convenient to adopt the notation

$$\widetilde{R}(\underline{c},\underline{b}) = \widetilde{R}(\xi,\omega) \quad . \tag{2.54}$$

2.2.2 Maximum Weighted Risk for Fixed $\underline{c}$

When the estimator $\hat{\gamma}$ is the Bayes estimator derived from a fixed prior $\xi(\underline{\theta})$, the set $\{c_1, c_2, \ldots, c_k\}$ is a set of fixed constants. The maximum value of $\widetilde{R}(\xi,\omega)$ possible, when $\xi(\underline{\theta})$ is fixed, can be obtained by considering expression (2.53) as a function of $\{b_1, b_2, \ldots, b_k\}$, the set of parameters specifying $\omega(\underline{\theta})$.

Note that for each i in the sum (2.53), the corresponding term contains one and only one member of the set $\{b_1, b_2, \ldots, b_k\}$. To find the set $\{b_1^*, b_2^*, \ldots, b_k^*\}$ that maximizes $\widetilde{R}(\underline{c},\underline{b})$, therefore, it is sufficient to maximize each term separately, i.e., maximize

$$\widetilde{R}_i(c_i, b_i) = \frac{\alpha_i^2\,(4c_i^2 - n_i)}{4(n_i+2c_i)^2(2b_i+3)} + \frac{n_i\,\alpha_i^2}{4(n_i + 2c_i)^2}\,, \quad i = 1,2,\ldots,k. \tag{2.55}$$

Now consider the behavior of $\widetilde{R}_i(c_i, b_i)$ as a function of b_i, for the following cases:

(a) $c_i^2 > n_i/4$. In expression (2.55), if $c_i^2 > n_i/4$, then $4c_i^2 - n_i > 0$ and that term is positive. In order to maximize $\widetilde{R}_i(c_i, b_i)$, b_i should be as small as possible, so $b_i^* = -1$. Thus,

$$\sup_{b_i} \tilde{R}_i(c_i, b_i) = \frac{4\alpha_i^2 c_i^2}{4(n_i + 2c_i)^2} \quad . \tag{2.56}$$

(b) $c_i^2 < n_i/4$. For this case, $4c_i^2 - n_i < 0$, and the term in (2.55) that contains b_i is negative. Maximizing $\tilde{R}_i(c_i, b_i)$ for this case requires minimizing the term containing b_i, so that $b_i^* = \infty$. Thus

$$\sup_{b_i} \tilde{R}_i(c_i, b_i) = \tfrac{1}{4} \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} \quad .$$

(c) $c_i^2 = n_i/4$. Here $4c_i^2 - n_i = 0$, so that, for any value of b_i,

$$\sup_{b_i} \tilde{R}_i(c_i, b_i) = \tfrac{1}{4} \frac{n_i \alpha_i^2}{(n_i + 2c_i)^2} \quad .$$

The weighting function that maximizes $\tilde{R}(\xi, \omega)$ for fixed prior $\xi(\underline{\theta})$, denoted by $\omega^*(\underline{\theta})$, is defined by

$$\omega^*(\underline{\theta}) = \prod_{i=1}^{k} \frac{\theta_i^{b_i^*} (1 - \theta_i)^{b_i^*}}{B(b_i^*+1, b_i^*+1)} \quad ,$$

where b_i^*, $i = 1,2,\ldots,k$, is determined by case (a), (b), or (c), depending on the value of c_i.

2.2.3 The Value of $\underline{c}$ That Minimizes the Maximum Weighted Risk

Notice that, in (2.53), if $\tilde{R}(\xi, \omega)$ is evaluated at $\omega(\underline{\theta}) = \omega^*(\underline{\theta})$, then

$$\tilde{R}(\xi, \omega^*) = \tilde{R}(\underline{c}, \underline{b}^*) = \sum \frac{\alpha_i^2 (4c_i^2 - n_i)}{4(n_i+2c_i)^2(2b_i^*+3)} + \tfrac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i+2c_i)^2} \quad . \tag{2.57}$$

To determine the set $\{c_1^*, c_2^*, \ldots, c_k^*\}$ that minimizes $\tilde{R}(\underline{c}, \underline{b}^*)$, it is clear that each term in (2.57) can be minimized separately, i.e., minimize

$$\tilde{R}_i(c_i, b_i^*) = \frac{\alpha_i^2 (4c_i^2 - n_i)}{4(n_i + 2c_i)^2(2b_i^* + 3)} + \frac{1}{4}\frac{n_i\,\alpha_i^2}{(n_i + 2c_i)^2} . \tag{2.58}$$

Consider again the following three cases:

(a) $c_i^2 > n_i/4$. From case (a) of Section 2.2.2

$$\tilde{R}_i(c_i, b_i^*) = \frac{4\alpha_i^2\, c_i^2}{4(n_i + 2c_i)^2}$$

$$= \frac{\alpha_i^2}{\frac{n_i^2}{c_i^2} + \frac{4n_i}{c_i} + 4} .$$

As c_i increases, the denominator decreases, and $\tilde{R}_i(c_i, b_i^*)$ increases.

(b) $c_i^2 < n_i/4$. From case (b) of 2.2.2,

$$\tilde{R}_i(c_i, b_i^*) = \frac{1}{4}\frac{n_i\,\alpha_i^2}{(n_i + 2c_i)^2} .$$

As c_i decreases toward its minimum value of zero, the denominator decreases, and $\tilde{R}_i(c_i, b_i^*)$ increases.

(c) $c_i^2 = n_i/4$. From case (c) of 2.2.2,

$$\tilde{R}_i(c_i, b_i^*) = \frac{1}{4}\frac{n_i\,\alpha_i^2}{(n_i + 2c_i)^2}$$

$$= \frac{1}{4}\frac{n_i\,\alpha_i^2}{(n_i + \sqrt{n_i})^2} .$$

These three cases can be summarized by the graph of $\tilde{R}_i(c_i, b_i^*)$ as a function of c_i, in Figure 2.5, which indicates that $\tilde{R}_i(c_i, b_i^*)$ is minimized when $c_i = \sqrt{n_i}/2$, $i = 1,2,\ldots,k$. Thus, the set $\{c_1^*, c_2^*, \ldots, c_k^*\}$ which minimizes the maximum weighted risk is $\{\sqrt{n_1}/2, \sqrt{n_2}/2, \ldots, \sqrt{n_k}/2\}$.

From (2.53) observe that the value of the weighted risk corresponding to $\{c_1^*, \ldots, c_k^*\}$, for any weighting function $\omega(\underline{\theta})$ in Ω, is

$$\tilde{R}(\underline{c}^*, \underline{b}) = \tfrac{1}{4} \sum \frac{n_i \alpha_i^2}{(n_i + \sqrt{n_i})^2} = \tfrac{1}{4} \sum \frac{\alpha_i^2}{(1 + \sqrt{n_i})^2} .$$

Note that, although the integrated risk is invariant for all weighting functions in Ω, the risk function itself is not constant over Θ. The reason for the invariance of the integrated risk can be seen by writing the risk function (2.52) in matrix form.

$$R(\hat{\gamma}, \underline{z}) = \sum \frac{n_i \alpha_i^2}{4(n_i + 2c_i)^2} +$$

$$\underline{z}' \begin{bmatrix} \frac{(4c_1^2 - n_1)\alpha_1}{(n_1+2c_1)^2} & \frac{4c_1 c_2 \alpha_1 \alpha_2}{(n_1+2c_1)(n_2+2c_2)} & \cdots & \frac{4c_1 c_k \alpha_1 \alpha_k}{(n_1+2c_1)(n_k+2c_k)} \\ \frac{4c_1 c_2 \alpha_1 \alpha_2}{(n_1+2c_1)(n_2+2c_2)} & \frac{(4c_2^2 - n_2)\alpha_2}{(n_2+2c_2)^2} & \cdots & \frac{4c_2 c_k \alpha_2 \alpha_k}{(n_2+2c_2)(n_k+2c_k)} \\ \vdots & & \ddots & \\ \frac{4c_1 c_k \alpha_1 \alpha_k}{(n_1+2c_1)(n_k+2c_k)} & & \cdots & \frac{(4c_k^2 - n_k)\alpha_k}{(n_k+2c_k)^2} \end{bmatrix} \underline{z} .$$

Let $\hat{\gamma}^* = \sum \alpha_i \dfrac{x_i + c_i^*}{n_i + 2c_i^*}$, where $c_i^* = \sqrt{n_i}/2$. Then

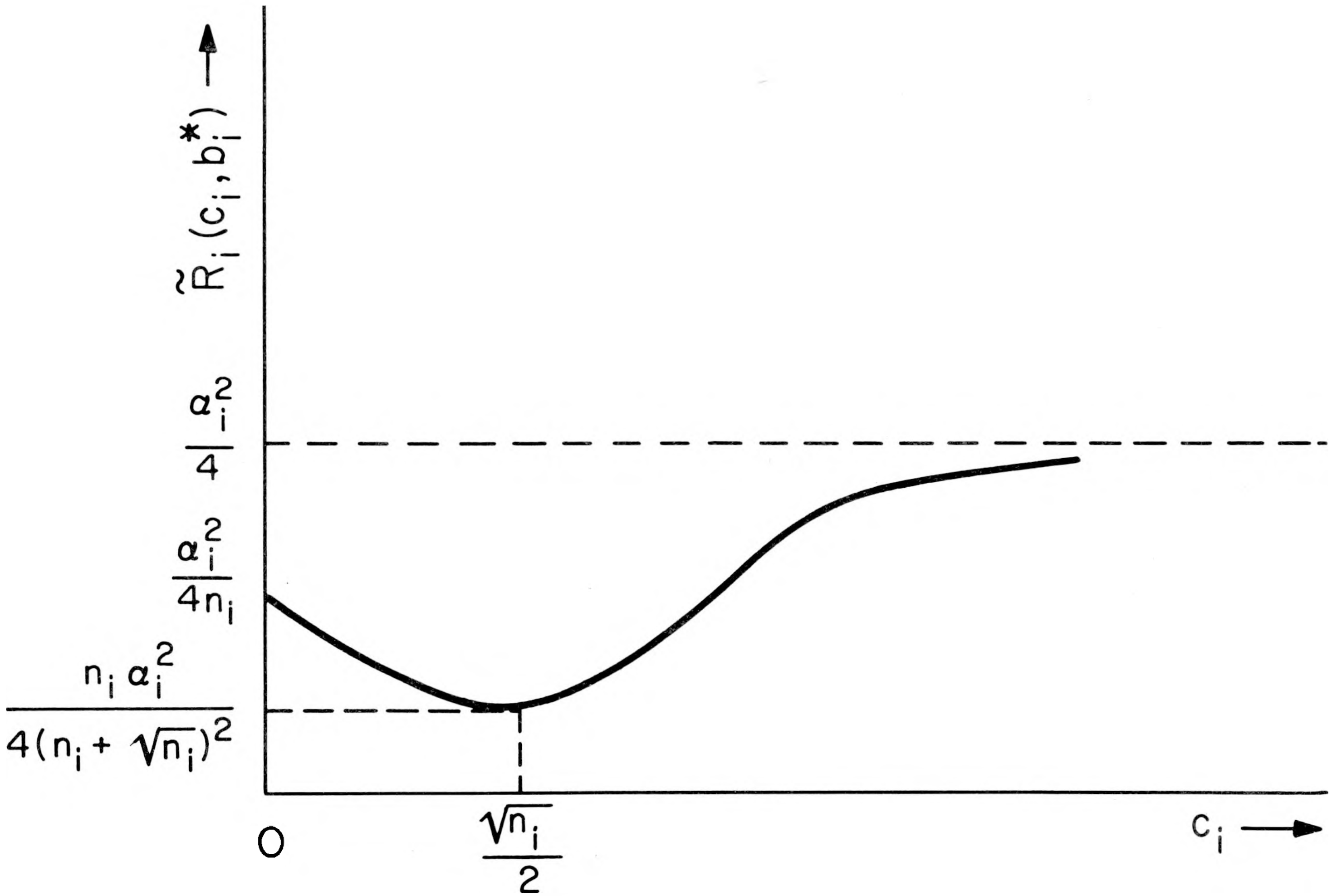

Figure 2.5. Minimization of $\tilde{R}_i(c_i, b_i^*)$

$$R(\hat{\gamma}^*, \underline{z}) = \sum \frac{\alpha_i^2}{4(1+\sqrt{n_i})^2} +$$

$$\underline{z}' \begin{bmatrix} 0 & \frac{\alpha_1\alpha_2}{(1+\sqrt{n_1})(1+\sqrt{n_2})} & \cdots & \frac{\alpha_1\alpha_k}{(1+\sqrt{n_1})(1+\sqrt{n_k})} \\ \frac{\alpha_1\alpha_2}{(1+\sqrt{n_1})(1+\sqrt{n_2})} & 0 & \cdots & \frac{\alpha_2\alpha_k}{(1+\sqrt{n_2})(1+\sqrt{n_k})} \\ \vdots & & \ddots & \\ \frac{\alpha_1\alpha_k}{(1+\sqrt{n_1})(1+\sqrt{n_k})} & \cdots & & 0 \end{bmatrix} \underline{z} ,$$

which is not constant over Θ. However, if $R(\hat{\gamma}^*, \underline{z})$ is weighted by any symmetric beta function, then the last term in $R(\hat{\gamma}^*, \underline{z})$, when weighted and integrated, can be written as a linear combination of expectations $\mathcal{E}(z_i z_j)$ taken over the density defined by the weighting function. These expectations are all zero, as shown in Section 2.2.1, because

(i) $\mathcal{E}_\omega(z_i) = \mathcal{E}_\omega(z_j) = 0$, and

(ii) $\mathcal{E}_\omega(z_i z_j) = \mathcal{E}_\omega(z_i)\, \mathcal{E}_\omega(z_j)$.

In fact, for any weighting function $\omega(\underline{\theta})$ such that $\omega(z_1, \ldots, z_k) = \omega_1(z_1) \ldots \omega_k(z_k)$ where $\mathcal{E}_{\omega_i}(z_i) = 0$, $i = 1,2,\ldots,k$, the integrated risk will be invariant.

2.3 Estimates Corresponding to the Least Favorable Prior

In this section estimates are found that correspond to the "least favorable" prior distribution in Ξ, the class of independent symmetric beta priors. The "least favorable" prior, denoted by $\xi^*(\underline{\theta})$, is the prior in Ξ that maximizes the Bayes risk. The Bayes risk for any prior $\xi(\underline{\theta})$ is found by evaluating the weighted risk $\widetilde{R}(\xi,\omega)$, given in (2.53) at $\omega(\underline{\theta}) = \xi(\underline{\theta})$ (i.e., at $b_i = c_i - 1$).

$$\widetilde{R}(\xi,\xi) = \sum \frac{\alpha_i^2\,(4c_i^2 - n_i)}{4(n_i+2c_i)^2(2c_i+1)} + \tfrac{1}{4}\sum \frac{n_i\,\alpha_i^2}{(n_i+2c_i)^2}$$

$$= \sum \frac{\alpha_i^2(4c_i^2 - n_i) + (2c_i + 1)(n_i\,\alpha_i^2)}{4(n_i + 2c_i)^2\,(2c_i + 1)}$$

$$= \sum \frac{\alpha_i^2\, c_i}{2(n_i + 2c_i)(2c_i + 1)} \quad . \tag{2.59}$$

The least favorable prior, $\xi^*(\underline{\theta})$, can be obtained by finding the values of its parameters, $\{c_1^*, c_2^*, \ldots, c_k^*\}$, that maximize the Bayes risk, $\widetilde{R}(\xi,\xi)$. The value of c_j that maximizes $\widetilde{R}(\xi,\xi)$ is found by setting $\frac{\partial\widetilde{R}(\xi,\xi)}{\partial c_j} = 0$, $j = 1,2,\ldots,k$.

$$\frac{\partial\widetilde{R}(\xi,\xi)}{\partial c_j} = \frac{(2n_j + \cancel{4n_jc_j} + \cancel{4c_j} + 8c_j^2)\alpha_j^2 - \alpha_j^2 c_j(\cancel{4n_j} + \cancel{4} + 16c_j)}{4(n_j + 2c_j)^2\,(2c_j + 1)^2}$$

$$= \frac{(2n_j - 8c_j^2)\,\alpha_j^2}{4(n_j + 2c_j)^2\,(2c_j + 1)^2} = 0$$

$$\Rightarrow c_j^{*2} = n_j/4 \ .$$

Since $\frac{\partial\widetilde{R}(\xi,\xi)}{\partial c_j}$ is greater than zero for $c_j^2 < n_j/4$ and less than zero for $c_j^2 > n_j/4$, it is clear that the stationary point at $c_j^2 = n_j/4$ is a maximum and not a minimum. Therefore, the least favorable prior, $\xi^*(\underline{\theta})$, is given by

$$\xi^*(\underline{\theta}) = \prod_{i=1}^{k} \frac{\theta_i^{c_i^*-1}(1-\theta_i)^{c_i^*-1}}{B(c_i^*, c_i^*)} ,$$

where $c_i^* = \sqrt{n_i}/2$, for $i = 1,2,\ldots,k$.

The value of the Bayes risk for the least favorable prior is found by evaluating (2.59) at $c_i = \sqrt{n_i}/2$, for $i = 1,2,\ldots,k$. Notice that, since the $\{c_i^*\}$ corresponding to the least favorable prior is the same set $\{c_j^*\}$ that minimizes the maximum weighted risk in Section 2.2, the Bayes risk here is the same as that minimax weighted risk, i.e.,

$$\tilde{R}(\xi^*, \xi^*) = \tfrac{1}{4} \sum \frac{\alpha_i^2}{(1 + \sqrt{n_i})^2} .$$

2.4 Summary of Results

For the criteria discussed in Chapter I, optimum SBP estimators of the form

$$\hat{\gamma} = \sum_{i=1}^{k} \alpha_i \left(\frac{x_i + c_i}{n_i + 2c_i} \right) ,$$

where $\gamma = \sum_{i=1}^{k} \alpha_i \theta_i$ have been found. They can be categorized according to the optimum values of $\{c_i\}$, as follows.

In Section 2.1, the general form of the risk function for an SBP estimator $\hat{\gamma}$ was discussed. For two special cases, estimators were found that produced the minimax risk in the class C of SBP estimators.

When $|\alpha_i| = 1$ and $n_i = n$, for $i = 1,2,\ldots,k$, the estimator $\hat{\gamma}$ based on the set of constants $\{c_i = \sqrt{\frac{n}{4k}}\}$ is C-minimax and universal minimax. When $k = 2$, $\alpha_1 = -\alpha_2 = 1$, and $n_1 \neq n_2$, then the SBP estimator $\hat{\gamma}$ based on the constants $\{c_1, c_2\}$, obtained as the solution to the set of equations

$$\frac{c_1(n_1 + 2c_1)^3}{n_1^2} = \frac{c_2(n_2 + 2c_2)^3}{n_2^2}$$

$$\frac{4c_1^2}{n_1} + \frac{4c_2^2}{n_2} = 1 \ ,$$

is C-minimax.

In Section 2.2, it was found that estimates with $\{c_i = \sqrt{\frac{n_i}{4}}\}$ minimized the maximum weighted risk, when the risk function was weighted by any member of the class of independent symmetric beta functions defined on Θ. For $\{c_i = \sqrt{\frac{n_i}{4}}\}$, it was determined that the weighted risk had value $\frac{1}{4}\sum_{i=1}^{k} \frac{\alpha_i^2}{(1 + \sqrt{n_i})^2}$, whatever the weighting function in this class.

The values of $\{c_i\}$ corresponding to the least favorable symmetric beta prior were found in Section 2.3 to again be $\{c_i = \sqrt{n_i}/4\}$. The Bayes risk for the least favorable prior is also $\frac{1}{4}\sum_{i=1}^{k} \frac{\alpha_i^2}{(1 + \sqrt{n_i})^2}$.

The results of this chapter suggest that optimum estimators of a linear combination of binomial probabilities, $\gamma = \sum \alpha_i \theta_i$, can be obtained by using a linear combination of simple transformations of the maximum likelihood estimators of the θ_i, x_i/n_i. The transformations performed on the maximum likelihood estimators are determined by the constants $\{c_i\}$, and these constants may be chosen by the experimenter to produce certain desirable results. If the worth of an estimator is judged by the behavior of its risk function, then the results in Section 2.1 allow the experimenter to control, to some extent, the behavior of the risk function by the proper choice of $\{c_i\}$. In particular, for two special cases, he may minimize the maximum value of the risk by choosing certain c_i. If the worth of an estimator is measured by weighted risk, then the maximum weighted risk may be minimized by using the c_i developed in Section 2.2.

APPENDIX 2A

Let B be a k-dimensional square matrix of the form

$$B = \begin{bmatrix} 1-\beta_1 & 1 & 1 & \cdots & 1 \\ 1 & 1-\beta_2 & 1 & \cdots & 1 \\ \vdots & & & & \\ 1 & 1 & \cdots & & 1-\beta_k \end{bmatrix} \tag{2.60}$$

where $\beta_i \neq 0$, $i = 1,2,\ldots,k$. The determinant of B, denoted $|B|$, may be derived as follows.

First subtract the first row from each of the others in turn, leaving the determinant unchanged; then partition as shown.

$$|B| = \left|\begin{array}{c|cccc} 1-\beta_1 & 1 & 1 & \cdots & 1 \\ \hline \beta_1 & -\beta_2 & 0 & \cdots & 0 \\ \beta_1 & 0 & -\beta_3 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \beta_1 & 0 & \cdots & & -\beta_k \end{array}\right| .$$

In general, the determinant of the partitioned matrix

$$\left|\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right|$$

may be written (see [10], Theorem 8.2.1) as

$$|B_{22}| \cdot |B_{11} - B_{12} B_{22}^{-1} B_{21}| .$$

Hence

$$|B| = (-1)^{k-1} \beta_2\beta_3\cdots\beta_k \left\{(1 - \beta_1) + (1,1,\ldots,1)\begin{bmatrix} \frac{1}{\beta_2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\beta_3} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \frac{1}{\beta_k} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_1 \end{bmatrix}\right\}$$

$$= (-1)^{k-1} \beta_2\cdots\beta_k \left[(1 - \beta_1) + \frac{\beta_1}{\beta_2} + \frac{\beta_1}{\beta_3} + \cdots + \frac{\beta_1}{\beta_k}\right]$$

$$|B| = (-1)^{k-1} \beta_1\beta_2\cdots\beta_k \left[\frac{1}{\beta_1} + \frac{1}{\beta_2} + \cdots + \frac{1}{\beta_k} - 1\right] . \qquad (2.61)$$

CHAPTER III. FIXED PRECISION ESTIMATION OF THE BINOMIAL PARAMETER θ

3.1 Derivation and Evaluation of Estimates

In this chapter, estimators of the binomial parameter θ will be developed for the binary loss function

$$L(\hat{\theta}(x),\theta) = \begin{cases} 0 & \text{if } \hat{\theta}(x) - \Delta/2 \le \theta \le \hat{\theta}(x) + \Delta/2 \\ 1 & \text{if } \theta > \hat{\theta}(x) + \Delta/2 \text{ or if } \theta < \hat{\theta}(x) - \Delta/2 \end{cases} \tag{3.1}$$

where $0 \le \Delta \le 1$. That is, the estimate $\hat{\theta}$ is considered "right" if $\hat{\theta}$ is sufficiently close to θ (i.e., if $|\hat{\theta}-\theta| \le \Delta/2$) and "wrong" otherwise. For this loss function, the risk (expected loss) is given by

$$R(\hat{\theta},\theta) = \sum_{x=0}^{n} L(\hat{\theta}(x),\theta)\; p(x|\theta)$$

$$= \sum_{x=0}^{n} L(\hat{\theta}(x),\theta)\; \binom{n}{x}\, \theta^x\, (1-\theta)^{n-x}$$

$$= \sum_{D} \binom{n}{x}\, \theta^x\, (1-\theta)^{n-x} \tag{3.2}$$

where $D = \{x|\hat{\theta}(x) > \theta + \Delta/2 \text{ or } \hat{\theta}(x) < \theta - \Delta/2\}$. Note that this is simply the probability that the estimate $\hat{\theta}$ is "wrong."

