TECHNIQUES FOR IMPROVING ESTIMATORS WITH APPLICATIONS TO PROBLEMS WITH RESTRICTED PARAMETER SPACES

A Thesis

Submitted to the Faculty of Graduate Studies at

THE UNIVERSITY OF MANITOBA

in Partial Fulfilment of the Requirements

for the Degree of

MASTER OF SCIENCE

in the

DEPARTMENT OF STATISTICS

By

Gary Dukes, B.Sc.

Winnipeg, Manitoba

October 1975

THE UNIVERSITY OF MANITOBA LIBRARIES

"TECHNIQUES FOR IMPROVING ESTIMATORS
WITH APPLICATIONS TO
PROBLEMS WITH RESTRICTED PARAMETER SPACES"

by

GARY DUKES

A dissertation submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

© 1976

Permission has been granted to the LIBRARY OF THE UNIVERSITY OF MANITOBA to lend or sell copies of this dissertation, to the NATIONAL LIBRARY OF CANADA to microfilm this dissertation and to lend or sell copies of the film, and UNIVERSITY MICROFILMS to publish an abstract of this dissertation.

The author reserves other publication rights, and neither the dissertation nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

ABSTRACT

In many statistical estimation problems, an estimator which has desirable properties if the parameter it estimates is unrestricted, loses these properties when restrictions are imposed on the parameter space. These restrictions often occur naturally as orderings or physical constraints on parameters, for example in some estimation problems involved in process improvement, biological assay, and variance components.

This thesis is an attempt to present and unify the techniques which are useful in obtaining "better" estimators in problems with restricted parameter spaces, using the framework of statistical decision theory. This attempt has been motivated to some extent by the failure of many classical methods of estimation to provide reasonable estimators in such problems. The method of maximum likelihood, usually the most reliable of the classical methods, provides an estimator for ordered binomial parameters which Sackrowitz and Strawderman (1974) have shown to be inadmissible in most cases. Other methods provide estimators which take on values outside the parameter space.

The decision theoretic approach suggests a criterion for comparing estimators which makes use of the partial ordering of the risk functions. An estimator in a problem with a restricted parameter space will be considered reasonable if it is admissible and has uniformly smaller risk than the estimator which is considered optimal in the problem with the parameter space unrestricted.

In Chapter I, the basic concepts of statistical decision theory are introduced, and useful results concerning admissibility and minimaxity are established. General techniques for improving estimators

are presented in Chapter II, with examples of their use. The application of these techniques to a process improvement problem, that of estimating a Poisson parameter which is bounded either above or below, is discussed in detail in Chapter III. Some of these techniques are also applied to a quantal response problem, the estimation of ordered binomial parameters, in Chapter IV.

TABLE OF CONTENTS

ABSTRACT		1
CHAPTER I	SOME DECISION THEORETIC CONCEPTS	5
1.1	Parametric Statistical Decision Problem	5
1.2	Estimation of a Poisson Parameter	22
1.3	Remarks on Improving Estimators and Admissibility in Problems with Restricted Parameter Spaces	29
CHAPTER II	IMPROVING ESTIMATORS	34
2.0	Introduction and Summary	34
2.1	Projection onto the Closed Convex Hull of the Parameter Space	37
2.2	Optimizing within a Subclass of $\mathcal{D}$	41
2.3	Searching among a Complete Class	46
2.4	Constructing Reasonable Estimators — A Step-function Approach	49
2.5	Combining Estimators	52
CHAPTER III	APPLICATION OF TECHNIQUES FOR IMPROVING ESTIMATORS TO THE POISSON ESTIMATION PROBLEM WHEN THE PARAMETER SPACE IS TRUNCATED	55
3.1	Estimation of a Poisson Parameter Bounded Above	55
3.2	Estimation of a Poisson Parameter Bounded Below	69
CHAPTER IV	REMARKS ON THE PROBLEM OF ESTIMATING ORDERED BINOMIAL PROPENSITIES	79
APPENDIX	SOME USEFUL RESULTS	86
BIBLIOGRAPHY		87

ACKNOWLEDGEMENTS

From the author's point of view, the purpose of this research has been to exercise and refine his talents in a creative and interesting manner. In this connection the efforts of John Brewster were quite invaluable. His clarity and enthusiasm as a teacher inspired the author's interest in the subject, and this research would not have been possible without his insights, direction, and patient assistance.