As in Chapter I, the object of this chapter is to determine the "best" SBP estimates for θ, for the following three definitions of "best" estimates:

(i) Those estimates that minimize the maximum risk;

(ii) Those estimates that minimize the maximum weighted risk;

(iii) Those estimates that are derived from the "least favorable" symmetric beta prior.

3.1.1 Derivation of Bayes Estimates

For the problem of finding point estimates of θ that are optimal, in some sense, for the risk function defined by (3.2), the class of estimates considered was the class C of Bayes estimates derived from symmetric beta prior (SBP) distributions on Θ defined by

$$\xi(\theta) = \frac{\theta^{a}(1-\theta)^{a}}{B(a+1, a+1)}, \quad a > -1 . \tag{3.3}$$

The conditional distribution of the random variable X is binomial with parameters n and θ; i.e.,

$$p(x|\theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x} . \tag{3.4}$$

The Bayes estimator for θ, when the prior distribution of θ is $\xi(\theta)$, was defined in Section 1.2 to be the estimator that minimizes the weighted risk, when the risk function is weighted by the prior $\xi(\theta)$; i.e.,

$$\hat{\theta}_{\xi} = \inf_{\hat{\theta}} \int R(\hat{\theta}, \theta)\, \xi(\theta)\, d\theta .$$

Since the risk function can be expressed as

$$R(\hat{\theta}, \theta) = \sum_{x=0}^{n} L(\hat{\theta}(x), \theta)\, p(x|\theta) ,$$

the weighted risk is equal to

$$\int_{\Theta} \left[\sum_{x=0}^{n} L(\hat{\theta}(x), \theta)\, p(x|\theta) \right] \xi(\theta)\, d\theta . \tag{3.5}$$

But this may also be written as

$$\sum_{x=0}^{n} \left[\int_{\Theta} L(\hat{\theta}(x), \theta)\ \xi(\theta|x)\ d\theta \right] p(x) \quad , \tag{3.6}$$

where $p(x) = \int_{\Theta} f(x, \theta)\ d\theta = \int_{\Theta} p(x|\theta)\ \xi(\theta)\ d\theta$. Hence,

$$\inf_{\hat{\theta}} \int_{\Theta} R(\hat{\theta}, \theta)\ \xi(\theta)\ d\theta = \inf_{\hat{\theta}} \sum_{x=0}^{n} \left[\int_{\Theta} L(\hat{\theta}(x), \theta)\ \xi(\theta|x)\ d\theta \right] p(x)$$

$$= \sum_{x=0}^{n} \left[\inf_{\hat{\theta}(x)} \int_{\Theta} L(\hat{\theta}(x), \theta)\ \xi(\theta|x)\ d\theta \right] p(x) \quad ,$$

and the Bayes estimator, $\hat{\theta}_\xi$, when the prior distribution of θ is $\xi(\theta)$, is equal to the set of estimates that minimize the posterior expected loss; i.e.,

$$\hat{\theta}_\xi = \left\{ \hat{\theta}_\xi(x) \,\middle|\, \hat{\theta}_\xi(x) = \inf_{\hat{\theta}(x)} \int_{\Theta} L(\hat{\theta}(x), \theta)\ \xi(\theta|x)\ d\theta\ , \quad x = 0, 1, \ldots, n \right\} . \tag{3.7}$$

When the prior distribution of θ is the symmetric beta function given in (3.3) and the conditional distribution of X is binomial, as in (3.4), then the posterior distribution of θ, $\xi(\theta|x)$, is defined as

$$\xi(\theta|x) = \frac{\xi(\theta)\ p(x|\theta)}{\int_{\Theta} \xi(\theta)\ p(x|\theta)\ d\theta}$$

$$= \frac{\theta^{a+x}\ (1 - \theta)^{a+n-x}}{B(a + x + 1, a + n - x + 1)} \quad . \tag{3.8}$$

For the loss function defined in (3.1), the posterior expected loss for an estimate, $\theta(x)$, is given by

$$\mathcal{E}_{(\theta|x)}[L(\hat{\theta}(x),\theta)] = \int_0^1 \frac{L(\hat{\theta}(x),\theta)\ \theta^{a+x}\ (1-\theta)^{a+n-x}\ d\theta}{B(a+x+1, a+n-x+1)}$$

$$= \frac{1}{B(a+x+1, a+n-x+1)} \left\{ \int_0^{\hat{\theta}(x)-\Delta/2} \theta^{a+x}(1-\theta)^{a+n-x} d\theta + \int_{\hat{\theta}(x)+\Delta/2}^{1} \theta^{a+x}(1-\theta)^{a+n-x} d\theta \right\} . \tag{3.9}$$

The posterior expected loss in (3.9) may also be written as

$$\mathcal{E}_{(\theta|x)}[L(\hat{\theta}(x),\theta)] = 1 - \int_{\hat{\theta}(x)-\Delta/2}^{\hat{\theta}(x)+\Delta/2} \xi(\theta|x)\ d\theta \tag{3.10}$$

where $\xi(\theta|x)$ is the posterior density in (3.8). Because the general shape of the posterior density, $\xi(\theta|x)$, changes if either $(a + x)$ or $(a + n - x)$ is negative, minimization of the posterior expected loss with respect to $\hat{\theta}(x)$ must be conducted separately for the following two cases.

<u>Case 1</u>: $a + x > 0$, $a + n - x > 0$. This condition is fulfilled if $1 \le x \le n-1$ or if $a > 0$. The beta density with parameters $(a + x, a + n - x)$ increases from zero at $\theta = 0$ to a maximum at $\theta = (x+a)/(2a+n)$ (Lindley [17], p. 143), and decreases again to zero at $\theta = 1$. This curve is shown in Figure 3.1, where the posterior expected loss (3.9) is represented by the shaded area. Clearly, the value of $\hat{\theta}(x)$ that minimizes the posterior expected loss is the midpoint of the interval of width Δ containing the most probability in the posterior density, $\xi(\theta|x)$. This value of $\hat{\theta}_{\xi}(x)$ can be found by setting the derivative of (3.9), with respect to $\hat{\theta}(x)$, equal to zero.

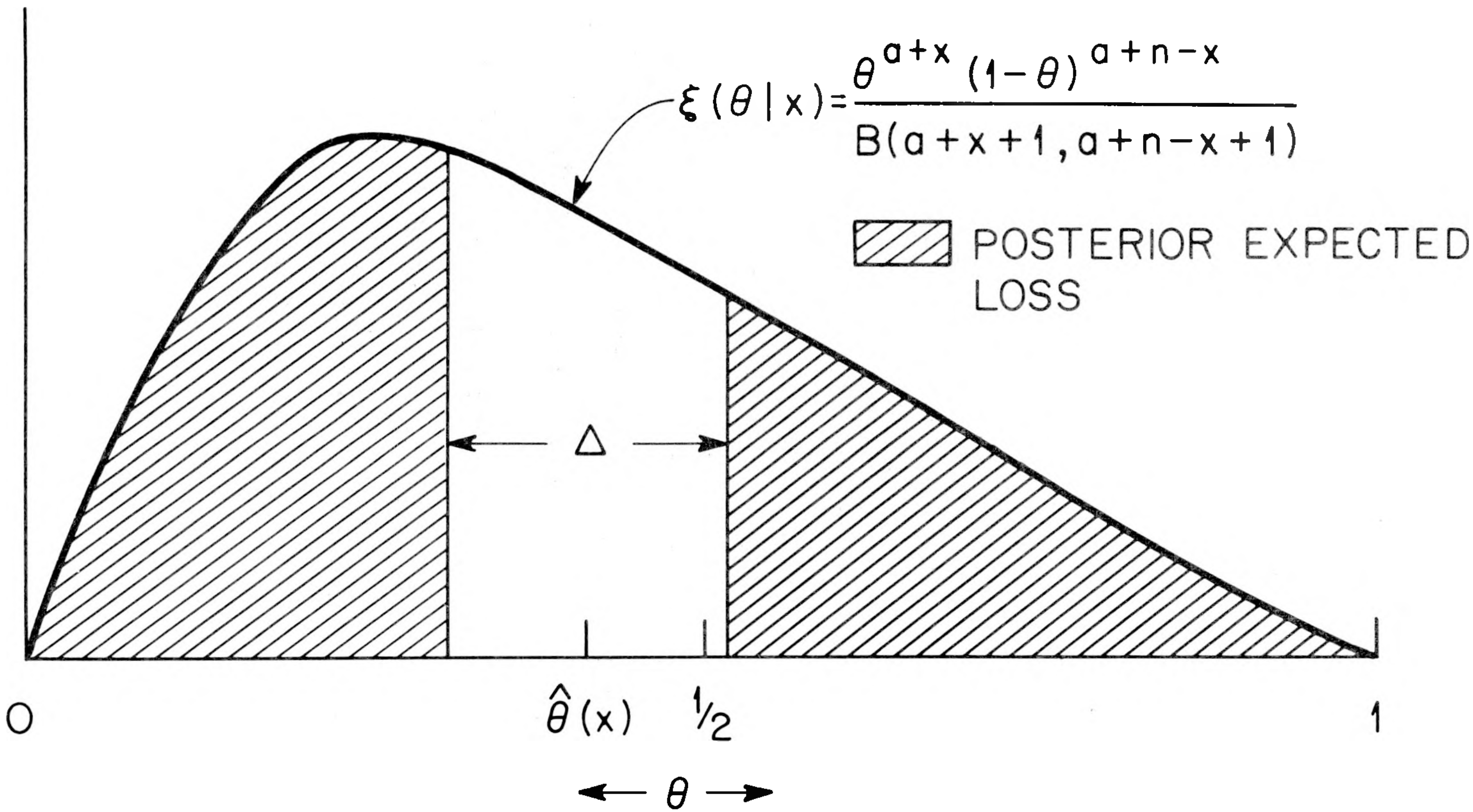

Figure 3.1. Posterior Density of θ: Case 1

$$\frac{\partial\, \mathcal{E}_{(\theta|x)}[L(\hat{\theta}(x),\theta)]}{\partial\, \hat{\theta}(x)} = \xi(\hat{\theta}(x) - \Delta/2|x) - \xi(\hat{\theta}(x) + \Delta/2|x) \quad .$$

Setting this equal to zero, we arrive at the equation

$$\xi(\hat{\theta}(x) - \Delta/2|x) = \xi(\hat{\theta}(x) + \Delta/2|x) \quad . \tag{3.11}$$

For this case, then, the Bayes estimate of θ, when the binomial observation is x, is the midpoint of the interval of width Δ whose endpoints are the values of θ having equal ordinates in the posterior density function. This is represented graphically in Figure 3.2. It is interesting to note that the same interval can be derived as the shortest interval containing a fixed probability in the posterior density (see [15], p. 237).

Case 2: $x = 0$ or n and $a \le 0$. For this case, one of the parameters of the beta function in expression (3.9) for the posterior expected loss is negative, and the beta density is either monotone increasing or decreasing. For example, if $x = 0$ and $a \le 0$, then the posterior density in (3.8) decreases as θ increases. If $a < 0$, then the maximum at $\theta = 0$ is infinitely large. Figure 3.3 shows the graphical representation of the posterior expected loss for this situation.

If $x = n$ and $a \le 0$, then the posterior density increases from zero at $\theta = 0$ to a maximum at $\theta = 1$, that maximum being infinitely large if $a < 0$. The graphical representation of the posterior expected loss for this situation is presented in Figure 3.4.

In these two situations, the value of $\hat{\theta}(x)$ that minimizes the posterior expected loss is again equal to the midpoint of the interval of width Δ that contains the most probability in the posterior density. If

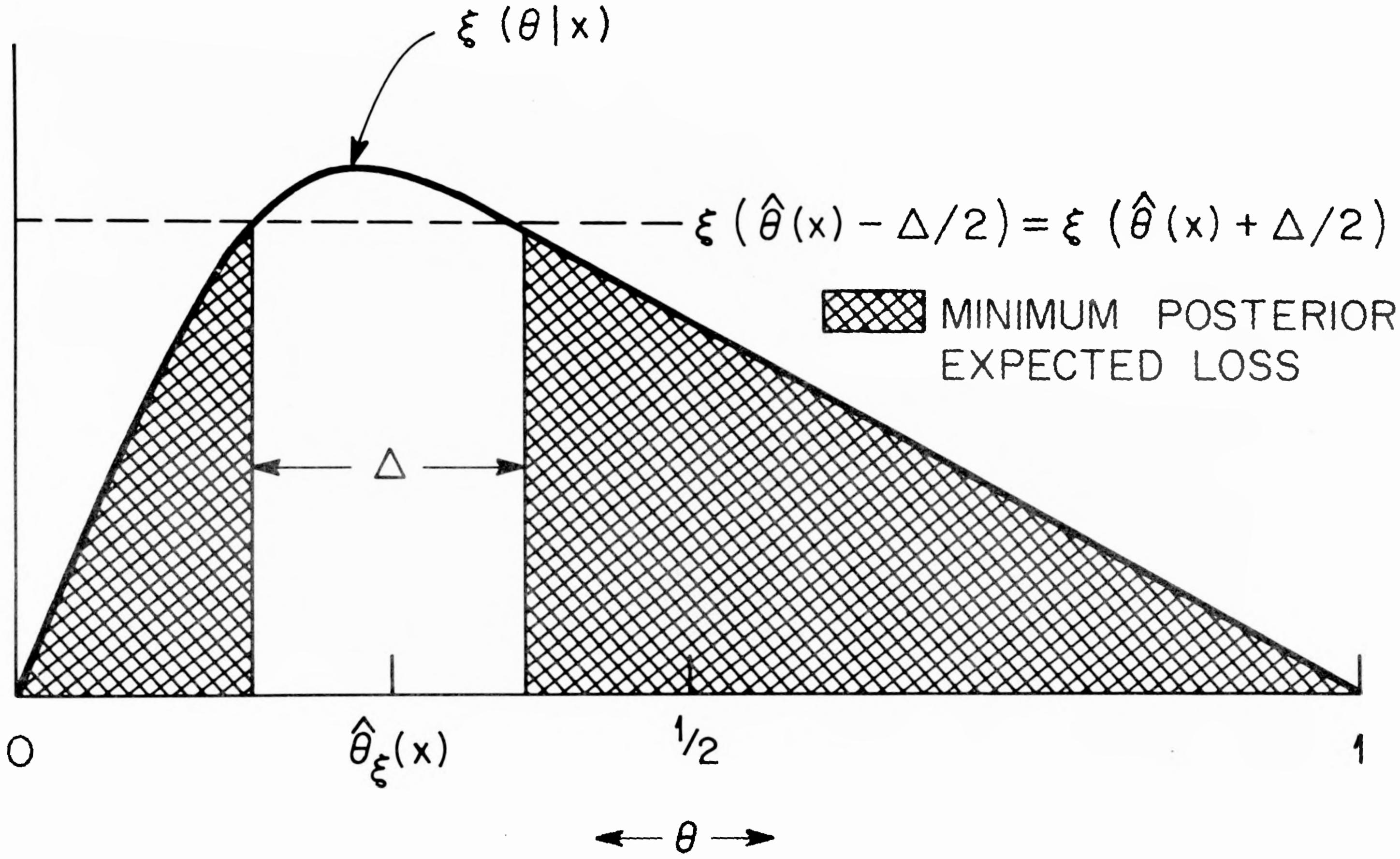

Figure 3.2. Posterior Density of θ: Case 1; Location of Bayes Estimate

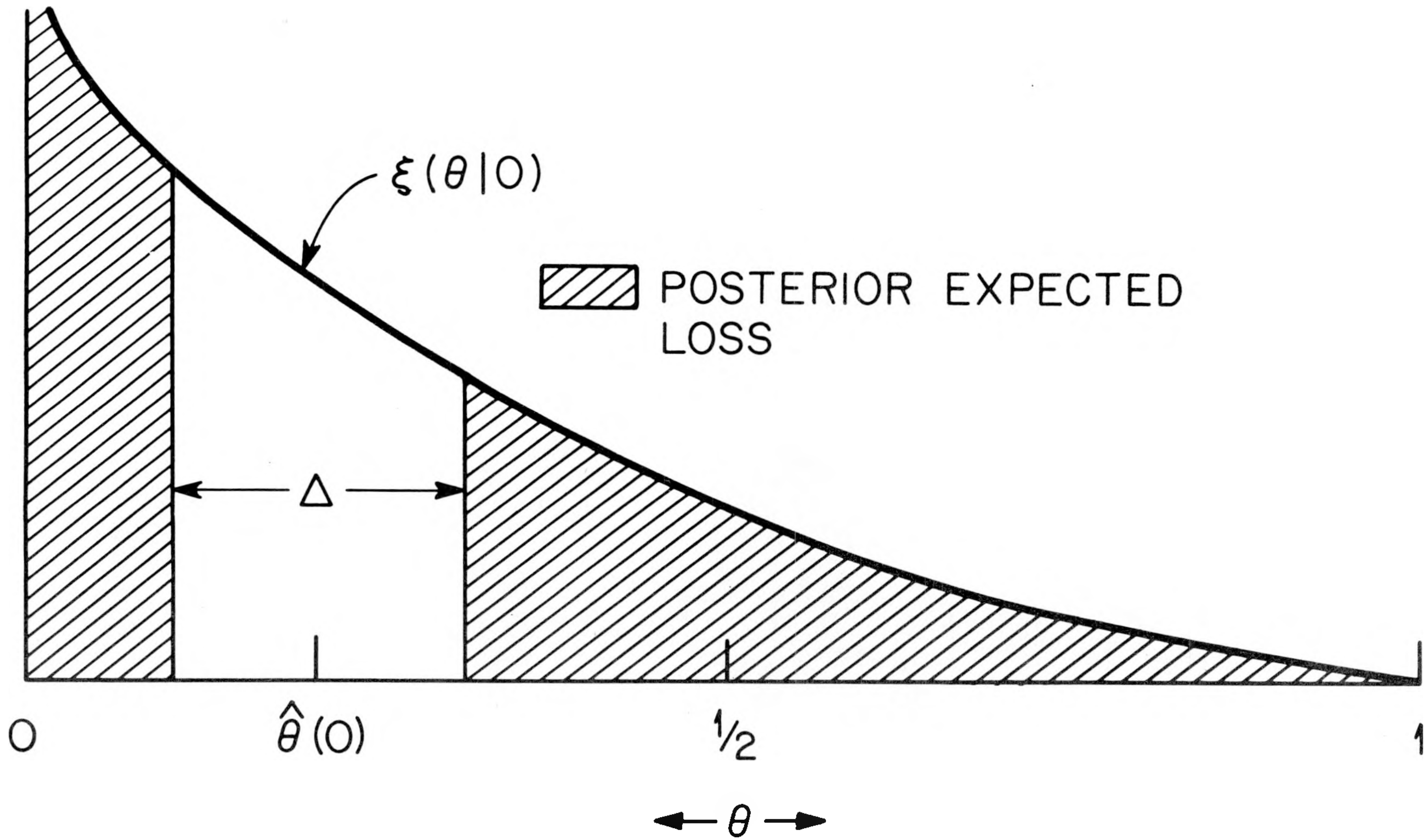

Figure 3.3. Posterior Density of θ: Case 2 (x = 0)

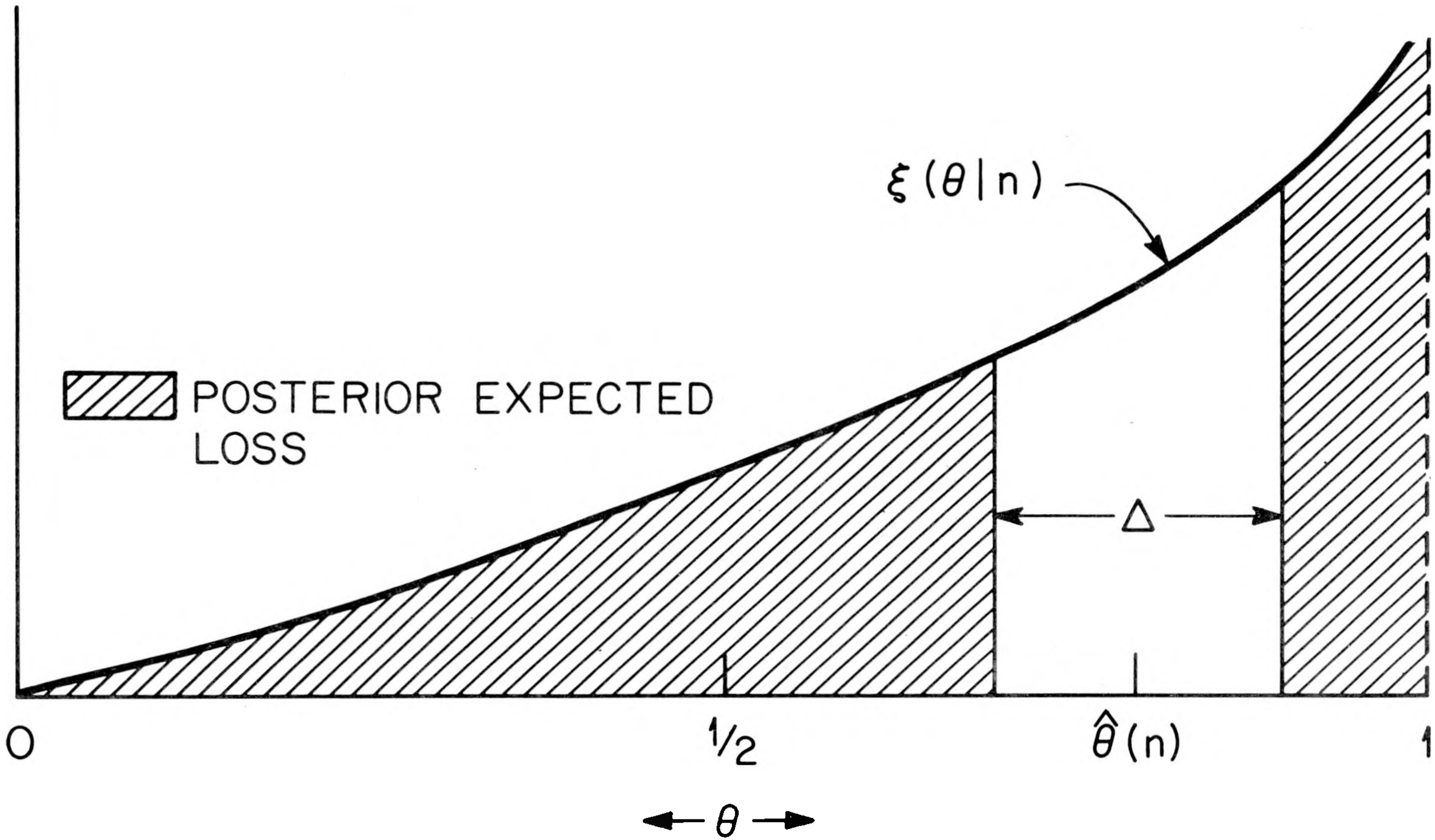

Figure 3.4. Posterior Density of θ: Case 2 (x = n)

$x = 0$ and $\alpha \le 0$, this interval is $[0,\Delta]$. If $x = n$ and $\alpha \le 0$, the interval containing the most posterior probability is $[1-\Delta,1]$.

Therefore, the Bayes estimates under the conditions of Case 2 are given by the rule:

$$\text{If } \alpha \le 0 \text{ and } \begin{cases} \text{(i)} \quad x = 0 \ , \quad \text{then } \hat{\theta}_\xi(0) = \Delta/2 \ ; \\ \text{(ii)} \quad x = n \ , \quad \text{then } \hat{\theta}_\xi(n) = 1 - \Delta/2 \ . \end{cases} \tag{3.12}$$

The Bayes estimator, $\hat{\theta}_\xi$, when the prior distribution of θ is $\xi(\theta)$, is given by the set of estimates

$$\hat{\theta}_\xi = \{\hat{\theta}_\xi(x) \ , \ x = 0,1,\ldots,n\}$$

where $\hat{\theta}_\xi(x)$ is found from (3.11) or (3.12), depending on which case applies.

3.1.2 Numerical Evaluation of Bayes Estimates

When the conditions of Case 2 in 3.1.1 hold, the value of the Bayes estimate, $\hat{\theta}_\xi(x)$, can be determined directly from the rule (3.12) for any value of n and Δ. However, when the conditions for Case 1 hold, $\hat{\theta}_\xi(x)$, which is the solution to equation (3.11), must be found by an iterative procedure. For the purposes of this study, equation (3.11) was solved using the technique described in this section.

It is obvious that $\Delta/2 \le \hat{\theta}_\xi(x) \le 1 - \Delta/2$, since $\hat{\theta}_\xi(x)$ represents the midpoint of an interval of width Δ, located in the interval $[0,1]$. Initially, let $\theta_L = \Delta/2$, and $\theta_U = 1 - \Delta/2$, so that $[\theta_L, \theta_U]$ represents an interval (initially of width $1 - \Delta$) that is known to contain $\hat{\theta}_\xi(x)$. The purpose of this iterative technique is to reduce the width of this interval containing $\hat{\theta}_\xi(x)$ until $\theta_L \doteq \theta_U$.

Starting with $\theta_L = \Delta/2$ and $\theta_U = 1 - \Delta/2$, the procedure is the same for every iteration:

(a) Let $\theta_M = (\theta_L + \theta_U)/2$;

(b) Evaluate $\xi(\theta_M - \Delta/2|x)$ and $\xi(\theta_M + \Delta/2|x)$ from (3.8);

(c) If $\xi(\theta_M - \Delta/2|x) > \xi(\theta_M + \Delta/2|x)$, then it must be true that $\theta_L \leq \hat{\theta}_\xi(x) \leq \theta_M$. Therefore, leave θ_L at its present value, but let $\theta_U = \theta_M$ and iterate again ;

(d) If $\xi(\theta_M + \Delta/2|x) > \xi(\theta_M - \Delta/2|x)$, then $\theta_M \leq \hat{\theta}_\xi(x) \leq \theta_U$. Leave θ_U at its present value, but let $\theta_L = \theta_M$ and iterate again.

After k iterations, the width of the interval $[\theta_L, \theta_U]$, which contains $\hat{\theta}_\xi(x)$, is equal to $(1-\Delta)/2^k$, and the value used for $\hat{\theta}_\xi(x)$ is $(\theta_L + \theta_U)/2$. Thus, the maximum error involved in finding the solution, $\hat{\theta}_\xi(x)$, is $(1-\Delta)/2^{k+1}$, and this error can be reduced to a satisfactory level by performing the appropriate number of iterations.

3.2 Estimates Which Minimize the Maximum Risk

For fixed values of n and Δ, and for a fixed prior $\xi(\theta)$, the Bayes estimates of θ may be obtained from (3.11) and/or (3.12), where (3.12) is solved numerically as in Section 3.1.2. Once the set of Bayes estimates is determined, the risk function for that set of estimates may be evaluated, using expression (3.2), at any value of θ. The maximum value of the risk function, $R(\hat{\theta}_\xi, \theta)$, over Θ may then be obtained by numerical search.

The search for the maximum value of the risk over Θ is complicated by the fact that the risk function is a discontinuous function of θ, with discontinuities occurring at the points $\hat{\theta}_\xi(x) \pm \Delta/2$, $x = 0, 1, \ldots, n$. However, the search can be simplified as a result of the following theorem.

Theorem 3.1. Let $\{\hat{\theta}(x),\ x = 0,1,\ldots,n\}$ be a set of Bayes estimates of θ, defined by equations (3.11) and/or (3.12), with risk function equal to $R(\hat{\theta}, \theta)$ in (3.2). Then the maximum of the risk function occurs at one or more of the 2n + 2 points.

$$\theta = \hat{\theta}(x) \pm \Delta/2 , \qquad \text{for } x = 0,1,\ldots,n .$$

The proof of Theorem 3.1 will be deferred until Lemma 3.1, below, has been proved. To introduce the lemma, first note that for any set of estimates of θ, $\{\hat{\theta}(i)\}$, $i = 1,2,\ldots,n$, there is an associated set of intervals in Θ, $\{(\hat{\theta}(i) - \Delta/2,\ \hat{\theta}(i) + \Delta/2)\}$. Denote these intervals by $\{I_0, I_1, \ldots, I_n\}$.

Lemma 3.1. $I_{x-1} \cap I_{x+1} \subset I_x$, $x = 1,2,\ldots,n-1$; i.e., any θ that is covered by I_{x-1} and I_{x+1} is also covered by I_x.

Proof: $$\theta \in I_{x-1} \Rightarrow \hat{\theta}(x - 1) - \Delta/2 \le \theta \le \hat{\theta}(x - 1) + \Delta/2 ,$$

$$\theta \in I_{x+1} \Rightarrow \hat{\theta}(x + 1) - \Delta/2 \le \theta \le \hat{\theta}(x + 1) + \Delta/2 .$$

Now suppose $\hat{\theta}(x + 1) > \hat{\theta}(x)$ for $x = 0,1,\ldots,n-1$. Then

$$\theta \in I_{x-1} \Rightarrow \theta \le \hat{\theta}(x) + \Delta/2 , \quad \text{and}$$

$$\theta \in I_{x+1} \Rightarrow \theta \ge \hat{\theta}(x) - \Delta/2 , \quad \text{so}$$

$$\theta \in I_{x-1} \cap I_{x+1} \Rightarrow \hat{\theta}(x) - \Delta/2 \le \theta \le \hat{\theta}(x) + \Delta/2 \Rightarrow \theta \in I_x .$$

That is, a sufficient condition for Lemma 3.1 to hold is that

$$\hat{\theta}(x + 1) > \hat{\theta}(x) , \qquad \text{for } x = 0,1,\ldots,n-1 . \tag{3.13}$$

To show (3.13) is true, consider the definition of $\hat{\theta}(x)$ given in Section 3.1.1. Assume $\alpha + x > 0$ and $\alpha + n - x > 1$. Then the estimates $\hat{\theta}(x)$ must satisfy equation (3.11), i.e., $\xi(\hat{\theta}(x) - \Delta/2|x) = \xi(\hat{\theta}(x) + \Delta/2|x)$, where $\xi(\theta|x)$ is the posterior density of θ, given by

$$\xi(\theta|x) = \frac{\theta^{\alpha+x}(1-\theta)^{\alpha+n-x}}{B(\alpha+x+1, \alpha+n-x+1)} .$$

Thus,

$$1 = \frac{\xi(\hat{\theta}(x) - \Delta/2|x)}{\xi(\hat{\theta}(x) + \Delta/2|x)} = \left[\frac{\hat{\theta}(x) - \Delta/2}{\hat{\theta}(x) + \Delta/2}\right]^{\alpha+x} \left[\frac{1 - \hat{\theta}(x) + \Delta/2}{1 - \hat{\theta}(x) - \Delta/2}\right]^{\alpha+n-x} .$$

Consider now the ratio of the values of the posterior density based on $X = x + 1$, evaluated at the endpoints of the interval defined by the Bayes estimate for $X = x$. (Again equation (3.11) can be used since $\alpha + (x + 1) > 0$ and $\alpha + n - (x + 1) > 0$.)

$$\frac{\xi(\hat{\theta}(x)-\Delta/2|x+1)}{\xi(\hat{\theta}(x)+\Delta/2|x+1)} = \left[\frac{\hat{\theta}(x)-\Delta/2}{\hat{\theta}(x)+\Delta/2}\right]^{\alpha+x+1} \left[\frac{1-\hat{\theta}(x)+\Delta/2}{1-\hat{\theta}(x)-\Delta/2}\right]^{\alpha+n-(x+1)}$$

$$= \left[\frac{\hat{\theta}(x)-\Delta/2}{\hat{\theta}(x)+\Delta/2}\right]^{\alpha+x} \left[\frac{1-\hat{\theta}(x)+\Delta/2}{1-\hat{\theta}(x)-\Delta/2}\right]^{\alpha+n-x} \frac{(\hat{\theta}(x)-\Delta/2)(1-\hat{\theta}(x)-\Delta/2)}{(\hat{\theta}(x)+\Delta/2)(1-\hat{\theta}(x)+\Delta/2)}$$

$$= 1 \cdot \left[\frac{\hat{\theta}(x)-\Delta/2}{\hat{\theta}(x)+\Delta/2}\right] \left[\frac{1-\hat{\theta}(x)-\Delta/2}{1-\hat{\theta}(x)+\Delta/2}\right] .$$

Since $\Delta > 0$, $\left[\frac{\hat{\theta}(x)-\Delta/2}{\hat{\theta}(x)+\Delta/2}\right] < 1$ and $\left[\frac{1-\hat{\theta}(x)-\Delta/2}{1-\hat{\theta}(x)+\Delta/2}\right] < 1$, so that

$\frac{\xi(\hat{\theta}(x)-\Delta/2|x+1)}{\xi(\hat{\theta}(x)+\Delta/2|x+1)} < 1$, or

$$\xi(\hat{\theta}(x) - \Delta/2|x+1) < \xi(\hat{\theta}(x) + \Delta/2|x+1) . \tag{3.14}$$

Because the function $\xi(\theta|x+1)$ increases monotonically from 0 at $\theta = 0$ to a single mode (at $\theta = \frac{x + \alpha + 1}{2\alpha + n}$), and then decreases monotonically to 0 at $\theta = 1$, the difference $D = \xi(\theta + \Delta/2|x+1) - \xi(\theta - \Delta/2|x+1)$ is positive for $\theta < \hat{\theta}(x + 1)$ and negative for all $\theta > \hat{\theta}(x + 1)$. (When $\theta = \hat{\theta}(x + 1)$ D is 0, by (3.11).) Thus, (3.14) implies that $\hat{\theta}(x) < \hat{\theta}(x + 1)$. This is illustrated in Figure 3.5.

It is still necessary to show that $\hat{\theta}(x) < \hat{\theta}(x + 1)$ when

(i) $\alpha + x \le 0$, i.e., $x = 0$, $\alpha \le 0$, and when

(ii) $\alpha + n - x \le 1$, i.e., $x = n-1$, $\alpha \le 0$.

In case (i), $\hat{\theta}(0) = \Delta/2$, from (3.12). Equation (3.11) gives

$$\xi(\hat{\theta}(1) - \Delta/2|x=1) = \xi(\hat{\theta}(1) + \Delta/2|x=1) \quad .$$

Now if $\hat{\theta}(1) = \Delta/2$, then $\xi(\Delta|x=1) = 0$, since $\xi(0|x=1) = 0$. But $\xi(\Delta|x=1) > 0$ for $0 < \Delta < 1$, so the assumption $\hat{\theta}(1) = \Delta/2$ leads to a contradiction, and is therefore false. Since $\Delta/2$ is the minimum value $\hat{\theta}(x)$ can assume, $\hat{\theta}(1) > \Delta/2 = \hat{\theta}(0)$. In case (ii), an analogous argument shows $\hat{\theta}(n) > \hat{\theta}(n - 1)$, when $\alpha \le 0$. This completes the proof of (3.13), and Lemma 3.1 is proved.

<u>Proof of Theorem 3.1</u>: Consider first the nontrivial case, in which $\bigcup_{x=0}^{n} I_x = \Theta$, i.e., every value of θ is covered by at least one I_x. The parameter space $\Theta = [0,1]$ is then the union of mutually exclusive subintervals S_i, each of which is covered by a nonempty subset of $\{I_o, I_1, \ldots, I_n\}$. For example, if $n = 2$, the interval $\Theta = [0,1]$ may be broken up into the five subintervals shown in Figure 3.6. By virtue of Lemma 3.1, each S_i covered by more than one interval from $\{I_o, I_1, \ldots, I_n\}$ is defined by the intersection of consecutive intervals, i.e.,

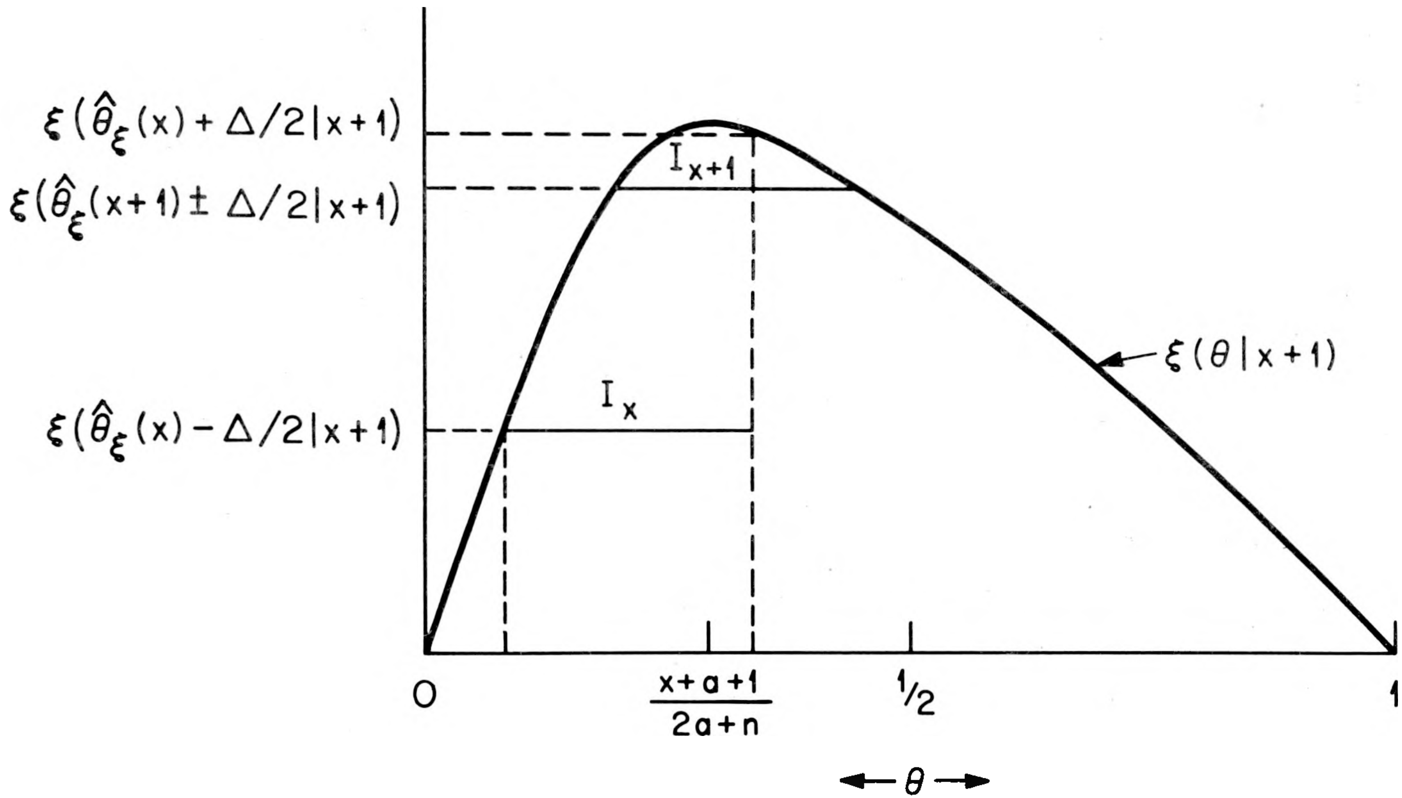

Figure 3.5. Posterior Density of θ for $X = x + 1$

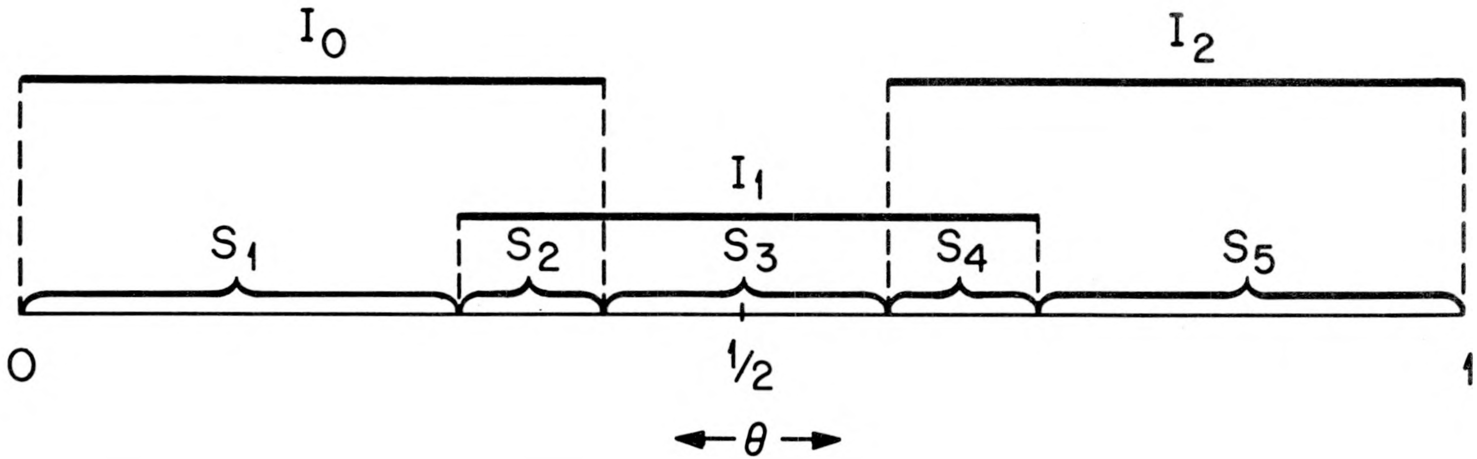

Figure 3.6. Decomposition of Θ Into Subintervals S_i (n = 2)

$S_i = \bigcap_{j=0}^{k_i - 1} I_{r_i + j}$, where k_i is the number of intervals which intersect to form S_i, and r_i the smallest value of X that defines an interval covering S_i. Delete the subscript i, and consider a general subinterval, $S = \bigcap_{j=0}^{k-1} I_{r+j}$. The behavior of the risk function over S will now be considered for several cases, depending on the value of r.

Case 1: r = 0. When S is covered by the first k intervals, the risk function over S from (3.2) is

$$R_S(\hat{\theta}, \theta) = 1 - \sum_{x=0}^{k-1} \binom{n}{x} \theta^x (1 - \theta)^{n-x} .$$

It will be convenient to define $R_S(\hat{\theta}, \theta)$ (abbreviated R_S) over $\Theta = [0,1]$, although it coincides with the risk function $R(\hat{\theta}, \theta)$ only over S.

The partial derivative of R_S with respect to θ is

$$\frac{\partial R_S}{\partial \theta} = n(1 - \theta)^{n-1} - \sum_{x=1}^{k-1} \binom{n}{x} \left[-(n - x) \theta^x (1 - \theta)^{n-x-1} + x \theta^{x-1} (1 - \theta)^{n-x}\right].$$

Let $b(x;n-1,\theta) = \frac{(n-1)!}{x!(n-x-1)!} \theta^x (1 - \theta)^{n-x-1}$, so that

$$\theta^x (1 - \theta)^{n-x-1} = \frac{b(x;n-1,\theta)\ x!(n-x-1)!}{(n-1)!} .$$

Similarly,

$$\theta^{x-1} (1 - \theta)^{n-x} = \frac{b(x-1;n-1,\theta)\ (x-1)!\ (n-x)!}{(n-1)!}$$

Then

$$\frac{\partial R_S}{\partial \theta} = n(1-\theta)^{n-1} - \sum_{x=1}^{k-1} \frac{n!}{x!(n-x)!}\left[\frac{(-1)(n-x)\ b(x;n-1,\theta)\ x!(n-x-1)!}{(n-1)!}\right]$$

$$- \sum_{x=1}^{k-1} \frac{n!}{x!(n-x)!}\left[\frac{x\ b(x-1;n-1,\theta)(x-1)!\ (n-x)!}{(n-1)!}\right]$$

$$= n(1-\theta)^{n-1} - n \sum_{x=1}^{k-1} [-b(x;n-1,\theta) + b(x-1;n-1,\theta]$$

$$= n\ b(k-1;n-1,\theta)$$

$$\frac{\partial R_S}{\partial \theta} = n \frac{(n-1)!}{(k-1)!\ (n-k)!}\ \theta^{k-1}\ (1-\theta)^{n-k} \quad .$$

This derivative is positive over $\Theta = [0,1]$, except at the points $\theta = 0$ and $\theta = 1$, where it is zero. This indicates that, for $0 < \theta < 1$, $R_S(\hat{\theta}, \theta)$ is an increasing function of θ. In particular, $R_S(\hat{\theta}, \theta)$ is an increasing function of θ over the interval S, where it coincides with the risk $R(\hat{\theta}, \theta)$. Thus, the maximum risk over S must occur at its right end point.

<u>Case 2</u>: $r = n - k + 1$. This occurs when the last k intervals cover S. The risk function, over S, is

$$R_S(\hat{\theta}, \theta) = 1 - \sum_{x=n-k+1}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$= \sum_{x=0}^{n-k} \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad .$$

Again, R_S is defined over Θ, although it coincides with the risk $R(\hat{\theta}, \theta)$ only over S. Now differentiate and collect terms just as in Case 1. The result is

$$\frac{\partial R_S}{\partial \theta} = - n \frac{(n-1)!}{(n-k)!\,(k-1)!} \theta^{n-k} (1 - \theta)^{k-1} .$$

This derivative is negative for $0 < \theta < 1$, and zero at $\theta = 0$ and $\theta = 1$, indicating the risk is a decreasing function of θ for $0 < \theta < 1$. In particular, $R_S(\hat{\theta}, \theta)$ is a decreasing function of θ over the interval S, where it coincides with the risk $R(\hat{\theta}, \theta)$. Thus, the maximum risk over S must occur at its left end point.

<u>Case 3</u>: $1 \le r \le n-k$. For this case the risk, over S, is

$$R_S(\hat{\theta}, \theta) = 1 - \sum_{j=0}^{k-1} \binom{n}{r+j} \theta^{r+j} (1 - \theta)^{n-r-j}$$

$$\frac{\partial R_S}{\partial \theta} = - \sum_{j=0}^{k-1} \binom{n}{r+j} \left[-\theta^{r+j}(n-r-j)(1-\theta)^{n-r-j-1} + (r+j)\, \theta^{r+j-1}(1-\theta)^{n-r-j}\right] .$$

Again, let $b(r+j;n-1,\theta) = \frac{(n-1)!}{(r+j)!\,(n-r-j-1)!} \theta^{r+j} (1 - \theta)^{n-r-j-1}$, and $b(r+j-1;n-1,\theta) = \frac{(n-1)!}{(r+j-1)!\,(n-r-j)!} \theta^{r+j-1} (1 - \theta)^{n-r-j}$, so that

$$\frac{\partial R_S}{\partial \theta} = - \sum_{j=0}^{k-1} \binom{n}{r+j} \left[(-1)(n-r-j) \frac{b(r+j;n-1,\theta)\,(r+j)!\,(n-r-j-1)!}{(n-1)!} + (r+j) \frac{b(r+j-1;n-1,\theta)\,(r+j-1)!\,(n-r-j)!}{(n-1)!}\right]$$

$$= - n \sum_{j=0}^{k-1} [b(r+j-1;n-1,\theta) - b(r+j;n-1,\theta)]$$

$$\frac{\partial R_S}{\partial \theta} = - n\, [b(r-1;n-1,\theta) - b(r+k-1;n-1,\theta)] . \tag{3.15}$$

This derivative is zero when $b(r-1;n-1,\theta) = b(r+k-1;n-1,\theta)$, or

$$\frac{(n-1)!}{(r-1)!\,(n-r)!}\,\theta^{r-1}\,(1-\theta)^{n-r} = \frac{(n-1)!}{(r+k-1)!\,(n-r-k)!}\,\theta^{r+k-1}\,(1-\theta)^{n-r-k}\ .$$

Equivalently, $\frac{\partial R_S}{\partial\theta}$ is zero when

$$\frac{(r+k-1)!\,(n-r-k)!}{(r-1)!\,(n-r)!} = \left[\frac{\theta}{1-\theta}\right]^k\ .$$

$\left[\frac{\theta}{1-\theta}\right]^k$ increases from 0 to ∞ as θ goes from 0 to 1, indicating that it is equal to the constant $\frac{(r+k-1)!\,(n-r-k)!}{(r-1)!\,(n-r)!}$ at only one point in Θ. This one stationary point can be diagnosed to be a minimum by studying expression (3.15). For θ sufficiently close to 0, $b(r-1;n-1,\theta) > b(r+k-1;n-1,\theta)$, so $\frac{\partial R_S}{\partial\theta}$ is negative. For θ sufficiently close to 1, $b(r-1;n-1,\theta) < b(r+k-1;n-1,\theta)$ and $\frac{\partial R_S}{\partial\theta}$ is positive. Thus the derivative of the risk is negative to the left of the stationary point and positive to the right of it, so the stationary point must be a minimum. This implies that $R_S(\hat{\theta},\theta)$ decreases monotonically to a minimum and then increases monotonically as θ goes from 0 to 1. As a result, the maximum of $R_S(\hat{\theta},\theta)$ over any interval in $[0,1]$ must occur at one of the endpoints of the interval. In particular, this is true for the interval S, where $R_S(\hat{\theta},\theta)$ coincides with the risk function $R(\hat{\theta},\theta)$. Thus the maximum risk over S must occur at one of the endpoints of S.