The author also wishes to thank Helen Rhoda for her excellent typing, the Department of Statistics for financial support and the opportunity to teach, his fellow graduate students for the co-operative effort in learning Statistics, and his friends and family for tolerance and moral support.

CHAPTER I

Some Decision Theoretic Concepts

§1.1. The Parametric Statistical Decision Problem

A random variable (or vector), X, takes on values in a sample space, $\mathcal{X}$, and is distributed by a unique but unknown member of a class of possible underlying distributions, $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$, which is indexed by a parameter, θ, belonging to the parameter space, Θ. Based on the observed value of X, an action, a, belonging to the space of all possible actions, $\mathcal{A}$, is taken with the help of a decision rule, d, chosen from the class of all (nonrandomized) decision rules, $\mathcal{D} = \{d | d:\mathcal{X} \to \mathcal{A}\}$. If we were being more formal, we should require decision rules to be measurable. In this thesis, however, we shall ignore measurability considerations. The problem is to find a decision rule which, "on the whole, has the least disastrous consequences". To measure how disastrous the consequences are of choosing an action a when in fact the true parameter is θ, we use a nonnegative real-valued loss function, $L:\mathcal{A} \times \Theta \to [0, \infty)$. The average loss incurred by using the decision rule d, when in fact the true parameter is θ, is given by the risk function $r_d(\theta) = E_\theta[L(d(X), \theta)]$. The problem can now be stated as minimizing $r_d(\theta)$ in some manner over all θ, for variations in d. Except in pathological problems, there exists no decision rule which uniformly minimizes $r_d(\theta)$. However, the risk functions do provide a partial ordering of $\mathcal{D}$. A rule d is <u>as good as</u> a rule d' if

(1.1.1) $$r_d(\theta) \leq r_{d'}(\theta), \text{ for all } \theta \in \Theta.$$

A rule d is *better than* d' (or d *improves* d') if (1.1.1) holds and the inequality is strict for at least one θ in $\textcircled{H}$; d is *uniformly better than* d' if (1.1.1) holds with strict inequality at each θ in $\textcircled{H}$; and d is *equivalent to* d' if (1.1.1) is an equality for each θ in $\textcircled{H}$. If there exists a rule d' which is better than d, then d is *inadmissible*; otherwise d is *admissible*. Since admissibility is a property that any good decision rule must have, the question arises: Which decision rules are admissible? It turns out that the answer to this question is closely related to the Bayes or "almost Bayes" nature of the decision rule under consideration. For this reason we now turn our attention to the Bayes principle and its extensions.

Consider a *prior (probability) distribution function* Π on $\textcircled{H}$. The *Bayes risk* of a decision rule d with respect to Π is

$$R_\Pi(d) = E[r_d(\Theta)] = \int_{\textcircled{H}} \int_{\mathcal{X}} L[d(x),\theta]\, dF_\theta(x) d\Pi(\theta)$$

$$= \int_{\textcircled{H}} \int_{\mathcal{X}} L[d(x),\theta]\, f_\theta(x)\, d\mu(x)\, \pi(\theta)\, d\nu(\theta)\ ,$$

where $f_\theta(x)$ and $\pi(\theta)$ represent the density functions (if they exist) of X and Θ with respect to some σ-finite measures μ and ν on $\mathcal{X}$ and $\textcircled{H}$, respectively. From now on, we will not refer to μ and ν, and act as though $\mathcal{X}$ and $\textcircled{H}$ were subsets of Euclidean spaces and μ and ν were Lebesgue measures. However, it is to be understood that the integrals could be replaced by summation signs in what follows, if, for example, the variables X and Θ were discrete and μ and ν were therefore counting measures. Here Θ will be used to denote the random variable whose distribution is given by Π — whether we wish to think of the parameter as random or not. Now, a decision rule d_Π is *Bayes with*