For the three cases considered, it has been demonstrated that, for any subinterval S covered by a specific nonempty subset of the intervals $\{I_0, I_1, \ldots, I_n\}$, the maximum risk over that subinterval occurs at one or both of its endpoints. But since S is of the form $\bigcap_{j=0}^{k-1} I_{r+j}$, the left endpoint of S is the left endpoint of I_r and the right endpoint of S is the right endpoint of I_{r+k-1}. The endpoints of any such subinterval S must therefore coincide with two of the 2n + 2 points $\theta = \theta(x) \pm \Delta/2$, $x = 0,1,\ldots,n$. Since the union of all possible subintervals, S_i, is equal to $[0,1]$, the search over the parameter space $\Theta = [0,1]$ for the maximum risk may be restricted to the 2n + 2 points

$$\theta = \hat{\theta}_\xi(x) \pm \Delta/2 , \quad x = 0,1,\ldots,n .$$

Theorem 3.1 has now been proved for the nontrivial case, in which $\bigcup_{x=0}^{n} I_x = \Theta$. In the trivial case, there are some intervals in $[0,1]$ which are not included in any I_x. Denote any such interval by S. Then, from (3.2), the risk $R(\hat{\theta},\theta)$ attains its maximum (one) at all θ in S. In particular, the maximum risk occurs at both endpoints of S, each of which must coincide with an endpoint of one of the intervals $\{I_x\}$. That is, each endpoint of S must be from the set $\{\hat{\theta}(x) \pm \Delta/2\}$, $x = 0,1,\ldots,n$, and Theorem 3.1 holds for the trivial case.

The symmetric beta prior density in (3.3) is a function of only one parameter, α. Thus, the search for the set of SBP estimates that minimize the maximum risk is equivalent to the search over its range for that value of the parameter α (and the prior indexed by it) whose Bayes estimates are minimax.

In order for the beta function to be a true density, it is necessary that $\alpha > -1$. Moreover, if $\alpha > 0$, the maximum risk is 1. This can be seen by noting that if x = 0 or n and $\alpha > 0$, then $\hat{\theta}(0) > \Delta/2$, and $\hat{\theta}(n) < 1 - \Delta/2$, since the posterior density of θ takes the form shown in Figure 3.2. This means that, for sufficiently small ϵ, values of θ such that $\theta < \epsilon$ and $\theta > 1 - \epsilon$ are not covered by any of the intervals $\theta(x) \pm \Delta/2$, so the risk at these values of θ is equal to one, the maximum possible. Therefore, if α is to minimize the maximum risk, it cannot exceed 0, and the search for optimal α can be restricted to $-1 < \alpha \leq 0$.

The numerical search was conducted for the 21 combinations of the following values of n and Δ in which $\Delta > 1/(n+1)$:

Δ: .05, .10, .15;

n: 9, 16, 25, 36, 49, 64, 81, 100.

The condition $\Delta > 1/(n+1)$ is necessary because, if $\Delta < 1/(n+1)$, any n+1 intervals defined by $x = 0,1,\ldots,n$ would not cover all points in Θ, and the maximum risk would always be equal to one. If $\Delta = 1/(n+1)$, there is only one set of estimates with maximum risk less than one, that set defining n+1 intervals that are adjacent but nonoverlapping.

For all 21 cases considered it was determined that $\alpha = 0$ corresponded to the symmetric beta prior whose Bayes estimates produced the minimax risk. The results of the 21 numerical searches are summarized in Table II.

The 21 cases studied covered a fairly wide range of values for n and Δ, and in all 21 cases the maximum risk was a decreasing function of α, for $-1 < \alpha \leq 0$. Because of these numerical results, it is conjectured that, for general n and Δ, the uniform prior (i.e., $\alpha = 0$) gives Bayes estimates that produce minimax risk among the class of SBP estimates.

3.3 Estimates Which Minimize the Maximum Weighted Risk

In this section, the optimum estimator will be defined to be the one that minimizes the maximum possible weighted risk, where the class of weighting functions is the set of symmetric beta functions on $\Theta = [0,1]$. The estimates $\hat{\theta}(x)$ are again restricted to the class of SBP estimates, i.e., Bayes estimates derived from symmetric beta prior distributions on Θ. The risk function is given in (3.2).

Table II

Fixed Precision Estimation of θ

C-Minimax SBP Estimates

Δ	n	α^*	C-Minimax Risk
.05	25	0	.84365
.05	36	0	.86800
.05	49	0	.77611
.05	64	0	.70935
.05	81	0	.65875
.05	100	0	.61902
.10	16	0	.80453
.10	25	0	.69140
.10	36	0	.61850
.10	49	0	.56819
.10	64	0	.45459
.10	81	0	.37625
.10	100	0	.31941
.15	9	0	.74560
.15	16	0	.62160
.15	25	0	.54947
.15	36	0	.40672
.15	49	0	.31872
.15	64	0	.26066
.15	81	0	.18325
.15	100	0	.13437

α^* = value of α in (3.3) whose corresponding Bayes estimates found by (3.11) and (3.12) are minimax in the class C of SBP estimates.

If $\omega^*(\theta)$ denotes the weighting function in Ω that produces the maximum weighted risk for the set of SBP estimates based on the prior $\xi(\theta)$, and if $\tilde{R}(\xi,\omega)$ is the weighted risk, then the objective of this section is to find ξ^* such that

$$\tilde{R}(\xi^*,\omega^*) = \inf_{\xi\in\Xi} \tilde{R}(\xi,\omega^*) .$$

Since the function $\omega(\theta)$ is specified by the single parameter b and $\xi(\theta)$ is specified by the single parameter a, it is convenient to consider the weighted risk $\tilde{R}$ as a function of a and b, so the objective is to determine a^* such that $\tilde{R}(a^*,b^*) = \inf_a \sup_b \tilde{R}(a,b)$.

For fixed values of n and Δ, the search for a^* was conducted numerically. For given a, the estimates $\{\hat{\theta}(x),\ x = 0,1,\ldots,n\}$ based on $\xi(\theta)$ were obtained from (3.11) and (3.12) as before. For that set of estimates if the risk function is weighted by $\omega(\theta)$, then the weighted risk is, by definition,

$$\tilde{R}(a,b) = \int_0^1 \left[\sum_{x=0}^{n} L(\hat{\theta}(x),\theta)\ p(x|\theta)\right] \omega(\theta)\ d\theta$$

$$= \sum_{x=0}^{n}\left[\int_0^1 L(\hat{\theta}(x),\theta)\ \omega(\theta|x)\ d\theta\right] p(x) , \qquad (3.16)$$

where

$$p(x) = \int_0^1 p(x|\theta)\ \omega(\theta)\ d\theta = \binom{n}{x} \frac{B(b+x+1,b+n-x+1)}{B(b+1,b+1)} \qquad (3.17)$$

and

$$\omega(\theta|x) = \frac{p(x|\theta)\ \omega(\theta)}{p(x)} = \frac{\theta^{b+x}\,(1-\theta)^{b+n-x}}{B(b+x+1,b+n-x+1)} .$$

By analogy to (3.9),

$$\int_0^1 L(\hat{\theta}(x),\theta)\ \omega(\theta|x)\ d\theta = \frac{1}{B(b+x+1, b+n-x+1)}\left\{\int_0^{\hat{\theta}(x)-\Delta/2} f(\theta)\ d\theta + \int_{\hat{\theta}(x)+\Delta/2}^1 f(\theta)\ d\theta\right\}, \tag{3.18}$$

where $f(\theta) = \theta^{b+x}(1-\theta)^{b+n-x}$.

For fixed values of (a, b) the corresponding weighted risk, $\tilde{R}(a, b)$, was obtained from (3.16), using (3.17), (3.18), and an algorithm by Amos [1] for evaluating the complete and incomplete beta functions.

For SBP estimates based on the prior $\xi(\theta)$ with parameter a, the maximum weighted risk was found by numerical search through the class of symmetric beta weighting functions indexed by b, for $-.99 \le b \le 100$. That is, b^* was determined such that $\tilde{R}(a, b^*) = \sup_b \tilde{R}(a, b)$. Then the interval $-.99 \le a \le 100$ was searched to find a^* such that $\tilde{R}(a^*, b^*) = \inf_a \tilde{R}(a, b^*)$. Thus, a^* indexes the prior $\xi^*(\theta)$ that gives SBP estimates that minimize the maximum weighted risk, and b^* indexes the symmetric beta function $\omega^*(\theta)$ that maximizes the weighted risk for the estimator based on a^*.

The results, for the same combinations of n and Δ studied in the previous section, are given in Table III. Note that in every case the uniform prior ($a = 0$) is the one whose Bayes estimates produce the minimax weighted risk. The same estimates were found, in the previous section, to be optimal for the C-minimax criterion. Thus, in the cases considered, the set of Bayes estimates based on the uniform prior are the SBP estimates which minimize both the maximum risk and the maximum weighted risk. Figure 3.7 is a graph of the maximum risk and maximum

Table III

Fixed Precision Estimation of θ

Minimax Weighted Risk SBP Estimators

Δ	n	a^*	b^*	Minimax Weighted Risk
.05	25	0	14	.80052
.05	36	0	14	.76109
.05	49	0	14	.72229
.05	64	0	14	.68392
.05	81	0	14	.64641
.05	100	0	14	.61033
.10	16	0	16	.68636
.10	25	0	15	.61134
.10	36	0	15	.54085
.10	49	0	15	.47498
.10	64	0	15	.41336
.10	81	0	15	.35612
.10	100	0	14	.30394
.15	9	0	17	.65030
.15	16	0	17	.54043
.15	25	0	16	.44102
.15	36	0	16	.35358
.15	49	0	16	.27887
.15	64	0	16	.21595
.15	81	0	16	.16312
.15	100	0	15	.11905

a^* = value of a in (3.3) whose corresponding Bayes estimates found by (3.11) and (3.12) have minimax weighted risk in the class C of SBP estimates.

b^* = value of b defining the symmetric beta weighting function which, when used to weight the risk function for the SBP estimator based on $a = a^*$ (as in (3.16)), maximizes the weighted risk.

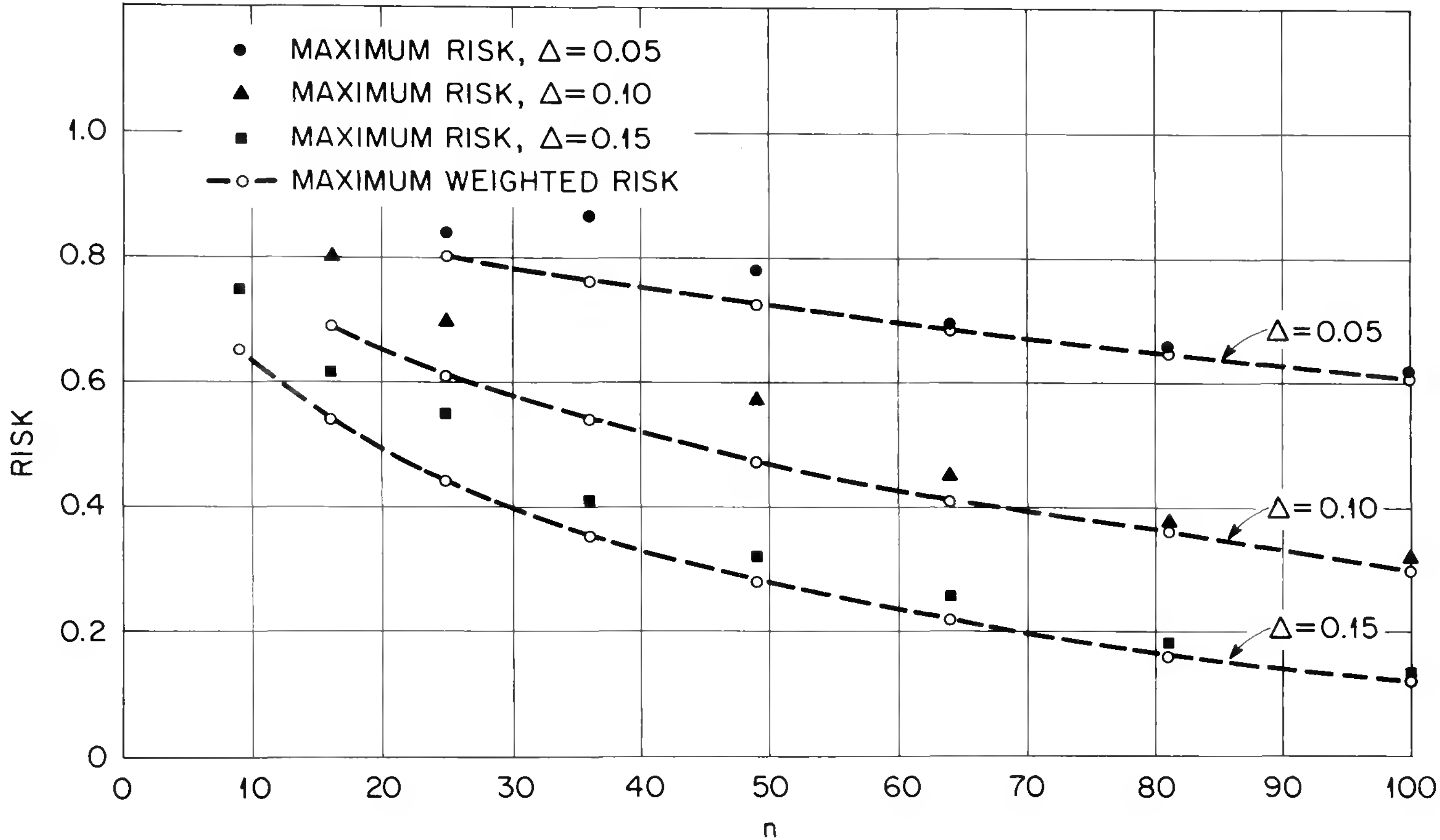

Figure 3.7. Maximum Risk and Maximum Weighted Risk as a Function of n

weighted risk as a function of n. Note that, as n gets large, there is very little difference between the minimax risk and minimax weighted risk.

3.4 Estimates Based on the Least Favorable Prior

When the weighting function in Section 3.3 coincides with the prior from which the estimates are derived, then the weighted risk, $\widetilde{R}(\xi,\xi)$, is called the Bayes risk. The prior that produces the maximum Bayes risk is designated the "least favorable" prior.

Since ξ is completely specified by the parameter a, where $-1 < a < \infty$, it is convenient to consider the Bayes risk $\widetilde{R}(\xi,\xi)$ simply as $\widetilde{R}(a)$. $\widetilde{R}(a)$ can be determined from (3.16), (3.17), and (3.18) by replacing b with a. Thus,

$$\widetilde{R}(a) = \sum_{x=0}^{n} \frac{\binom{n}{x}}{B(a+1,a+1)} \int_{0}^{\hat{\theta}(x)-\Delta/2} f(\theta)\, d\theta + \int_{\hat{\theta}(x)+\Delta/2}^{1} f(\theta)\, d\theta \qquad (3.19)$$

where $f(\theta) = \theta^{a+x}(1-\theta)^{a+n-x}$. Again, $\hat{\theta}(x)$ is obtained using (3.12) or the numerical solution of (3.11).

For fixed values of n and Δ, a numerical search was conducted over the interval $-.99 \le a \le 100$, in order to find the value of a corresponding to the "least favorable" prior. Numerical evaluation of (3.19) was again achieved by using the algorithm by Amos [1].

The results are presented in Table IV. Comparisons of the risk functions for the set of estimates suggested in Sections 3.2 and 3.3 ($a = 0$), and the set of Bayes estimates corresponding to the "least favorable" symmetric beta prior are given in Figures 3.8, 3.9, and 3.10 for particular values of n and Δ. The exact behavior of the risk is

Table IV

Fixed Precision Estimation of θ

SBP Estimator Corresponding to the Least Favorable Prior

Δ	n	a^*	Bayes Risk
.05	25	2.22	.76403
.05	36	2.72	.72625
.05	49	3.16	.68909
.05	64	3.70	.65266
.05	81	4.20	.61705
.05	100	4.69	.58231
.10	16	1.70	.61649
.10	25	2.16	.54813
.10	36	2.61	.48368
.10	49	3.08	.42354
.10	64	3.52	.36801
.10	81	3.95	.31725
.10	100	4.44	.27136
.15	9	1.20	.54657
.15	16	1.62	.45213
.15	25	2.02	.36742
.15	36	2.45	.29325
.15	49	2.86	.22982
.15	64	3.25	.17680
.15	81	3.61	.13351
.15	100	3.95	.09901

a^* = value of a in (3.3) having maximum Bayes risk.

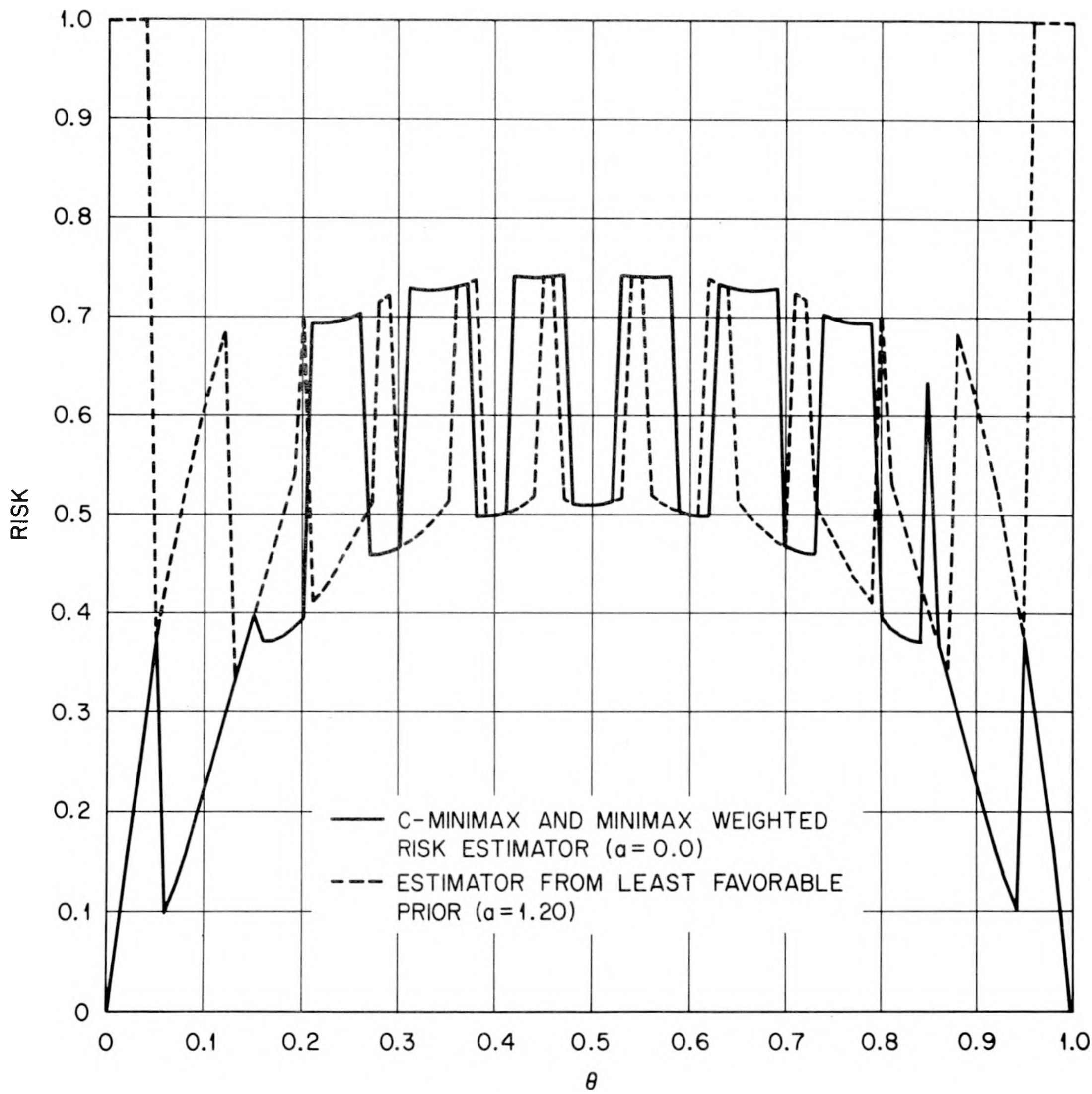

Figure 3.8. Risk Functions for $n = 9$, $\Delta = .15$

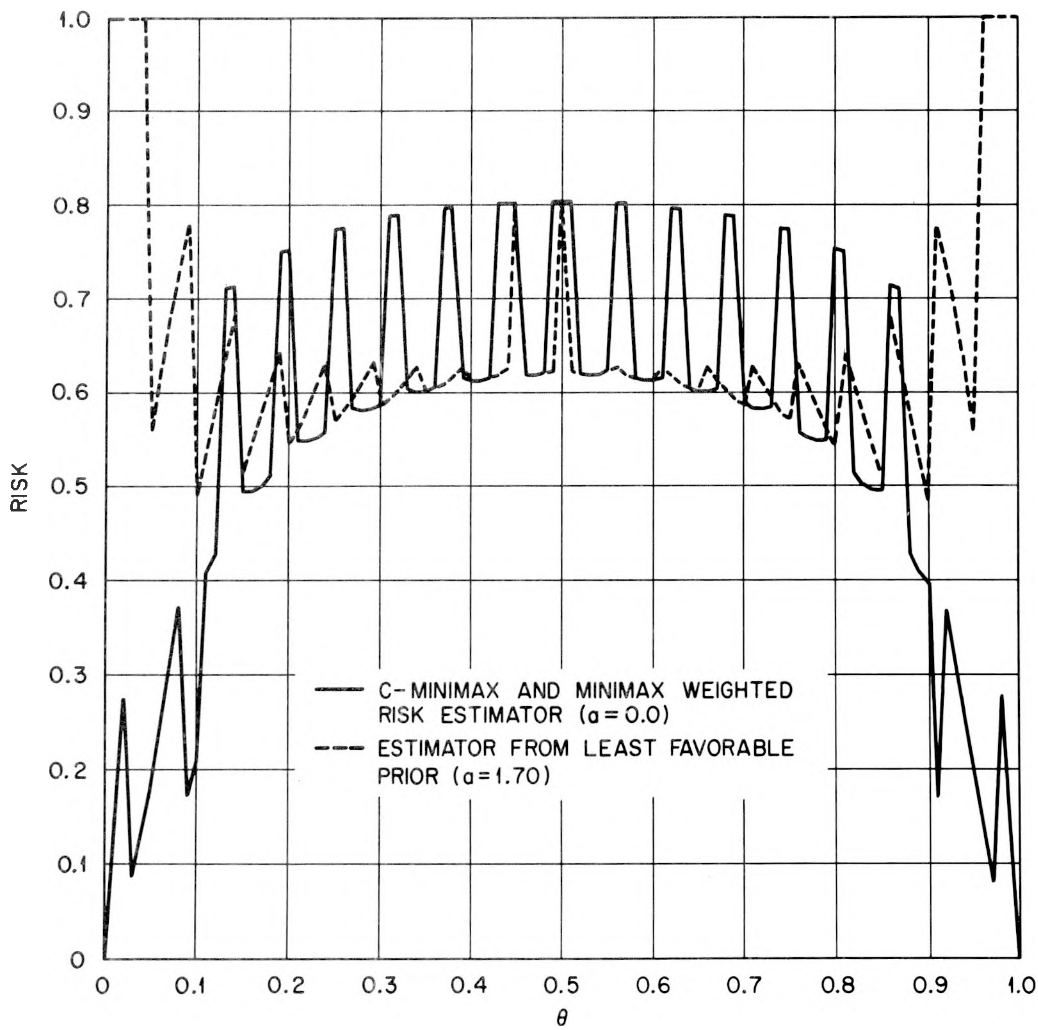

Figure 3.9. Risk Functions for n = 16, Δ = .10

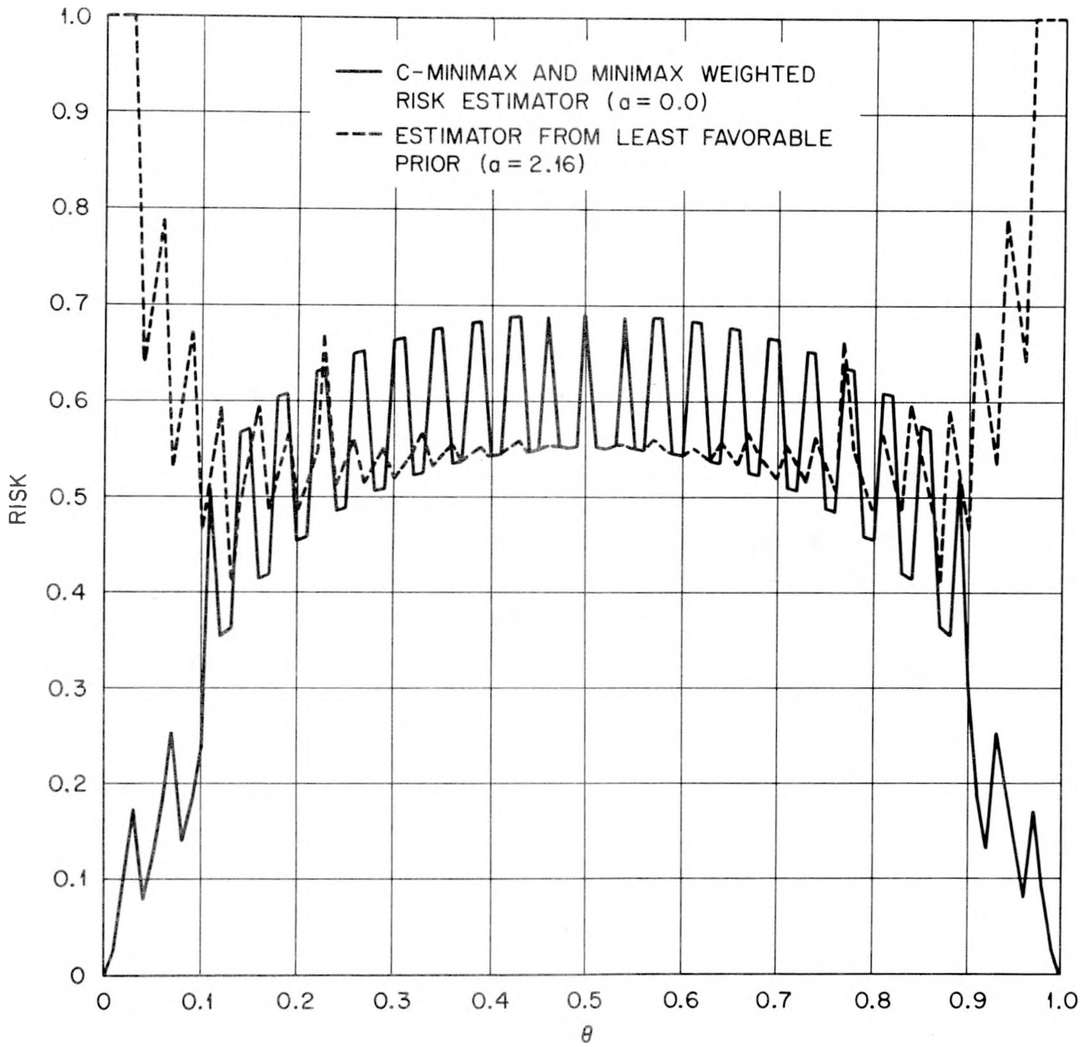

Figure 3.10. Risk Functions for n = 25, Δ = .10

difficult to determine in these three figures, because the risk was evaluated at a finite number of points for equal intervals in θ. However, it appears that the risk for the "least favorable" estimator may be smaller than the risk for the estimator based on $\alpha = 0$ for intervals near $\theta = 1/2$. This advantage, though, is somewhat overshadowed by the behavior near θ equal to 0 and 1, where the risk for the "least favorable" estimator is 1.

3.5 Fixed Width Confidence Intervals for θ

Although this chapter is concerned with fixed precision estimation of the binomial parameter θ, the set of estimates $\hat{\theta}(x)$, $x = 0,\ldots,n$, immediately gives a set of fixed width confidence intervals for θ. The interval corresponding to the outcome x is simply $[\hat{\theta}(x) - \Delta/2, \hat{\theta}(x) + \Delta/2]$ and the confidence $C(\theta)$ is simply $1 - R(\hat{\theta}, \theta)$, where $R(\hat{\theta}, \theta)$ is the risk function (3.2).

The traditional confidence interval for θ has width that is dependent on the value of the sample observation obtained from the binomial distribution. Various methods are available for obtaining such confidence intervals. For references, the reader should consult Greenwood and Hartley's Guide to Tables in Mathematical Statistics [11], Section 3.32 on "Confidence Limits for Binomial Distribution."

The importance of being able to obtain a confidence interval for θ whose width is independent of the outcome of the experiment was discussed by Steinhaus [20]. On the other hand, an argument against considering fixed-width confidence intervals is that certain outcomes of a binomial experiment may well carry more "information" about θ than others.

Intuitively it seems that those confidence intervals which are associated with the more "informative" outcomes should be narrower.

The object of this section is not to argue the merits of fixed-width confidence intervals, but to compare the intervals derived from the "optimal" fixed precision estimates of the previous sections with those given in the literature by Naddeo [18]. Naddeo suggests using the maximum likelihood estimate, x/n, as the midpoint of the interval. This leads to the interval

$$x/n - \Delta/2 \leq \theta \leq x/n + \Delta/2 \quad , \tag{3.20}$$

where Δ is the fixed interval width. For a fixed value of θ, the confidence level of this set of intervals is

$$C(\theta) = \sum_{E} \binom{n}{x} \theta^x (1 - \theta)^{n-x} \tag{3.21}$$

where $E = \{x | x/n - \Delta/2 \leq \theta \leq x/n + \Delta/2\}$. In order to determine the value of n necessary to make $\min_{\theta} C(\theta)$ sufficiently large, Naddeo gave the following approximation:

$$\bar{n}\,\Delta^2 \cong z^2_{\alpha/2} \tag{3.22}$$

where $z_{\alpha/2}$ is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{z_{\alpha/2}}^{\infty} e^{-\frac{1}{2}t^2}\, dt = \frac{\alpha}{2} \tag{3.23}$$

and $\bar{n}$ is the minimum n necessary to insure the

$$\Pr\{x/n - \Delta/2 \leq \theta \leq x/n + \Delta/2\} \geq 1 - \alpha \quad . \tag{3.24}$$

These results were based on the asymptotic distribution of Pearson's χ^2 goodness of fit statistic and the relationship between the χ_1^2 distribution and the standard normal distribution. In this section, the accuracy of Naddeo's approximation (3.22) will be examined. At the same time, a comparison will be made with the Bayes-suggested intervals based on the uniform prior, which were found to be optimal with respect to the criteria of Sections 3.2 and 3.3

Notice that, when $a = 0$ in (3.3) (i.e., the beta prior is simply the uniform density) the posterior density, $\xi(\theta|x)$ in (3.8), is just the likelihood function, $p(x|\theta)$. Thus the Bayes-suggested fixed-width confidence intervals based on the uniform prior are such that every value of θ in an interval has a greater likelihood than every point outside the interval. Naddeo's intervals do not have this property, since they are centered on the maximum values of the likelihood functions, which are not generally symmetric.

Another apparent disadvantage of Naddeo's intervals stems from the fact that, when $x = 0$ or n, they include values outside the range of θ. If $x = 0$, Naddeo's method gives the interval $[-\Delta/2, \Delta/2]$, whereas the interval defined in (3.12) for $a = 0$ is $[0, \Delta]$. A similar comparison holds if $x = n$. For $1 \le x \le n-1$, the two sets of interval estimates differ only slightly, and they become almost identical as n gets large.

A comparison of the two sets of interval estimates and a study of the accuracy of the approximate relationship (3.22) in predicting the minimum confidence level for Naddeo's set of interval estimates were carried out simultaneously. The exact minimum confidence level was plotted as a function of n for both sets of estimates, along with the predicted minimum confidence level for Naddeo's estimates, $1 - \alpha$, where α is determined by (3.22). These graphs are given in Figures 3.11, 3.12, and 3.13 for $1 \leq n \leq 100$ and interval widths $\Delta = .05$, .10, and .15, respectively.

In no case is the minimum confidence greater for Naddeo's intervals than for the Bayes-suggested intervals based on the uniform prior, indicating that the fixed-width intervals with x/n as midpoints are inadmissible. The gain in confidence achieved by using the intervals with midpoints obtained from the solutions of (3.11) and (3.12), instead of the intervals with midpoints equal to x/n, for $x = 0,1,\ldots,n$, may or may not be worth the extra effort involved. The amount of improvement depends on the values of Δ and n, as indicated in Figures 3.11, 3.12, and 3.13.

The degree of inaccuracy of the approximation (3.22) for the minimum confidence level for Naddeo's estimates is also indicated in these three figures. This inaccuracy is probably due to the continuous χ^2 approximation for the Pearson goodness of fit statistic, which is discrete. Unfortunately, Naddeo's approximation overestimates the minimum confidence for given n. Equivalently, if (3.22) is to be used to select n to achieve a specified confidence, the value of n obtained is too small. A rule-of-thumb for correcting this error can be formulated by inspecting Figures 3.11, 3.12, and 3.13, where it is noted that the "jumps" in the confidence level for Naddeo's estimates occur when n is the smallest integer greater than or equal to J/Δ, $J = 1,2,3,\ldots$. Denote each such n by n_J. For a specified confidence $1 - \alpha$, use (3.22) to find $\bar{n}$. Then let the sample size n be the smallest n_J which is greater than or equal to $\bar{n}$. An added advantage of choosing n at one of these "jump points"

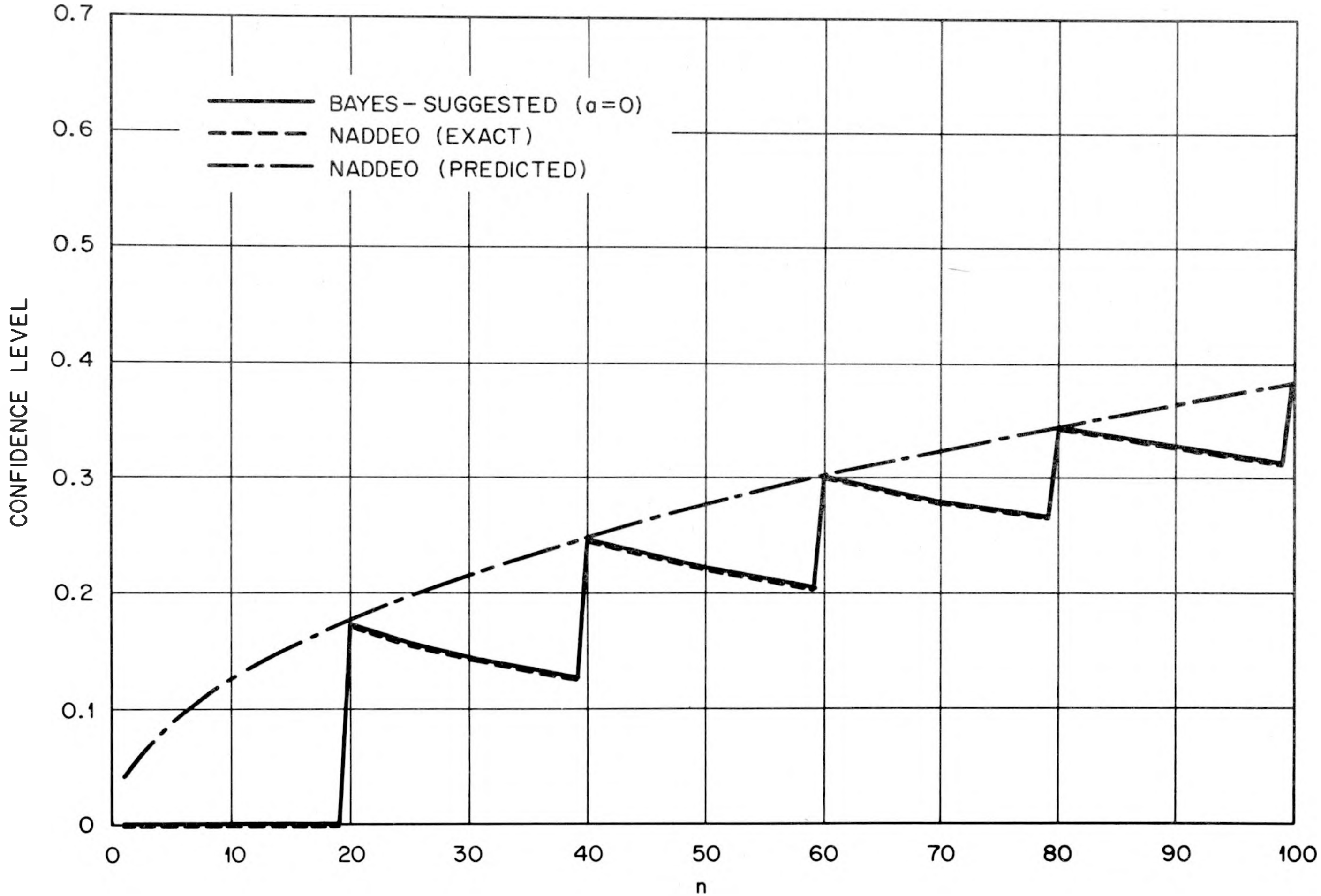

Figure 3.11. Confidence Level for Fixed Width Intervals: Δ = .05

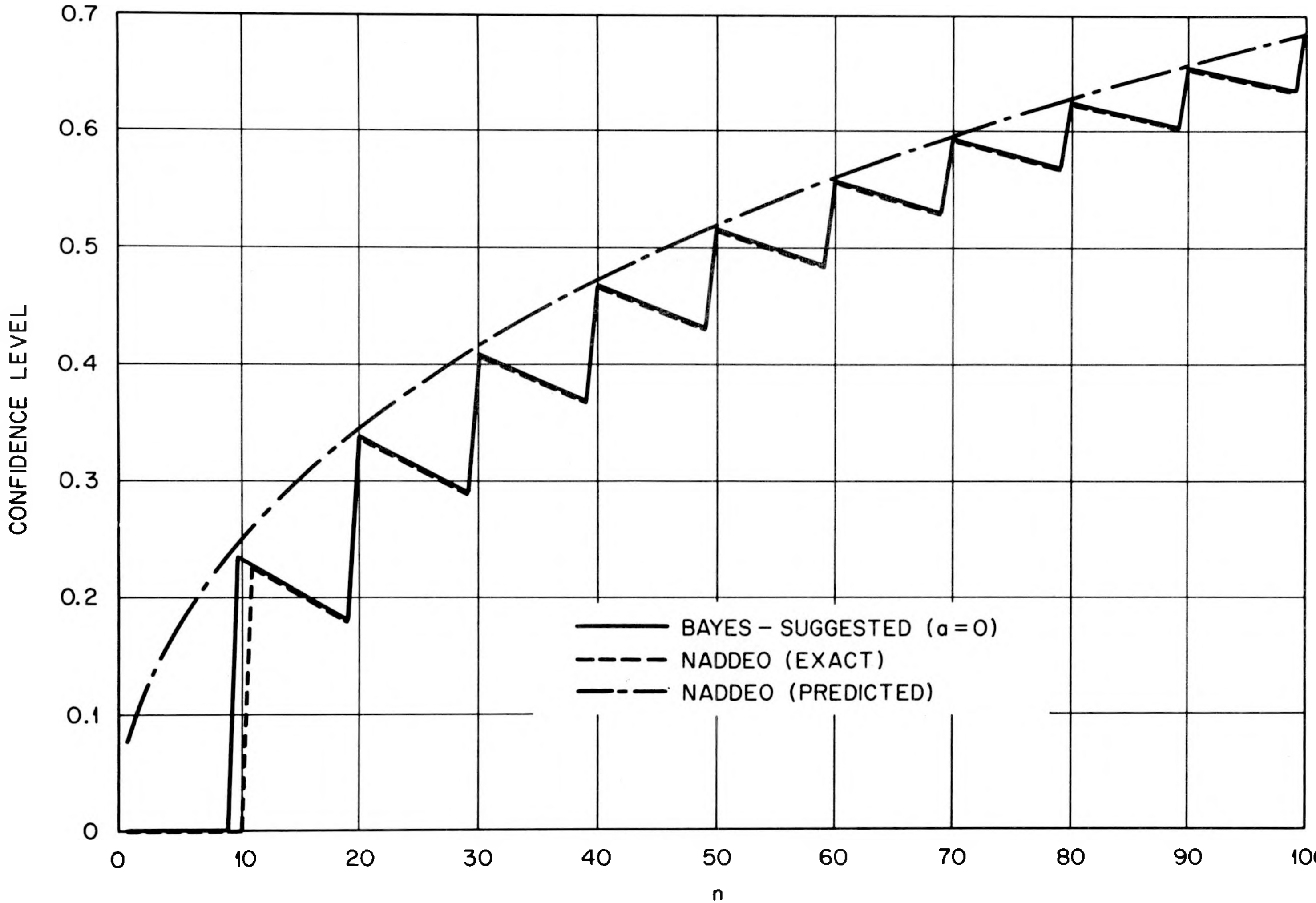

Figure 3.12. Confidence Level for Fixed Width Intervals: $\Delta = .10$

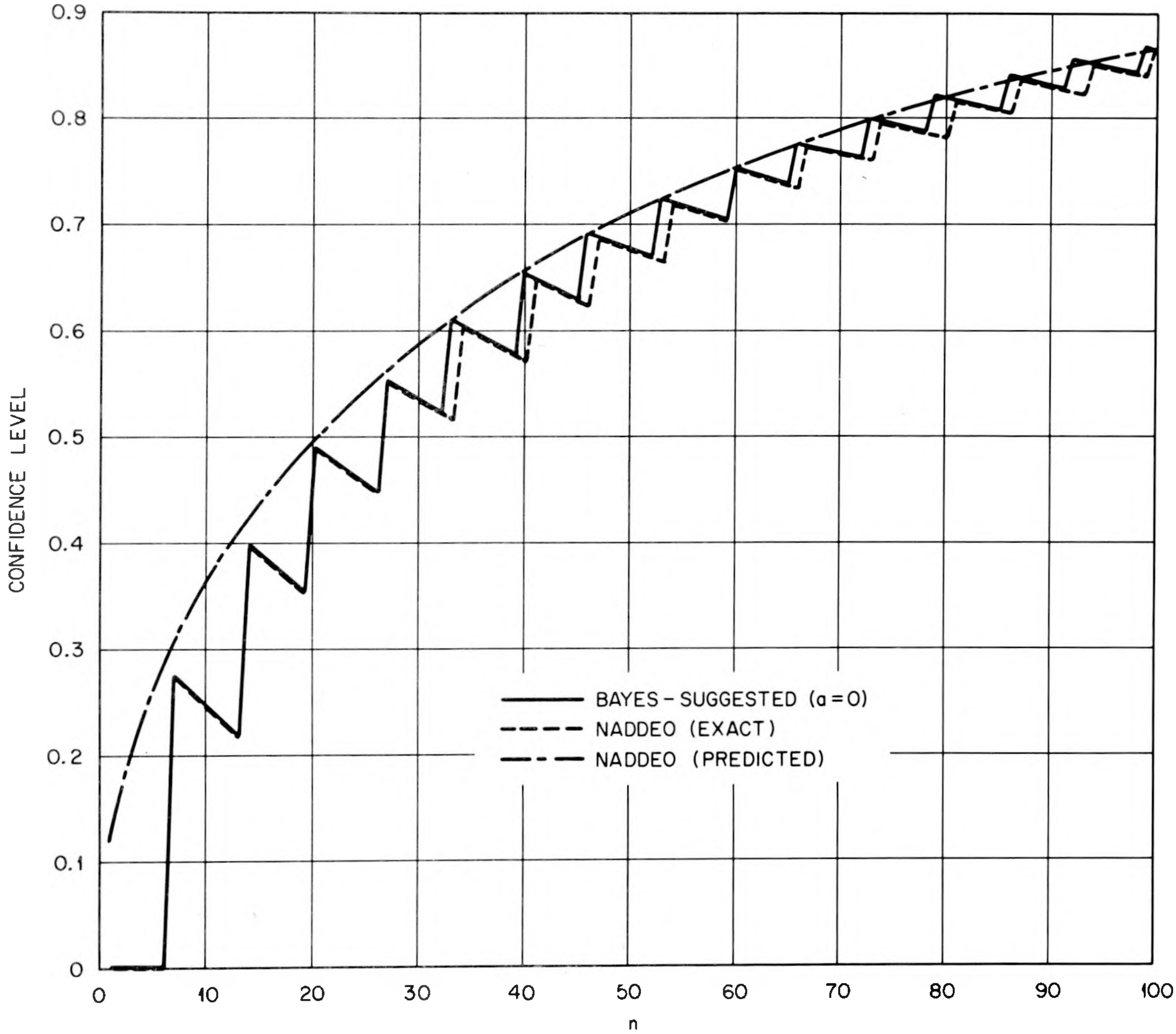

Figure 3.13. Confidence Level for Fixed Width Intervals: $\Delta = .15$

is that the minimum confidence for Naddeo's intervals then appears to be very close to that of the "optimal" Bayes-suggested ($\alpha = 0$) intervals, which are difficult to construct without the aid of a computer.

Figures 3.11, 3.12, and 3.13 indicate that, between "jumps," the confidence level for both sets of intervals decreases as n increases. This unusual behavior can perhaps be best understood by looking at a simple example.

Consider the estimation of θ with intervals of fixed width equal to 0.4. For n = 2, the Bayes-suggested intervals of width 0.4 derived from the uniform prior are given in Figure 3.14.

For this case, the minimum confidence level occurs at $\theta = .3 - \epsilon$, covered only by I_0, and at $\theta = .7 + \epsilon$, covered only by I_2, where ϵ is some arbitrarily small positive number. The confidence level at both of these points is 0.48.

Now consider the set of intervals for n = 3, $\Delta = 0.4$, given in Figure 3.15. The minimum confidence level for this case shifts to $\theta = .446 - \epsilon$, covered only by I_1, and at $\theta = .554 + \epsilon$, covered only by I_2, where the probability of coverage is 0.41.

For n = 4, the set of Bayes-suggested intervals for the uniform prior are given in Figure 3.16. Here the minimum confidence occurs at $\theta = .48 + \epsilon$ and $.52 - \epsilon$, both covered only by I_2, with probability of coverage equal to 0.374.

In each case (n = 2,3,4), the value of θ at which the minimum confidence occurs is covered by an interval corresponding to only one outcome. As shown above, the probability of this particular outcome happens to <u>decrease</u> from .48 to .41 to .374 as n <u>increases</u>.