respect to Π if it minimizes and makes finite the Bayes risk, that is, if

$$R_\Pi(d_\Pi) = \inf_{d \in \mathcal{D}} R_\Pi(d) < \infty .$$

But
$$R_\Pi(d) = \int_{\Theta} \int_{X} L[d(x),\theta]\, f_\theta(x)\, dx\, \pi(\theta)\, d\theta$$
$$= \int_{\Theta} \int_{X} L[d(x),\theta]\, h(x,\theta)\, dx\, d\theta$$
$$= \int_{\Theta} \int_{X} L[d(x),\theta]\, g(\theta|x)\, q(x)\, dx\, d\theta$$
$$= \int_{X} \int_{\Theta} L[d(x),\theta]\, g(\theta|x)\, d\theta\, q(x)\, dx ,$$

if the interchange of integrals is valid, and thus $R_\Pi(d)$ is minimized when the (conditional) posterior risk, $\int_{\Theta} L[d(x),\theta]\, g(\theta|x)\, d\theta$, is minimized for each given x, where $h(x,\theta)$ is the joint probability density function of X and Θ, $q(x)$ is the marginal density of X, and $g(\theta|x)$ the conditional posterior density of θ, given $X = x$. Thus the Bayes rule with respect to Π, $d_\Pi(X)$, can be determined by minimizing the posterior risk at each x, providing that $R_\Pi(d_\Pi) < \infty$. (Nonnegativity of the loss function will validate the required interchange in order of integration). Now suppose that we are estimating a real-valued function $\tau(\theta)$, that is, $\tau: \Theta \to (-\infty, \infty)$, subject to a weighted squared error loss function $L(a,\theta) = \gamma(\theta)[a - \tau(\theta)]^2$, where the weighting function $\gamma(\theta)$ is everywhere positive and finite. Then, for each $x \in X$, the posterior risk,

$$E\{L[d(X), \theta] \mid X = x\} = \int_{\Theta} [d(x) - \tau(\theta)]^2\, \gamma(\theta)\, g(\theta|x)\, d\theta$$
$$= [d(x)]^2 \int_{\Theta} \gamma(\theta)\, g(\theta|x)\, d\theta - 2d(x) \int_{\Theta} \tau(\theta)\, \gamma(\theta)\, g(\theta|x)\, d\theta$$
$$+ \int_{\Theta} [\tau(\theta)]^2\, \gamma(\theta)\, g(\theta|x)\, d\theta,$$

is an upward-facing parabola in $d(x)$ and is minimized for

$$d_{\Pi}(x) = \int_{\textcircled{H}} \tau(\theta)\, \gamma(\theta)\, g(\theta|x)\, d\theta \;\Big/\; \int_{\textcircled{H}} \gamma(\theta)\, g(\theta|x)\, d\theta \quad ,$$

which defines the Bayes rule with respect to Π. But

$$h(x,\theta) = g(\theta|x)\, q(x) = f(x|\theta)\, \pi(\theta) \quad ,$$

so we have proved the following theorem.

Theorem 1.1.1. If the loss function is of the form $L(a,\theta) = \gamma(\theta)[a - \tau(\theta)]^2$, where $0 < \gamma(\theta) < \infty$, then the Bayes estimator of the real-valued function $\tau(\theta)$ with respect to the prior distribution Π is given by

$$d_{\Pi}(x) = \frac{\int_{\textcircled{H}} \tau(\theta)\, \gamma(\theta)\, g(\theta|x)\, d\theta}{\int_{\textcircled{H}} \gamma(\theta)\, g(\theta|x)\, d\theta} = \frac{\int_{\textcircled{H}} \tau(\theta)\, \gamma(\theta)\, \pi(\theta)\, f(x|\theta)\, d\theta}{\int_{\textcircled{H}} \tau(\theta)\, \pi(\theta)\, f(x|\theta)\, d\theta} \quad ,$$

providing that the numerator and denominator both exist, and $R_{\Pi}(d_{\Pi}) < \infty$.

Theorem 1.1.2. If the Bayes rule with respect to Π, d_{Π}, is unique up to equivalence, then it is admissible.