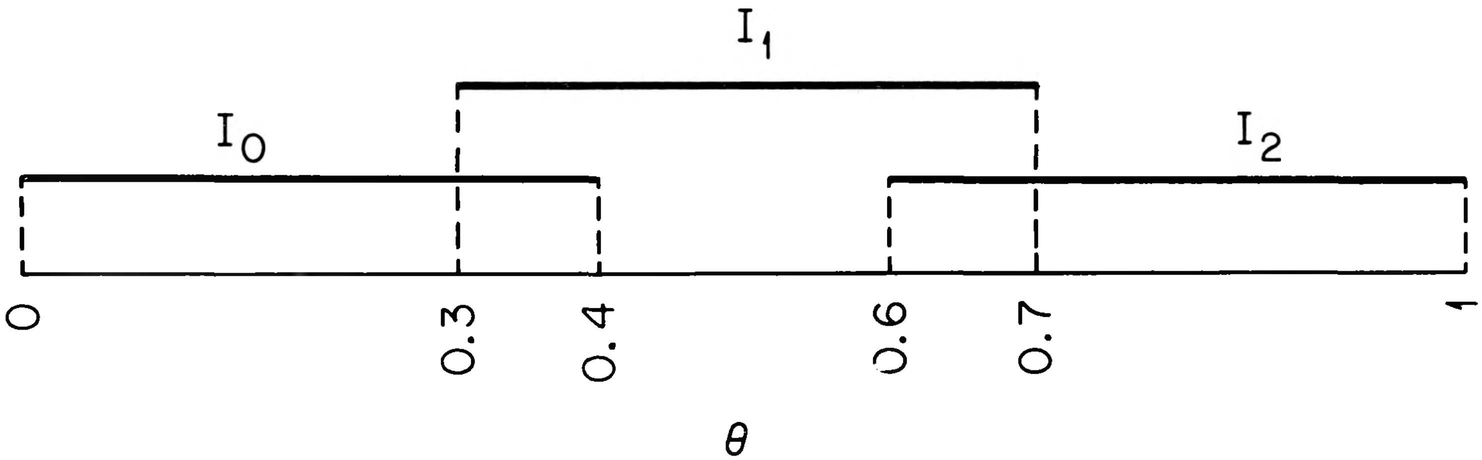

Figure 3.14. Bayes-Suggested Intervals: $n = 2$, $\Delta = .4$

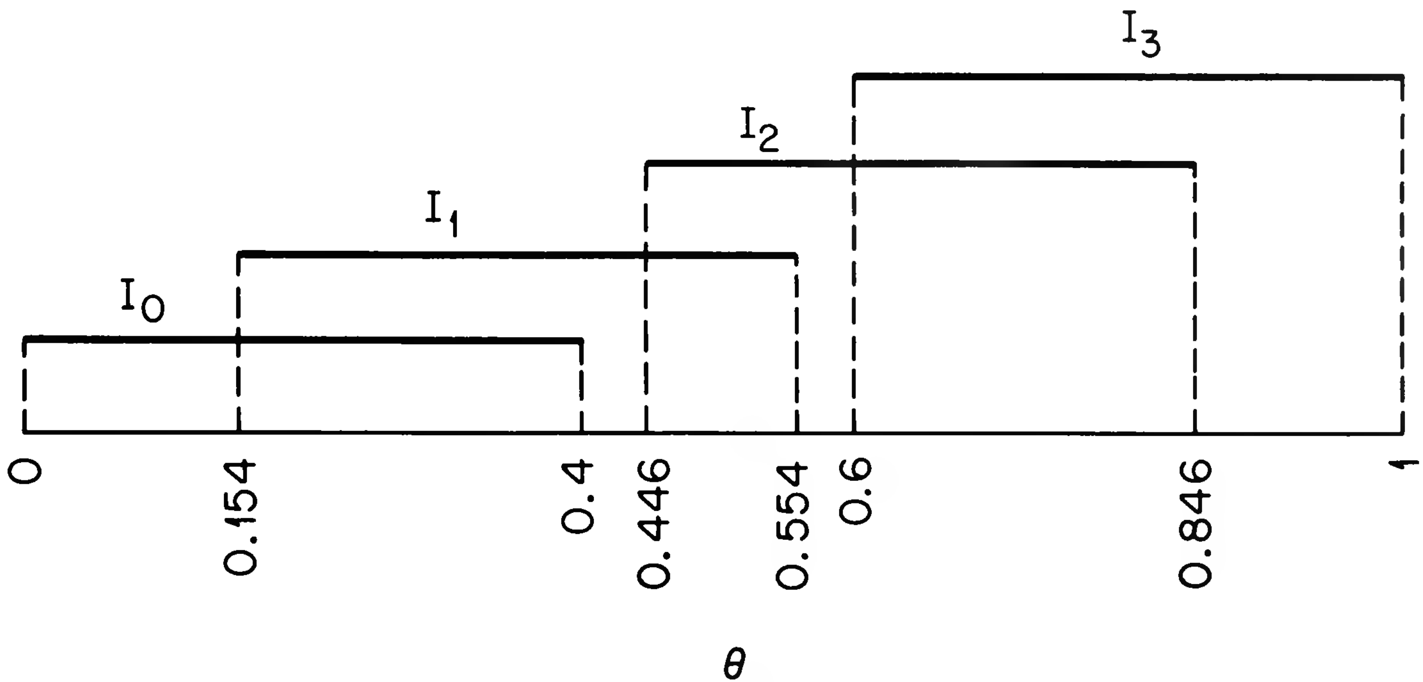

Figure 3.15. Bayes-Suggested Intervals: $n = 3$, $\Delta = .4$

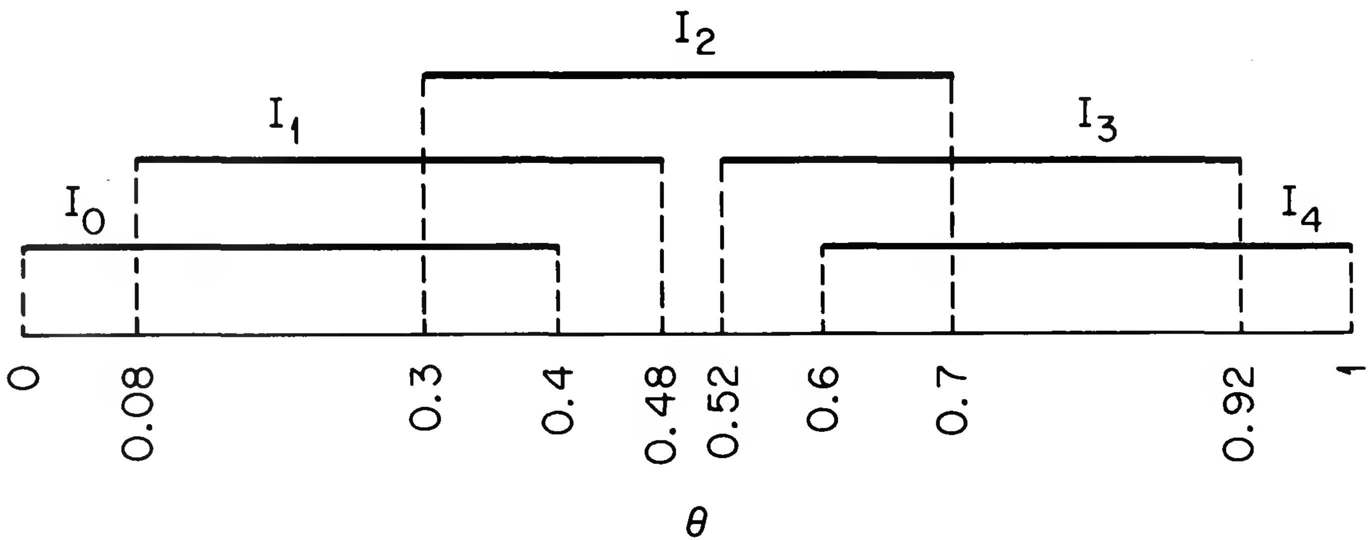

Figure 3.16. Bayes-Suggested Intervals: $n = 4$, $\Delta = .4$

For $n = 5$, the Bayes-suggested intervals are as shown in Figure 3.17, where the minimum confidence occurs at $\theta = .45 + \varepsilon$ and $.549 - \varepsilon$, and is equal to 0.613. The sudden increase in confidence level from .374 to .613 is due to the fact that the values of θ at which the minimum confidence occurs are now covered by two intervals, rather than one. In general, the point of minimum confidence is covered by k intervals for several consecutive values of n and then is suddenly covered by k + 1 intervals, resulting in the "decline and jump" appearance of Figures 3.11, 3.12, and 3.13.

3.6 Summary and Conclusions

In the search for optimum fixed precision estimates of θ, the binomial probability of "success," the class C of estimates derived from symmetric beta prior distributions was considered. The estimates were compared by three criteria, and the optimum set of estimates obtained for each criterion.

In Section 3.4, the sets of estimates which maximize the Bayes risk are given for several values of n and Δ. These estimates have risk functions which, in Figures 3.8, 3.9, and 3.10, indicate that they are fairly good when θ is in a reasonably wide interval around $\theta = 1/2$. However, if θ is near zero or one, this set of estimates has risk equal to one, so their use requires some prior knowledge of the location of θ.

In Sections 3.2 and 3.3, one set of estimates is found that minimizes both the maximum risk and the maximum weighted risk. These optimum estimates are the Bayes estimates derived from the uniform symmetric beta prior.

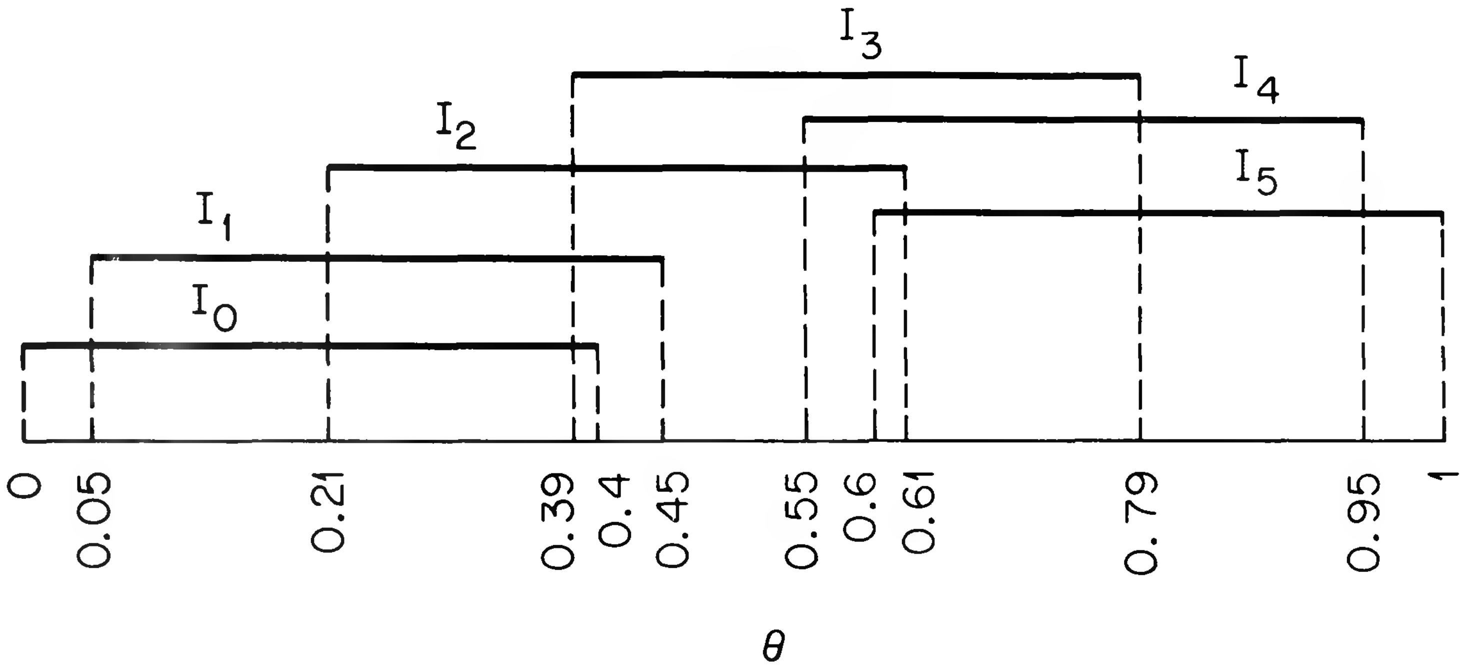

Figure 3.17. Bayes-Suggested Intervals: $n = 5$, $\Delta = .4$

Estimation with fixed precision may also be interpreted as interval estimation with intervals of fixed width. Using this approach, fixed-width confidence intervals with midpoints equal to Bayes-suggested estimates derived from the uniform beta prior are compared with fixed-width confidence intervals defined by midpoints x/n, $x = 0,1,\ldots,n$. This comparison, in Section 3.5, indicates that the Bayes-suggested intervals have a higher confidence level, but the amount of improvement is often not worth the extra trouble involved in finding the Bayes-suggested intervals.

CHAPTER IV. ESTIMATION OF THE LOGIT

The logit of the binomial probability θ is defined to be the natural logarithm of the ratio $\theta/(1-\theta)$. This transformation was introduced by Berkson [3] for application to quantal bioassay, where the response of each experimental unit may be regarded as "0" or "1." Use of the logit transformation has been extended to multidimensional contingency tables (Woolf [23]) and factorial experiments with proportions as observations (Dyke and Patterson [6]). Under the logit transformation, the interval $[0,1]$ is transformed to the real line $[-\infty,\infty]$, which makes it especially suitable for expressing binomial data in terms of a response surface.

Various modifications of the original estimator for the logit proposed by Berkson have been suggested (see, for example, Anscombe [2], Haldane [12], and Hitchcock [13]). Almost all modifications proposed have had as their objective a reduction in bias for special situations. Once an estimator is suggested, its performance is often measured by its bias or variance, and it is usually compared with other estimators using one of these two criteria (see Gart and Zweifel [7]). This suggests that an estimator's overall performance should perhaps be measured by mean squared error = bias^2 + variance. This chapter is concerned with the search for estimators of the logit that are optimum for the squared error loss function. Throughout the chapter, the logit of θ will be designated by λ_θ, i.e.,

$$\lambda_\theta = \ell n \left(\frac{\theta}{1-\theta}\right) . \tag{4.1}$$

4.1 Modified Squared Logit Error

The purpose of this study is to find good estimators of λ_θ when loss is measured by squared error, i.e.,

$$L(\hat{\lambda}(x), \lambda_\theta) = (\hat{\lambda}(x) - \lambda_\theta)^2 \quad \text{for } x = 0,1,\ldots,n \quad , \tag{4.2}$$

where $\hat{\lambda}(x)$ is the estimate of λ_θ when x is the observed value of the binomial random variable, X. Using this "squared logit error" loss function, the risk function is defined as

$$R(\hat{\lambda}, \lambda_\theta) = \sum_{x=0}^{n} L(\hat{\lambda}(x), \lambda_\theta) \; p(x|\theta)$$

$$= \sum_{x=0}^{n} (\hat{\lambda}(x) - \lambda_\theta)^2 \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad . \tag{4.3}$$

As in Chapters II and III, the first goal is to find an estimator that minimizes the maximum risk. However, close inspection of (4.3) reveals that every estimator for λ_θ is minimax, since the maximum risk is always infinitely large. This can be seen by looking at two complementary classes of estimators for λ_θ.

The first class is the set of estimators for which the estimate of λ_θ, when x = 0, is defined to be $-\infty$. This is not an unreasonable estimate to use, since, when $\theta = 0$, $\lambda_\theta = \ell n(0) = -\infty$. However, if $\hat{\lambda}(0) = -\infty$, consider what happens to the risk function in (4.3) at any positive value of θ.

$$R(\hat{\lambda}, \lambda) = (-\infty - \lambda_\theta)^2 \; p(0|\theta) + \sum_{x=1}^{n} (\hat{\lambda}(x) - \lambda_\theta)^2 \; p(x|\theta)$$

$$= \infty \cdot p(0|\theta) + (\text{positive constant}) \quad . \tag{4.4}$$

Since $p(0|\theta) > 0$ if $\theta > 0$, then $R(\hat{\lambda}, \theta) = \infty$.

The alternative class of estimators is composed of those estimators for which $\hat{\lambda}(0) = c$ (any finite number). At $\theta = 0$, the risk function in (4.3) for this class of estimators has value

$$R(\hat{\lambda}, \lambda_0) = (c - \lambda_0)^2 \, p(0|0) = \sum_{x=1}^{n} (\hat{\lambda}(x) - \lambda_0)^2 \, p(x|0) \tag{4.5}$$

$$= (c + \infty)^2 \cdot (1) + \sum_{x=1}^{n} (\hat{\lambda}(x) - \lambda_0)^2 \cdot (0)$$

$$= \infty \; .$$

The same problem occurs at $\theta = 1$ when $\hat{\lambda}(n)$ is finite, and when $\hat{\lambda}(n) = \infty$. We see, then, that the maximum of the risk function in (4.3) is infinity, no matter what estimator is used.

The problem of estimating λ_θ when $x = 0$ or n was encountered by Berkson in his development of the minimum logit χ^2 estimates for the parameters of the logistic model in bioassay. For $x = 1,2,\ldots,n-1$, Berkson uses the logit transformation of the observed proportion of response, x/n, as his estimate of the true logit. This is just the maximum likelihood estimator of λ_θ. However, if $x = 0$ or n, then the logit of the observed response is $\pm\infty$. This does not affect the calculation of Berkson's "logit χ^2" statistic, which essentially ignores observations of $x = 0$ and $x = n$. In order to retain data points at which the response is "all or nothing," Berkson devised his "2n rule" for $x = 0$ and $x = n$. This rule defines $\hat{\lambda}(0) = \ln(\tilde{p}_0/(1 - \tilde{p}_0))$ and $\hat{\lambda}(n) = \ln(\tilde{p}_n/(1 - \tilde{p}_n))$, where $\tilde{p}_0 = 1/2n$ and $\tilde{p}_n = 1 - 1/2n$. For a more detailed discussion of the development of minimum logit χ^2 estimates and Berkson's "2n rule" see DeRouen [5].

In order to estimate λ_θ using the squared logit error loss function and, at the same time, avoid the problems caused by a risk function with infinitely large maximum value, the estimation of λ_θ was studied using the following modified form of the squared logit error loss function:

$$L(\hat{\lambda}(x), \lambda_\theta) = \begin{cases} (\hat{\lambda}(x) - \lambda_\theta)^2 & \text{for } x = 1,2,\ldots,n-1 , \\ 0 & \text{for } x = 0,n . \end{cases} \quad (4.6)$$

This may be viewed as a refusal by the statistician to make a point estimate of λ_θ when x is 0 or n. By assigning zero loss to estimates based on x = 0 or x = n, the problem of an infinitely large risk is avoided. The loss for estimates based on any other value of x is still squared error, so that useful properties of the squared error loss function may still be applied in deriving the appropriate Bayes-suggested estimator.

The study of the estimation of λ_θ, using the "modified squared logit error" loss function in (4.6), was undertaken with several objectives in mind. First of all, it was desired to find the estimators that were optimum with respect to the same three criteria used in previous chapters; namely, minimax risk, minimax weighted risk, and maximum Bayes risk (corresponding to the least favorable prior). Once these optimum estimators had been obtained, the next objective was to compare their risk functions with that of the maximum likelihood estimator of λ_θ. This comparison can be made, not only for the modified squared logit error loss function, but also for the conventional squared logit error loss function (4.3), with two alternatives for the estimation of λ_θ when x = 0 or n. One alternative is to replace estimates based on x = 0 or n, for both the maximum likelihood and Bayes-suggested estimates, with estimates using Berkson's "2n rule." The other alternative is to use the "2n rule" for the maximum likelihood estimator, and to use the Bayes-suggested estimates for $x = 0,1,2,\ldots,n$ for the other estimator, since the Bayes-suggested estimator does define estimates of λ_θ when $x = 0$ and n.

As in previous chapters, the estimators under consideration were restricted to those in the class of Bayes estimators of λ_θ, derived from symmetric beta prior distributions defined on $\Theta = [0,1]$ (i.e., the class C of SBP estimators). This approach produced a new and interesting class of estimators which can be evaluated rather easily. From this class, optimum estimators were chosen for each of the three criteria proposed in Chapter I.

4.2 The Form of an SBP Estimator for λ_θ

The risk function for any estimator, $\hat{\lambda}$, of the logit λ_θ, and for any loss function, is defined to be

$$R(\hat{\lambda}, \lambda_\theta) = \sum_{x=0}^{n} L(\hat{\lambda}(x), \lambda_\theta)\ p(x|\theta) \quad .$$

For the modified squared logit error loss function defined in (4.6), the corresponding risk function is

$$R(\hat{\lambda}, \lambda_\theta) = \sum_{x=1}^{n-1} (\hat{\lambda}(x) - \lambda_\theta)^2 \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad . \tag{4.7}$$

To obtain SBP estimates, the prior distribution of θ is defined to be the symmetric beta function

$$\xi(\theta) = \frac{\theta^{a}(1-\theta)^{a}}{B(a+1,a+1)} \quad , \qquad a > -1 \quad . \tag{4.8}$$

Then $\hat{\lambda}_{\xi}$, the Bayes estimator for λ_{θ} corresponding to $\xi(\theta)$, is defined to be that set of estimates which minimizes the expected or weighted risk, when the risk function is weighted by the prior, $\xi(\theta)$, i.e.,

$$\int_0^1 R(\hat{\lambda}_{\xi}, \lambda_{\theta})\ \xi(\theta)\ d\theta = \inf_{\hat{\lambda}} \int_0^1 R(\hat{\lambda}, \lambda_{\theta})\ \xi(\theta)\ d\theta \quad .$$

For any estimator $\hat{\lambda} = \{\hat{\lambda}(0), \hat{\lambda}(1), \ldots, \hat{\lambda}(n)\}$, and for the modified squared logit error risk function in (4.7), the weighted risk may be written

$$\int_0^1 \left[\sum_{x=1}^{n-1} (\hat{\lambda}(x) - \lambda_{\theta})^2\ p(x|\theta)\right] \xi(\theta)\ d\theta \quad .$$

But this may also be written as

$$\sum_{x=1}^{n-1} \left[\int_0^1 (\hat{\lambda}(x) - \lambda_{\theta})^2\ \xi(\theta|x)\ d\theta\right] p(x) \quad , \tag{4.9}$$

so that

$$\inf_{\hat{\lambda}} \int_0^1 \left[\sum_{x=1}^{n-1} (\hat{\lambda}(x) - \lambda_{\theta})^2\ p(x|\theta)\right] \xi(\theta)\ d\theta$$

$$= \sum_{x=1}^{n-1} \left[\inf_{\hat{\lambda}(x)} \int_0^1 (\hat{\lambda}(x) - \lambda_{\theta})^2\ \xi(\theta|x)\ d\theta\right] p(x) \quad .$$

Therefore, the SBP estimator, $\hat{\lambda}_{\xi}$, is defined by

$$\hat{\lambda}_\xi = \{\hat{\lambda}_\xi(x) \ , \ x = 1,2,\ldots,n-1\} \tag{4.10}$$

where $\hat{\lambda}_\xi(x)$ is the estimate that minimizes the posterior risk, i.e., $\hat{\lambda}_\xi(x)$ minimizes

$$\int_0^1 (\hat{\lambda}(x) - \lambda_\theta)^2 \ \xi(\theta|x) \ d\theta \ , \quad \text{for } x = 1,2,\ldots,n-1. \tag{4.11}$$

Thus, even though a modified squared logit error loss function is being used, the individual estimates for $x = 1,2,\ldots,n-1$, are those estimates which minimize the posterior expected squared logit error. From the properties of squared error loss functions, it is known (see Lehmann [16], Chapter 4, p. 31) that the mean of the posterior distribution is the value of the estimate which minimizes the posterior expected squared error, so that

$$\hat{\lambda}_\xi(x) = \int_0^1 \lambda_\theta \ \xi(\theta|x) \ d\theta \quad . \tag{4.12}$$

If the prior distribution is $\xi(\theta)$ as given in (4.8), and the conditional distribution of the random variable X is binomial with parameters n and θ, then the posterior distribution, $\xi(\theta|x)$, is given by

$$\xi(\theta|x) = \frac{\theta^{\alpha+x} \ (1 - \theta)^{\alpha+n-x}}{B(\alpha+x+1, \alpha+n-x+1)} \quad . \tag{4.13}$$

The posterior expectation of $\lambda_\theta = \ell n \left(\frac{\theta}{1-\theta}\right)$ can be found by first obtaining the characteristic function corresponding to the posterior distribution of λ_θ. If $y = \lambda_\theta$, then

$$e^{iyt} = \left(\frac{\theta}{1 - \theta}\right)^{it} \quad ,$$

and the posterior characteristic function of y is

$$\varphi_y(t) = \mathcal{E}_{(\theta|x)}(e^{iyt}) = \int_0^1 \frac{\left(\frac{\theta}{1-\theta}\right)^{it} \theta^{\alpha+x} (1-\theta)^{\alpha+n-x} \, d\theta}{B(\alpha+x+1, \alpha+n-x+1)} \tag{4.14}$$

$$= \frac{B(\alpha+x+it+1, \alpha+n-x-it+1)}{B(\alpha+x+1, \alpha+n-x+1)}$$

$$= \frac{\Gamma(\alpha+x+it+1) \cdot \Gamma(\alpha+n-x-it+1)}{\Gamma(2\alpha+n+2) \cdot B(\alpha+x+1, \alpha+n-x+1)} , \tag{4.15}$$

where $\Gamma(p)$ is the gamma function of p.