Proof: Suppose $d_{\Pi}(X)$ is inadmissible, that is, there exists a rule $\phi(X)$ such that $r_{\phi}(\theta) \le r_{d_{\Pi}}(\theta)$ for each $\theta \in \textcircled{H}$, with strict inequality for at least one θ. This implies

1) that d_{Π} and ϕ are not equivalent, and

2) that $R_{\Pi}(\phi) = \int_{\textcircled{H}} r_{\phi}(\theta)\, d\Pi(\theta) \le \int_{\textcircled{H}} r_{d_{\Pi}}(\theta)\, d\Pi(\theta)$

$$= R_{\Pi}(d_{\Pi}) = \inf_{\phi \in \mathcal{D}} R_{\Pi}(d) \quad ,$$

so ϕ is a Bayes rule with respect to Π, contradicting the assumption

that Bayes rules with respect to Π are unique up to equivalence, and therefore d_Π is admissible. ‖

Example 1.1.1. Consider the problem of estimating θ, the mean of a normal distribution when the variance is known to be one. Without loss of generality, suppose that the estimate is to be based on a single observation X from this distribution, and is subject to squared error loss. The estimator X is both the minimum variance unbiased estimator and the maximum likelihood estimator. Is it admissible? The use of Theorem 1.1.2 requires that X be Bayes with respect to some prior distribution Π, so that application of Theorem 1.1.1 must yield

$X = \int_{(H)} \theta\, g(\theta|X)\, d\theta = E(\theta|X)$. But X is unbiased,

so $E(\theta^2) = E\{\theta E(X|\theta)\} = E\{E(\theta X|\theta)\} = E(\theta X)$

$= E\{E(\theta X|X)\} = E\{XE(\theta|X)\} = E(X^2)$,

and $R_\Pi(X) = E[(X-\theta)^2] = E(X^2) - 2E(\theta X) + E(\theta^2) = 0$.

But this is impossible, for $R_\pi(X) = E\{E[(X-\theta)^2|\theta]\} = E\{Var_\theta(X)\} = 1$. Thus X is not a Bayes rule, and we cannot use Theorem 1.1.2 to determine its admissibility.

A similar situation occurs in many common estimation problems, so Theorem 1.1.2 is inapplicable in determining the admissibility of many estimators which, as we shall see, can be proved to be admissible using other methods. These other methods make use of the "almost Bayes" nature of an estimator. We will now consider some extensions to the Bayes principle.

Consider a sequence of prior distributions, $\{\Pi_n\}_{n=1}^{\infty}$. If the sequence of Bayes rules with respect to these priors,

$\{d_{\Pi_n}\}_{n=1}^{\infty}$, converges to the (nonrandomized) rule $d(X)$ almost surely in X (for each $\theta \in \Theta$), then $d(X)$ is the <u>limit of Bayes rules</u> $d_{\Pi_n}(X)$.

In Example 1.1, consider a normal prior distribution Π_σ on Θ with mean zero and variance σ^2 . The joint density of X and θ is

$$h(x,\theta) = f(x|\theta)\,\pi_\sigma(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\theta^2/\sigma^2}$$

$$= \frac{1}{2\pi\sigma} e^{-\frac{1}{2}(x-\theta)^2 - \frac{1}{2}\theta^2/\sigma^2} .$$

The marginal density of X is

$$q(x) = \frac{1}{2\pi\sigma} \int_{\Theta} \exp\{-\tfrac{1}{2}\sigma^{-2}[\theta^2(1+\sigma^2) + 2\sigma^2 x\theta - x^2\sigma^2]\}\, d\theta$$

$$= (2\pi)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}x^2/(1+\sigma^2)\} \int_{\Theta} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}[\theta - x\sigma^2/(1+\sigma^2)]^2(1+\sigma^2)/\sigma^2\}\, d\theta$$

$$= [2\pi(1+\sigma^2)]^{-1} \exp\{-\tfrac{1}{2}x^2/(1+\sigma^2)\} , \quad x \in X = (-\infty, \infty) .$$

Therefore the posterior density of θ, given $X = x$, is

$$g(\theta|x) = h(x,\theta)/q(x) = [2\pi\sigma^2/(1+\sigma^2)]^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(1+\sigma^2)/\sigma^2 \cdot [\theta - x\sigma^2/(1+\sigma^2)]^2\} ,$$

for $\theta \in \Theta = (-\infty, \infty)$, that is, the normal density with mean $x\sigma^2/(1+\sigma^2)$ and variance $\sigma^2/(1+\sigma^2)$.

By Theorem 1.1.1, the Bayes rule with respect to Π_σ is given by the mean of the posterior distribution of θ, given $X = x$, that is, $d_{\Pi_\sigma}(X) = X\sigma^2/(1+\sigma^2)$. ($R_{\Pi_\sigma}(d_{\Pi_\sigma})$ is shown below to be finite). As $\sigma \to \infty$, $d_{\Pi_\sigma}(X)$ converges pointwise (and thus almost surely for each $\theta \in \Theta$) to X. Thus X is a limit of Bayes rules.