The cumulant generating function of $y(= \lambda_\theta)$, $\eta_y(t)$, is equal to the natural logarithm of the characteristic function, $\varphi_y(t)$. Expressing $\varphi_y(t)$ as in (4.15),

$$\eta_y(t) = \ell n \, \Gamma(\alpha+x+it+1) + \ell n \, \Gamma(\alpha+n-x-it+1)$$

$$- \ell n \, \Gamma(2\alpha+n+2) - \ell n \, B(\alpha+x+1, \alpha+n-x+1) . \tag{4.16}$$

The Bayes estimates of λ_θ are equal to the posterior expectations of λ_θ, for $x = 1,2,\ldots,n-1$. These posterior expectations, in turn, are equal to $\frac{1}{i} \left. \frac{d \, \eta_y(t)}{dt} \right|_{t=0}$, for $x = 1,2,\ldots,n-1$. The Bayes estimates may then be expressed as

$$\hat{\lambda}_\xi(x) = \frac{1}{i} \left\{ i \, \psi(\alpha+x+it+1) \Big|_{t=0} - i \, \psi(\alpha+n-x-it+1) \Big|_{t=0} \right\} \tag{4.17}$$

$$= \psi(\alpha+x+1) - \psi(\alpha+n-x+1) , \tag{4.18}$$

where $\psi(p) = \frac{d}{dp} \ell n \, \Gamma(p)$ as defined in [9], p. 943. Since all estimates of λ_θ considered in this chapter are Bayes (SBP) estimates, for notational convenience the subscript ξ will not be included in future expressions for estimates of λ_θ, i.e., $\hat{\lambda}(x) = \hat{\lambda}_\xi(x)$.

Evaluation of the Bayes estimates given in (4.18) may be accomplished by using the following recursive relation, given on p. 945 of [9]:

$$\psi(p + 1) = \psi(p) + \frac{1}{p} \quad . \tag{4.19}$$

Thus,

$$\psi(\alpha+x+1) - \psi(\alpha+n-x+1) = \psi(\alpha+1) + \frac{1}{\alpha+1} + \ldots + \frac{1}{\alpha+x} - \psi(\alpha+1) - \frac{1}{\alpha+1} - \ldots - \frac{1}{\alpha+n-x}$$

and, for $1 \le x \le n-1$,

$$\begin{aligned} \hat{\lambda}(x) &= \psi(\alpha+x+1) - \psi(\alpha+n-x+1) \\ &= \sum_{j=1}^{x} \frac{1}{\alpha + j} - \sum_{j=1}^{n-x} \frac{1}{\alpha + j} \quad . \end{aligned} \tag{4.20}$$

Although estimates based on $x = 0$ and $x = n$ are not of interest for the modified squared logit error loss function, if the unmodified squared logit error loss function is employed, the estimates based on all values of x, including $x = 0$ and $x = n$, may be given by (4.18). In this case, evaluation of the estimates based on $x = 0$ and $x = n$ may be evaluated using

$$\begin{aligned} \hat{\lambda}(0) &= \psi(\alpha+1) - \psi(\alpha+n+1) \\ &= - \sum_{j=1}^{n} \frac{1}{\alpha + j} \quad , \end{aligned} \tag{4.21}$$

$$\begin{aligned} \hat{\lambda}(n) &= \psi(\alpha+n+1) - \psi(\alpha+1) \\ &= \sum_{j=1}^{n} \frac{1}{\alpha + j} \quad . \end{aligned} \tag{4.22}$$

Expressions (4.21) and (4.22) may be considered as Bayes-suggested alternatives to Berkson's "2n rule."

4.3 Estimates Which Minimize the Maximum Risk

For prior distributions of the form (4.8), let a^* denote the value of its parameter defining the distribution whose Bayes estimates of λ_θ are C-minimax. In other words, the SBP estimates based on $a = a^*$ are those estimates in the class C of SBP estimates that minimize the maximum value of the modified squared logit risk function in (4.7). For selected values of the sample size, n, a numerical search was conducted to determine a^*. At fixed values of n, estimates based on a specified value of a were calculated using (4.20), and the corresponding risk as a function of θ was calculated using (4.7).

The maximum risk, for fixed values of n and a, was obtained by a direct search, in which the risk function (4.7) was evaluated from $\theta = 0$ to 0.5 in steps of 1×10^{-3}. The risk function is symmetric about $\theta = .5$, allowing the search for maximum to be restricted to the interval [0,.5]. The risk was not always found to be a unimodal function of θ, and, for several combinations of n and a, two or more local maxima were located. However, the maximum risk, as a function of the parameter a for fixed n, was found to have a single minimum. This optimum value of a, which yields the minimax risk, was calculated to three decimal places.

Table V gives the results of this numerical search for a^*, the value of the parameter yielding C-minimax SBP estimates. Also given is the maximum value of the modified squared logit error risk function when maximum likelihood estimates are used for $x = 1,2,\ldots,n-1$.

Table V

Loss: Modified Squared Logit Error

Criterion: Minimax Risk

n	a^*	C-Minimax Risk	Maximum Risk for Maximum Likelihood Estimator
3	-.936	.6626	.7983
4	-.641	.6727	.7098
5	-.512	.6799	.7033
6	-.473	.6702	.7109
7	-.487	.6466	.6719
8	-.532	.6138	.6148
9	-.592	.5759	.5988
10	-.640	.5447	.5916
11	-.668	.5253	.5859
12	-.686	.5114	.5812
13	-.699	.5008	.5772
14	-.708	.4924	.5740
15	-.715	.4857	.5712
16	-.721	.4804	.5687
25	-.747	.4541	.5557
36	-.756	.4418	.5492
49	-.760	.4338	.5449
64	-.765	.4306	.5423
81	-.766	.4280	.5372
100	-.767	.4260	.5326

a^* = value of a corresponding to the set of SBP estimates of the form in (4.18) having minimax risk.

Note in Table V that the maximum risks for both the C-minimax SBP estimator and the maximum likelihood estimator decrease only slightly for $n > 25$. Although the risk functions for both estimators tend to decrease substantially over a wide range of Θ as n gets large, the maximum risk evidently decreases very little for n greater than 25. Thus far, a reason for this behavior has not been found.

The behavior of a^* and the maximum risk as a function of n, for $2 \leq n \leq 6$, is also very strange. First of all, a^* increases to a maximum of -.473 at $n = 6$, and decreases from then on. What is really disturbing is that the minimax risk, for both the C-minimax SBP estimator and the maximum likelihood estimator, actually increases as n increases for some values of n in this range. The only explanation for this erratic behavior is that it must be due to the modification performed on the squared error loss function.

4.4 Estimates Which Minimize the Maximum Weighted Risk

In this section, weighted risk is the criterion by which an estimator is judged. The risk function is again modified squared logit error, and the class of estimators considered is again the class C of SBP estimates (the class of Bayes estimators for λ_θ derived from symmetric beta priors). The class of weighting functions, as before, is also restricted to the class of symmetric beta functions, $\omega(\theta)$, defined on $\Theta = [0,1]$, where

$$\omega(\theta) = \frac{\theta^b (1-\theta)^b}{B(b+1, b+1)} , \qquad b > -1 , \; 0 \leq \theta \leq 1 . \tag{4.23}$$

If an estimator is based on the prior, $\xi(\theta)$, it is made up of the set of estimates

$$\hat{\lambda} = \{\hat{\lambda}(x) \;,\; x = 1,2,\ldots,n-1\} \;,$$

where

$$\hat{\lambda}(x) = \mathcal{E}^{(a)}_{(\theta|x)}(\lambda_\theta) \;. \tag{4.24}$$

This expression differs slightly from that in (4.12) in that the superscript (a) is included to show that the expectation of λ_θ is taken over the posterior distribution of θ, derived when $\xi(\theta)$ (with parameter a) is the prior. Similarly, the variance of λ_θ in the posterior distribution based on $\xi(\theta)$ will be denoted

$$\mathrm{Var}^{(a)}_{(\theta|x)}(\lambda_\theta) \;.$$

This extra notation is necessary because, in deriving the weighted risk for a set of estimates based on $\xi(\theta)$ whose risk function is weighted by $\omega(\theta)$, expressions will be encountered which can be viewed as the mean and variance of λ_θ for a distribution that could be called a posterior distribution, if $\omega(\theta)$ were considered a prior distribution. The mean and variance of λ_θ for a posterior distribution derived by treating $\omega(\theta)$ (with parameter b) as a prior will be denoted by $\mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta)$ and $\mathrm{Var}^{(b)}_{(\theta|x)}(\lambda_\theta)$, respectively.

The risk function for an SBP estimator, $\hat{\lambda}$, is given in (4.7). The weighted risk, when the risk function is weighted by $\omega(\theta)$, is given by

$$\widetilde{R}(\xi,\omega) = \int_0^1 \left[\sum_{x=1}^{n-1} (\hat{\lambda}(x) - \lambda_\theta)^2 \binom{n}{x} \theta^x (1-\theta)^{n-x}\right] \frac{\theta^{b}(1-\theta)^{b}\, d\theta}{B(b+1,b+1)} \qquad (4.25)$$

$$= \frac{1}{B(b+1,b+1)} \sum_{x=1}^{n-1} \left[\binom{n}{x} \int_0^1 (\hat{\lambda}(x) - \lambda_\theta)^2\, \theta^{b+x}(1-\theta)^{b+n-x}\, d\theta\right] . \qquad (4.26)$$

Now, add and subtract the constant, $\mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta)$, inside the parentheses of the squared logit error term in (4.26), yielding

$$\widetilde{R}(\xi,\omega) = \frac{1}{B(b+1,b+1)} \sum_{x=1}^{n-1} \binom{n}{x} \int_0^1 \theta^{b+x}(1-\theta)^{b+n-x}$$

$$\left\{\left[\hat{\lambda}(x) - \mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta)\right] + \left[\mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta) - \lambda_\theta\right]\right\}^2 d\theta . \qquad (4.27)$$

Since

$$\int_0^1 \theta^{b+x}(1-\theta)^{b+n-x}\, d\theta = B(b+x+1, b+n-x+1) \quad ,$$

$$\int_0^1 \lambda_\theta\, \theta^{b+x}(1-\theta)^{b+n-x}\, d\theta = B(b+x+1, b+n-x+1) \cdot \mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta) \quad ,$$

and

$$\int_0^1 \theta^{b+x}(1-\theta)^{b+n-x}\left[\lambda_\theta - \mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta)\right]^2 d\theta = B(b+x+1, b+n-x+1)\ \mathrm{Var}^{(b)}_{(\theta|x)}(\lambda_\theta),$$

expression (4.27) reduces to

$$\widetilde{R}(\xi,\omega) = \frac{1}{B(b+1,b+1)} \sum_{x=1}^{n-1} \binom{n}{x} B(b+x+1, b+n-x+1)$$

$$\left\{\left[\hat{\lambda}(x) - \mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta)\right]^2 + \mathrm{Var}^{(b)}_{(\theta|x)}(\lambda_\theta)\right\} . \qquad (4.28)$$

This expression for the weighted risk, when estimates are based on the prior $\xi(\theta)$, and the risk function is weighted by the function $\omega(\theta)$, is now relatively easy to evaluate. It involves the evaluation of complete beta functions, the estimates $\{\hat{\lambda}(x),\ x = 1,2,\ldots,n-1\}$, and the mean and variance of λ_θ from the posterior distribution of θ, when the prior is taken to be $\omega(\theta)$ (with parameter b). The beta functions were again evaluated using the algorithm suggested by Amos [1], and the estimates were calculated using (4.20).

Expressions for the posterior mean and variance of λ_θ, based on the prior density $\omega(\theta)$, were obtained as follows. If $\tau_y(t)$ denotes the posterior cumulant generating function of $y = \lambda_\theta$ when $\omega(\theta)$ is the prior, then, from (4.16),

$$\tau_y(t) = \ell n\ \Gamma(b+x+it+1) + \ell n\ \Gamma(b+n-x-it+1) - \ell n\ \Gamma(2b+n+2) - \ell n\ B(b+x+1, b+n-x+1) \quad . \tag{4.29}$$

By definition of the cumulant generating function,

$$\mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta) = \frac{1}{i}\left.\frac{d\ \tau_y(t)}{dt}\right|_{t=0} , \tag{4.30}$$

and

$$\mathrm{Var}^{(b)}_{(\theta|x)}(\lambda_\theta) = \frac{1}{i^2}\left.\frac{d^2\ \tau_y(t)}{dt^2}\right|_{t=0} . \tag{4.31}$$

Again employing the function $\psi(p) = \frac{d}{dp}\ \ell n\ \Gamma(p)$,

$$\mathcal{E}^{(b)}_{(\theta|x)}(\lambda_\theta) = \frac{1}{i}\left\{i\ \psi(b+x+it+1)\Big|_{t=0} - i\ \psi(b+n-x-it+1)\Big|_{t=0}\right\}$$

$$= \psi(b+x+1) - \psi(b+n-x+1) \quad ; \tag{4.32}$$

$$\mathrm{Var}^{(b)}_{(\theta|x)}(\lambda_\theta) = \frac{1}{i^2}\left\{i^2\ \psi'(b+x+it+1)\Big|_{t=0} + i^2\ \psi'(b+n-x-it+1\Big|_{t=0}\right\}$$

$$= \psi'(b+x+1) + \psi'(b+n-x+1) \qquad (4.33)$$

where $\psi'(p) = \frac{d}{dp}\psi(p)$.

Numerical evaluation of (4.32) is accomplished using (4.19), so that

$$\delta^{(b)}_{(\theta|x)}(\lambda_\theta) = \sum_{j=1}^{x}\frac{1}{b+j} - \sum_{j=1}^{n-x}\frac{1}{b+j} \quad . \qquad (4.34)$$

To obtain a similar expression for the posterior variance, first differentiate both sides of (4.19):

$$\psi'(p+1) = \psi'(p) - \frac{1}{p^2} \quad . \qquad (4.35)$$

The application of this recursive relation to (4.33) yields the expression:

$$\mathrm{Var}^{(b)}_{(\theta|x)}(\lambda_\theta) = 2\psi'(b+1) - \sum_{j=1}^{x}\frac{1}{(b+j)^2} - \sum_{h=1}^{n-x}\frac{1}{(b+h)^2} \ . \qquad (4.36)$$

From p. 944 of [9]

$$\psi'(b+1) = \sum_{m=0}^{\infty}\frac{1}{(b+1+m)^2} \quad . \qquad (4.37)$$

Evaluation of $\psi'(b+1)$ was accomplished by truncating the infinite series in (4.37) at some finite value, M. Observe that

$$\int_M^\infty \frac{dm}{(b+1+m)^2} > \sum_{m=M+1}^{\infty}\frac{1}{(b+1+m)^2}$$

$$\Rightarrow \frac{1}{b+1+M} > \sum_{m=M+1}^{\infty}\frac{1}{(b+1+m)^2} \qquad (4.38)$$

where the right-hand term in (4.38) is the error involved in evaluating $\psi'(b + 1)$ by the infinite series in (4.37), truncated at M. Since $b > -1$, accuracy to three decimal places was obtained by truncating the series at M = 1500.

The weighted risk, for prior $\xi(\theta)$ and weighting function $\omega(\theta)$, was evaluated using (4.28), in conjunction with (4.20), (4.34, (4.36), and (4.37).

For every prior $\xi(\theta)$ (indexed by the value of its parameter, a), the weighting function $\omega^*(\theta)$ (indexed by its parameter, b^*) that maximizes the weighted risk, $\widetilde{R}(\xi,\omega)$ was obtained numerically, using a direct search method. From this set of maximum weighted risks, $\{\widetilde{R}(\xi,\omega^*)\}$, the value of a that minimizes $\widetilde{R}(\xi,\omega^*)$, was selected, again by direct search. If we denote this optimum value of a by a^* and the corresponding symmetric beta prior by $\xi^*(\theta)$, then $\widetilde{R}(\xi^*,\omega^*)$ is the minimax weighted risk.

Table VI contains the results of the search for the estimates having minimax weighted risk for several values of n. For each n, the following are given:

(i) a^*, the optimum value of the parameter of the symmetric beta prior;

(ii) b^*, the value of the parameter in the symmetric beta weighting function that produces the maximum weighted risk for the set of estimates based on the prior with parameter a^*;

(iii) the minimax weighted risk, $\widetilde{R}(\xi^*,\omega^*)$.

Note that, as n increases, the minimax weighted risk decreases at a faster rate than the minimax risk in Table V. Since the optimum values of $a(a^*)$ for minimax weighted risk appear to approach the optimum

Table VI

Loss: Modified Squared Logit Error

Criterion: Minimax Weighted Risk

n	a^*	b^*	Minimax Weighted Risk
4	-.31	-.31	.545435
9	-.16	-.5	.450166
16	-.5	-.5	.349908
25	-.6	-.6	.291609
36	-.66	-.66	.254163
49	-.70	-.70	.228572
64	-.73	-.73	.210063
81	-.75	-.75	.196065
100	-.76	-.76	.185063

a^* = value of a in (4.18) yielding minimax weighted risk SBP estimates.

b^* = value of b defining the weighting function in (4.23) that maximizes the weighted risk for the SBP estimator based on $a = a^*$.

values of a for minimax risk as n gets large, the discrepancy between the values of the minimax risk and minimax weighted risk and the difference in their rates of decrease suggest a peculiar behavior for the risk functions. These differences in minimax risk and minimax weighted risk indicate that, as n gets large, the risk functions for the estimators based on a^* (optimum a for minimax risk or weighted risk) decrease substantially over a wide range of Θ, but, in some small interval, the risk increases to a maximum value that changes very little for $n > 25$. This observation was noted at the end of Section 4.3, but the explanation of its origin was omitted. As indicated before, the cause of this somewhat unusual behavior is not known.

It was also pointed out in Section 4.3 that the optimum value of a for minimax risk, as a function of n, increased to a maximum value at $n = 6$ and decreased from then on. It is interesting to note from Table VI that a^* as a function of n has maximum value at $n = 9$, and that $n = 9$ is the only case where $a^* \neq b^*$. This unusual behavior again has not been linked with any known cause, except that it is probably due to the use of the modified loss function. Because of this unusual behavior the values of a^* and b^* for all values of n between 2 and 16 would be of interest. However, because of the amount of additional computer time required to obtain these values of a^* and b^*, it was decided that the question would not be pursued further.

4.5 Estimates Based on the Least Favorable Prior

In order to find the least favorable prior, $\xi^*(\theta)$, in the class of symmetric beta priors, it is necessary to find the prior that has maximum

Bayes risk, $\widetilde{R}(\xi^*, \xi^*)$. For any prior, $\xi(\theta)$, the Bayes risk, $\widetilde{R}(\xi,\xi)$, is just the weighted risk, when the weighting function is the prior.

The Bayes risk for a prior $\xi(\theta)$ (indexed by parameter a) may be determined from expression (4.28) by replacing the parameter of the weighting function (b) with that of the prior (a). Thus,

$$\widetilde{R}(\xi,\xi) = \frac{1}{B(a+1,a+1)} \sum_{x=1}^{n-1} \binom{n}{x} B(a+x+1, a+n-x+1) \left\{ \left[\hat{\lambda}(x) - \mathcal{E}^{(a)}_{(\theta|x)}(\lambda_\theta) \right]^2 + \mathrm{Var}^{(a)}_{(\theta|x)}(\lambda_\theta) \right\} . \tag{4.39}$$

But, from equation (4.24)

$$\hat{\lambda}(x) = \mathcal{E}^{(a)}_{(\theta|x)}(\lambda_\theta)$$

so that

$$\widetilde{R}(\xi,\xi) = \frac{1}{B(a+1,a+1)} \sum_{x=1}^{n-1} \binom{n}{x} B(a+x+1, a+n-x+1)\ \mathrm{Var}^{(a)}_{(\theta|x)}(\lambda_\theta) . \tag{4.40}$$

This expression for the Bayes risk can be evaluated numerically, for given a, using expressions (4.36) and (4.37) for $\mathrm{Var}^{(a)}_{(\theta|x)}(\lambda_\theta)$; i.e.,

$$\mathrm{Var}^{(a)}_{(\theta|x)}(\lambda_\theta) = 2\psi'(a+1) - \sum_{j=1}^{x} \frac{1}{(a+j)^2} - \sum_{h=1}^{n-x} \frac{1}{(a+h)^2} \tag{4.41}$$

where

$$\psi'(a+1) = \sum_{m=0}^{\infty} \frac{1}{(a+1+m)^2} , \tag{4.42}$$

evaluated to 1500 terms.

The value of α that maximizes (4.40) was determined by direct search for several values of n. In each case, the solution $\alpha = \alpha^*$ identifies the least favorable symmetric beta prior for the estimation of λ_θ. The results are given in Table VII.

4.6 Comparisons With Maximum Likelihood Estimator

The purpose of this section is to present a brief comparison of the "optimum" estimates derived in Sections 4.2 and 4.3 with the maximum likelihood (m.l.) estimator $\ell n\left(\frac{x}{n - x}\right)$. The risk functions of these estimators will be examined for two cases: n = 16 and n = 100.

In Figure 4.1 the risk functions of the estimators, from the class C of SBP estimators, that minimize the maximum risk and maximum weighted risk, are plotted for the case n = 16. Also plotted is the risk function for the m.l. estimator. In this figure, the loss function is modified squared logit error, for which the optimum SBP estimators were developed. Note the relative proximity of the risk functions of the m.l. estimator and the SBP estimator based on $\alpha = -.5$ (minimax weighted risk).

In subsequent comparisons, conventional squared logit error will be the loss function employed. Since this loss function assigns non-zero loss to estimates based on x = 0 and x = n, these estimates are of key importance. For the m.l. estimator, the "2n rule" will be used to determine these estimates, as suggested by Berkson and discussed earlier in this chapter. When the "2n rule" is also used for the "optimum" SBP estimates, the risk functions are as shown in Fig. 4.2. The three risk functions are very close, especially the m.l. and minimax weighted risk ($\alpha = -.5$) SBP estimators. It seems, then, that this attempt to

Table VII

Loss: Modified Squared Logit Error

Criterion: Least Favorable Prior

n	a^*	Maximum Bayes Risk
4	-.31	.545435
9	-.36	.4424695
16	-.50	.349908
25	-.60	.291609
36	-.66	.254163
49	-.70	.228572
64	-.73	.210063
81	-.75	.196065
100	-.76	.185063

a^* = value of a in (4.18) giving SBP estimates from least favorable prior.

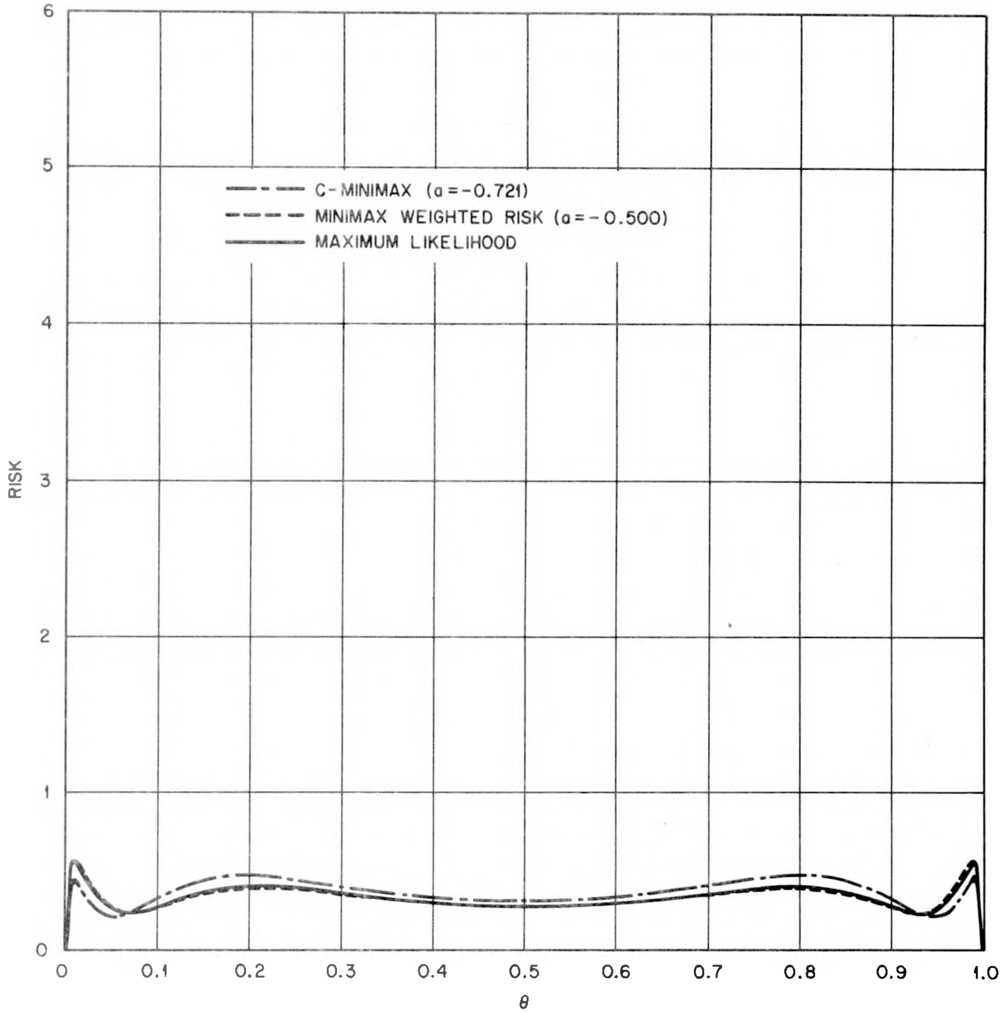

Figure 4.1. Risk Functions for Modified Squared Logit Error: n = 16

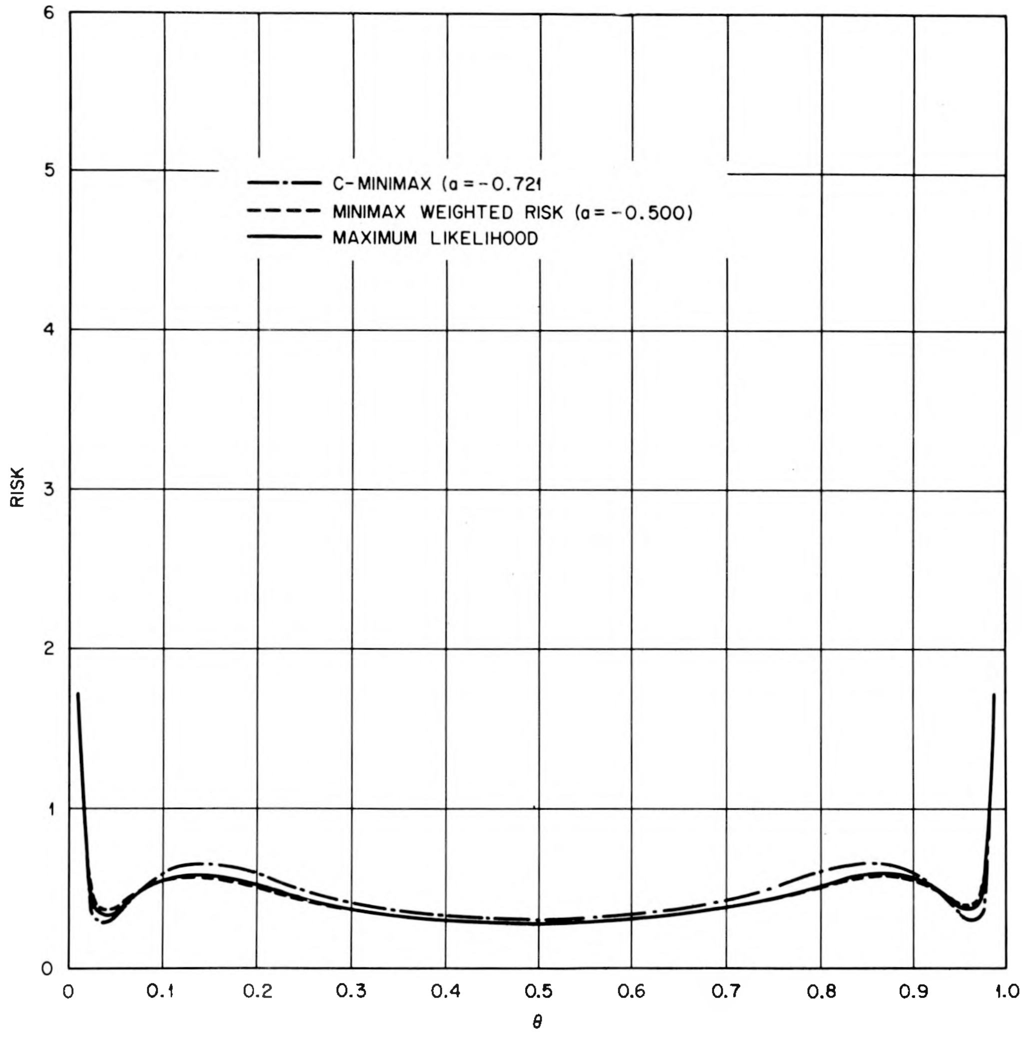

Figure 4.2. Risk Functions for Squared Logit Error (Using 2n Rule for all 3 Estimators): n = 16

find an estimator "better" than the m.l. estimator has led to two estimators whose risk functions are approximately the same as that of the m.l. estimator.

An alternative method of defining estimates based on x = 0 and n for the SBP estimators is to use those estimates suggested by the corresponding values of a (the Bayes estimates for that prior given in (4.21) and (4.22)). These SBP estimators are compared with the m.l. estimator (using the "2n rule") in Figure 4.3. It is obvious from this graph that the optimum SBP estimators are better than the m.l. estimator only in some tiny region near $\theta = 0$ and $\theta = 1$. However, it should be noted that Figure 4.3 gives the risk functions evaluated from .01 to .99 in θ scale. Since the logit is the parameter being estimated, the risk functions are also presented in logit scale, in Figures 4.4 and 4.5. Although a somewhat different picture is presented in these two figures, it is still clear that the SBP estimators are superior to the maximum likelihood estimator only when θ is extremely close to 0 or 1. Moreover, the range of θ over which the m.l. estimator is superior increases with n. In practice, it is likely that the sample size will be sufficient to ensure the superiority of the maximum likelihood estimator.

Much of the credit for the performance of the m.l. estimator should go to Berkson's "2n rule," as can be seen by comparing Figure 4.2, in which all estimators used the "2n rule," with Figure 4.3, in which only the m.l. estimator used the "2n rule."

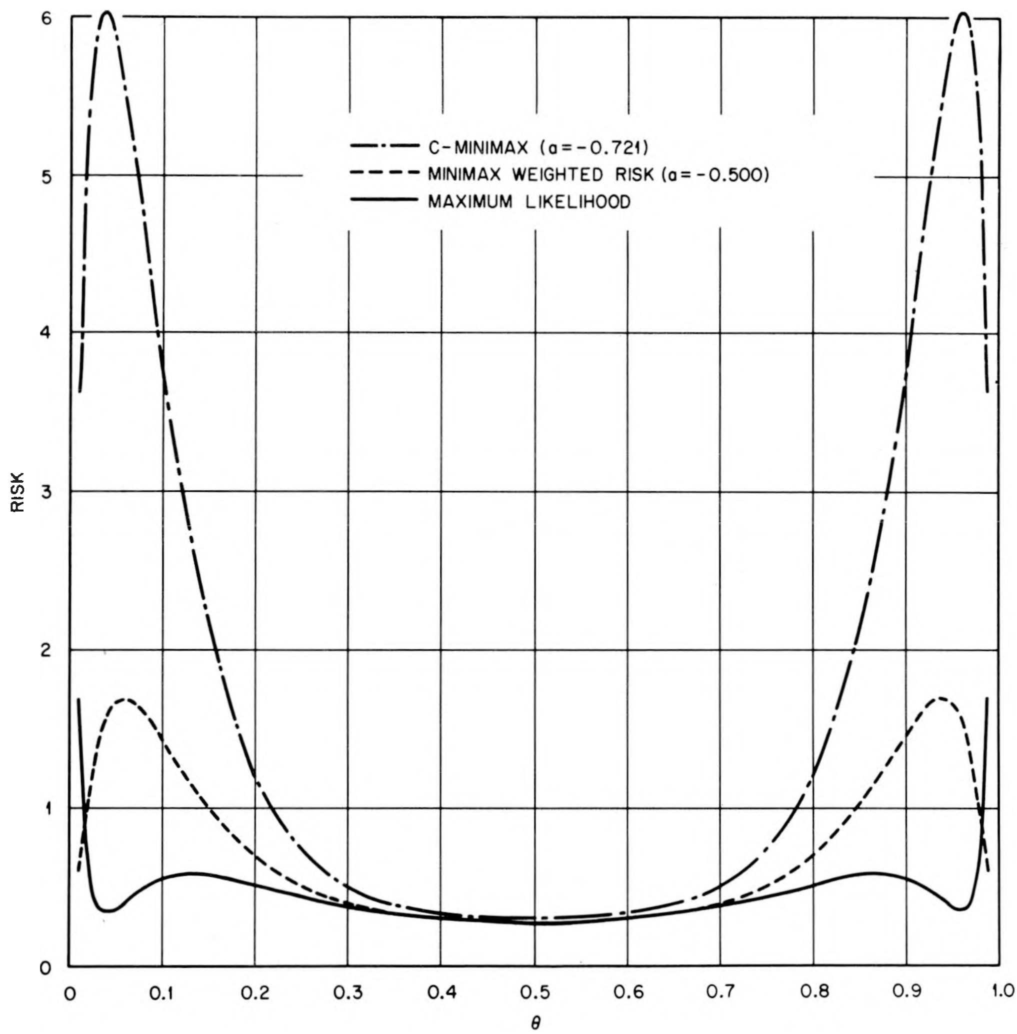

Figure 4.3. Risk Functions for Squared Logit Error: n = 16

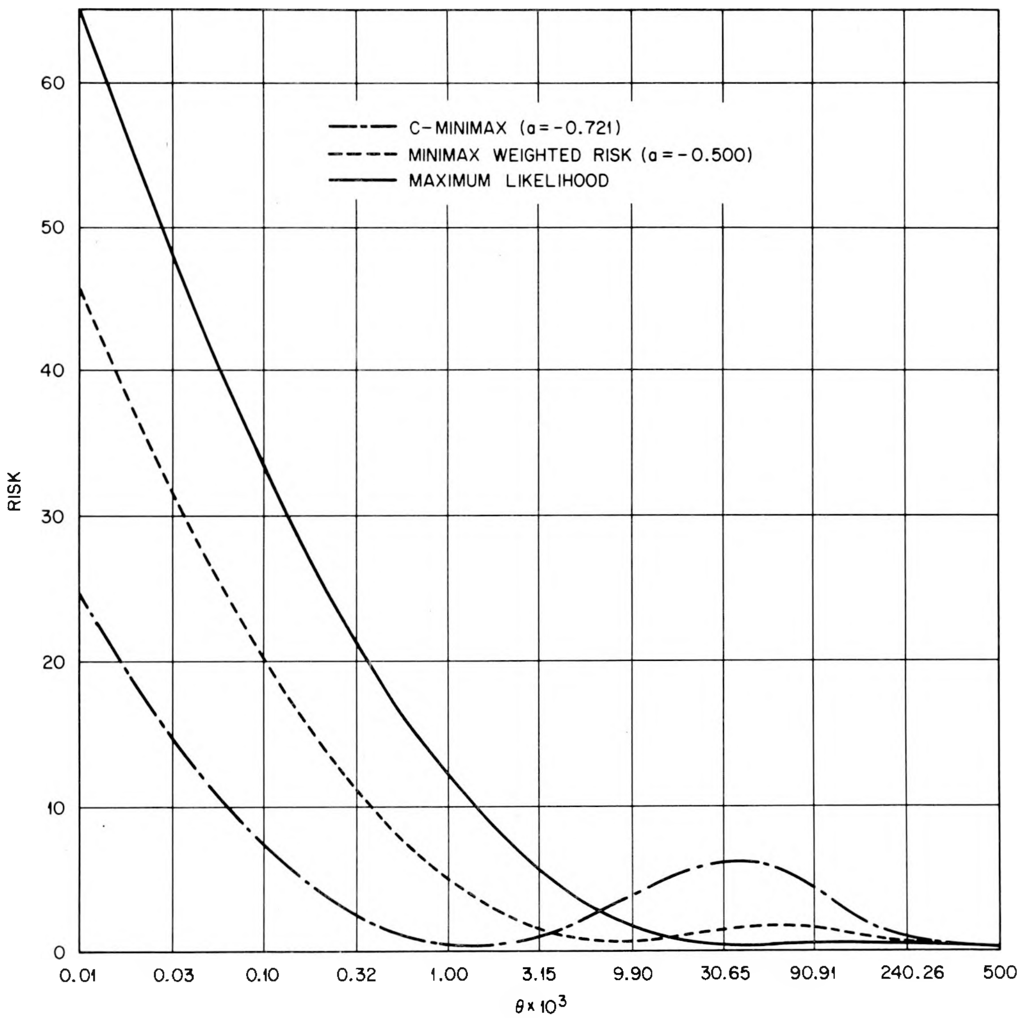

Figure 4.4. Risk Functions for Squared Logit Error (Logit Scale): n = 16

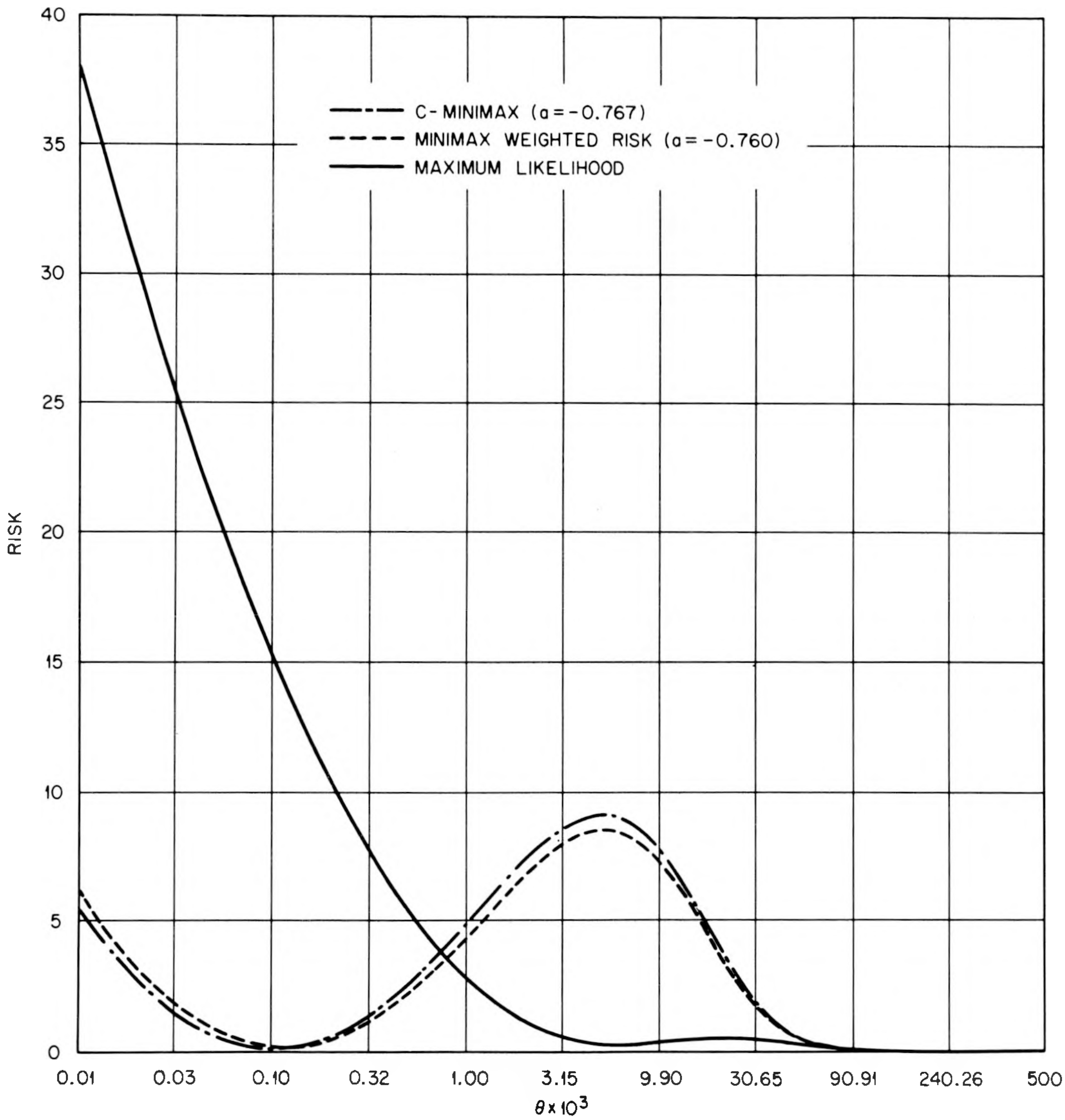

Figure 4.5. Risk Functions for Squared Logit Error (Logit Scale): n = 100

4.7 Summary and Conclusions

From the class C of SBP estimators of the form

$$\hat{\lambda} = \{\psi(\alpha+x+1) - \psi(\alpha+n-x+1),\ x = 0,1,\ldots,n\}$$

those estimators were found that minimized the maximum risk and weighted risk, for a modified form of the squared logit error loss function. These optimum SBP estimators were then compared with the m.l. estimator using the conventional squared logit error loss function.

Generally the maximum likelihood estimator appears to be the best of the three, although there is very little difference between them when the "2n rule" is used in all cases. When the Bayes-suggested alternative for the "2n rule" is used for the optimum SBP estimators, the maximum likelihood estimator is clearly superior (as is illustrated in the comparison of Figures 4.2 and 4.3), implying that the "2n rule" contributes a great deal to the good performance of the maximum likelihood estimator.

BIBLIOGRAPHY

1. Amos, D. E., "Additional Percentage Points for the Incomplete Beta Distribution." Biometrika, Vol. 50, 1963, pp. 449-451.

2. Anscombe, F. J., "On Estimating Binomial Response Relations." Biometrika, Vol. 43, 1956, pp. 461-464.

3. Berkson, Joseph, "Application of the Logistic Function to Bioassay." Journal of the American Statistical Association, Vol. 39, 1944, pp. 357-365.

4. Berkson, Joseph, "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function." Journal of the American Statistical Association, Vol. 50, 1955, pp. 130-162.

5. DeRouen, T. A., "The Logistic Model in the Analysis of Binomial Response Data." M.S. Thesis, Virginia Polytechnic Institute, 1969.

6. Dyke, G. V., and H. D. Patterson, "Analysis of Factorial Arrangements When the Data Are Proportions." Biometrics, Vol. 8, 1952, pp. 1-12.

7. Gart, John J., and James R. Zweifel, "On the Bias of Various Estimators of the Logit and Its Variance with Application to Quantal Bioassay." Biometrika, Vol. 54, 1967, pp. 181-187.

8. Good, I. J., The Estimation of Probabilities. Research Monograph No. 30, The M.I.T. Press, Cambridge, Massachusetts, 1965.

9. Gradshteyn, I. S., and I. M. Ryzhik, Table of Integrals, Series, and Products. Academic Press, Inc., New York, 1966.

10. Graybill, F. A., Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Co., Inc., Belmont, California, 1969.

11. Greenwood, J. A., and H. O. Hartley, Guide to Tables in Mathematical Statistics. Princeton University Press, Princeton, New Jersey, 1962.

12. Haldane, J. B. S., "The Estimation and Significance of the Logarithm of a Ratio of Frequencies." Annals of Human Genetics, Vol. 20, 1955, pp. 309-311.

13. Hitchcock, S. E., "A Note on the Estimation of the Parameters of the Logistic Function, Using the Minimum Logit χ^2 Method." Biometrika, Vol. 49, 1962, pp. 250-252.

14. Hodges, J. L., Jr., and E. L. Lehmann, "Some Problems in Minimax Point Estimation." Annals of Mathematical Statistics, Vol. 21, 1950, pp. 182-197.

15. Krutchkoff, R. G., Probability and Statistical Inference. Gordon and Breach Science Publishers, New York, 1970.

16. Lehmann, E. L., Notes on the Theory of Estimation. Associated Students Store, University of California, Berkeley, 1950.

17. Lindley, D. V., Introduction to Probability and Statistics From a Bayesian Viewpoint, Part 2, Inference. Cambridge University Press, London, 1965.

18. Naddeo, A., "Confidence Intervals for the Frequency Function and the Cumulative Frequency Function of a Sample Drawn from a Discrete Random Variable." Review of the International Statistical Institute, Vol. 36, No. 3, 1969, pp. 313-318.

19. Raiffa, H., and R. Schlaiffer, Applied Statistical Decision Theory. Graduate School of Business Administration, Harvard University, Boston, Massachusetts, 1961.

20. Steinhaus, H., "The Problem of Estimation." Annals of Mathematical Statistics, Vol. 28, 1957, pp. 633-648.

21. Wald, A., "Statistical Decision Functions Which Minimize the Maximum Risk." Annals of Mathematics, Vol. 46, No. 2, 1945, pp. 265-280.

22. Wald, A., "Statistical Decision Functions." Annals of Mathematical Statistics, Vol. 20, 1949, pp. 165-205.

23. Woolf, B., "On Estimating the Relationship Between Blood Group and Disease." Annals of Human Genetics, Vol. 19, 1954, pp. 251-253.

ORNL-4700
UC-32 — Mathematics and Computers

INTERNAL DISTRIBUTION

1-3. Central Research Library
4. ORNL — Y-12 Technical Library
Document Reference Section
5-24. Laboratory Records Department
25. Laboratory Records, ORNL R.C.
26. J. J. Beauchamp
27. K. O. Bowman
28. H. P. Carter
29. F. L. Culler
30-54. T. A. DeRouen
55. D. A. Gardiner
56. D. G. Gosslee
57. T. L. Hebble
58. R. F. Hibbs
59. W. H. Jordan
60. W. E. Lever
61-85. T. J. Mitchell
86. M. J. Skinner
87. A. H. Snell
88. D. A. Sundberg
89. V. R. R. Uppuluri
90. A. M. Weinberg

EXTERNAL DISTRIBUTION

91. Jesse Arnold, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060
92. P. L. Ford, Department of Mathematical Sciences, McNeese State University, Lake Charles, Louisiana 70601
93. Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060
94. Richard Krutchkoff, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060
95. Raymond Myers, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060
96. C. V. L. Smith, U.S. Atomic Energy Commission, Washington, D.C.
97. J. A. Swartout, Union Carbide Corporation, New York, N.Y. 10017
98. Charles Thigpen, Department of Statistics, University of Tennessee, Knoxville, Tennessee
99. Document Room, Computing Center, University of Notre Dame, Notre Dame, Indiana 46556
100. Laboratory and University Division, AEC, ORO
101. Patent Office, AEC, ORO
102-315. Given distribution as shown in TID-4500 under Mathematics and Computers category (25 copies — NTIS)