Another extension to the Bayes principle is given by extended

Bayes rules: Suppose $\epsilon > 0$. A decision rule ϕ is <u>ϵ-Bayes with respect to the prior distribution Π</u> if

$$R_\Pi(\phi) \leq \inf_{d\in\mathcal{D}} R_\Pi(d) + \epsilon \ .$$

A rule ϕ is <u>extended Bayes</u> if, for every $\epsilon > 0$, there exists a prior distribution with respect to which ϕ is ϵ-Bayes.

In our example, the Bayes risk with respect to Π_σ is minimized by the corresponding Bayes rule, $d_{\Pi_\sigma}(X) = X\sigma^2/(1+\sigma^2)$, and has the minimum value $R_{\Pi_\sigma}(d_{\Pi_\sigma}) = E[(\theta - d_{\Pi_\sigma}(X))^2]$

$= E[E\{(\theta - d_{\Pi_\sigma}(X))^2 | X\}] = E[\mathrm{Var}(\theta|X)] = \sigma^2/(1+\sigma^2)$.

But $R_{\Pi_\sigma}(X) = 1$, so $R_{\Pi_\sigma}(X) \leq \inf_{d\in\mathcal{D}} R_\Pi(d) + 1/(1+\sigma^2)$.

Thus, for each $0 < \epsilon < 1$, there exists a prior distribution, namely the normal distribution with mean zero and variance $\sigma^2 = (1-\epsilon)/\epsilon$, with respect to which X is ϵ-Bayes. (For $\epsilon \geq 1$, X is trivially ϵ-Bayes with respect to any prior distribution). Thus X is an extended Bayes rule.

Our final extension to the Bayes principle will be the most useful for our purposes. If there exists a measure Π on $\textcircled{H}$ (Π may be improper, that is, $\Pi(\textcircled{H})$ may be infinite) such that, for each x, $\int_{\textcircled{H}} L(d(x),\theta)\, f(x|\theta)\, d\Pi(\theta)$ takes on a finite minimum value when $d(x) = \phi_\Pi(x)$, then $\phi_\Pi(X)$ is a <u>generalized Bayes rule with respect to Π</u>. Here we are proceeding as we did for Bayes rules, by minimizing the "inside" integral. The difference is that if Π is improper, then usually the double integral (which would have been the Bayes risk), will be infinite.

In fact, the distinction between Bayes rules and generalized Bayes rules should not be made on the basis of whether Π is proper or improper. Instead, although both types of rules are chosen to minimize and make finite the posterior risk, the rule d_Π will be called Bayes if and only if the double integral $R_\Pi(d_\Pi)$ is finite. It should be noted that the proof of Theorem 1.1.2 is still valid for Bayes rules according to this definition, and that a rule may be Bayes with respect to an improper prior. Alternatively, the prior may be proper, and yet the rule d_Π, which minimizes and makes finite the posterior risk, may yield $R_\Pi(d_\Pi) = \infty$, and d_Π would then be called generalized Bayes. This distinction is useful when the loss function is of the form $\gamma(\theta)\, W(a,\theta)$, for then $\gamma(\theta)$ can formally be incorporated into the prior Π (see, for example, Section 1.2).

The prior measure with respect to which a particular rule ϕ is generalized Bayes may sometimes be suggested by the sequence of priors for which ϕ is a limit of Bayes rule. For example, in estimating the mean θ of a normal distribution with variance one on the basis of a single observation, X, subject to squared error loss, the estimator X is a limit of Bayes rules with respect to normal priors Π_σ with mean zero and variance σ^2, as $\sigma^2 \to \infty$. As σ^2 approaches infinity, Π_σ "approaches" Lebesgue measure, $d\Pi(\theta) = d\theta$, and the Bayes estimators with respect to Π_σ, $X\sigma^2/(1+\sigma^2)$, approach X. Proceeding as though Lebesque measure Π were a valid prior distribution, the generalized Bayes rule is obtained using Theorem 1.1.1 as

$$d_\Pi(x) = \Big[\int_{(H)} \theta\, f(x|\theta)\, d\theta\Big]\Big/\Big[\int_{(H)} f(x|\theta)\, d\theta\Big] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \theta\, e^{-\frac{1}{2}(x-\theta)^2} d\theta = x\,.$$

(Here, $R_\Pi(d_\Pi) = \int_{-\infty}^{\infty} 1\, d\theta = \infty$.) Thus X is also a generalized Bayes

rule with respect to the Lebesgue prior measure.

The three extensions presented above are often useful in characterizing admissibility. Their "almost Bayes" nature is not as powerful as the actual Bayes principle itself. In most problems, Bayes estimators are admissible (Theorem 1.1.2), but estimators which are "almost Bayes" need not be admissible. However, there are admissible estimators which are not Bayes. For example, X in the above situation is not Bayes, yet it is admissible (see application following Theorem 1.1.6 or Ferguson (1967) p.141 for proof), and in many problems all admissible estimators must be "almost Bayes" in some sense. Sacks (1963) shows that in the one-parameter exponential family, if an estimator is admissible, then it must be equivalent to a generalized Bayes estimator (if the loss function satisfies certain assumptions). Such characterizations of admissible rules are known as complete class theorems.

A class C of decision rules, $C \subset \mathcal{D}$, is defined to be <u>complete</u> if, given any rule $d \in \mathcal{D} \sim C$, there exists a rule $d' \in C$ which is better than d. C is <u>essentially complete</u> if, given any rule $d \in \mathcal{D} \sim C$, there exists a rule $d' \in C$ which is as good as d.

<u>Theorem 1.1.3</u>. If C is a complete class, and A denotes the class of admissible rules, then $A \subset C$.

<u>Proof</u>: Let $d \in A$ and suppose $d \notin C$. Then there exists a rule $d' \in C$ which is better than d. Therefore d is inadmissible, that is, $d \notin A$. This contradiction proves the theorem. ‖

<u>Theorem 1.1.4</u>. If C is essentially complete, and if there exists an admissible rule $d \notin C$, then there exists a rule $d' \in C$ which is

equivalent to d.

Proof: There exists a rule $d' \in C$ which is as good as d. Since d is admissible, d' cannot be better than d. Therefore d' is equivalent to d. ||

Since the parametric decision problem consists basically of choosing an optimal decision rule, a possible line of solution involves discarding as many inadmissible decision rules as possible, say by reducing $\mathcal{D}$ to some smaller complete (or essentially complete) class. One such reduction is implicit in our statement of the decision problem — we have considered only nonrandomized decision rules. It turns out that the class of nonrandomized decision rules is essentially complete in the problem which allows the decision space to contain randomized rules, if the action space is a convex subset of k-dimensional Euclidean space, E_k, and if $L(a,\theta)$ is convex in a for each fixed θ, with $\lim_{|a|\to\infty} L(a,\theta) = \infty$. (Consideration should be made to the measurability of the nonrandomized decision rule generated pointwise from a given randomized rule — for discussion see Ferguson (1967) pp. 76-79). The most generally applicable reduction of $\mathcal{D}$ uses the essential completeness of the class of rules based on a sufficient statistic, which always holds. (See Ferguson (1967) p.120 for proof).

Sacks (1963) shows that the class of generalized Bayes rules is essentially complete in the problem of estimating the natural parameter θ of a distribution with density $f_\theta(x) = \beta(\theta)e^{x\theta}$, with respect to a measure μ on X, belonging to the exponential family, if the loss function is nonnegative, finite, continuous, increasing to ∞ as $|a|\to\infty$, bowl shaped in a for fixed θ, and exponentially dominated in θ.

In problems where the conditions required for the essential completeness of the class of generalized Bayes rules are satisfied and the loss function and action space are convex, the class of nonrandomized generalized Bayes rules (which by the Factorization Theorem must be based on a sufficient statistic) is essentially complete. Although it is sometimes difficult to tell whether or not a given rule is generalized Bayes, it is trivial to determine whether or not the rule is nonrandomized, or based on a sufficient statistic, and so we shall henceforth consider only those rules which are nonrandomized and based on the minimal sufficient statistic. This class of rules is also essentially complete whenever the action space is a convex subset of E_k and $L(a,\theta)$ is convex in a for each fixed θ (Ferguson (1967) p.121).

Instead of characterizing the class of admissible decision rules using a complete class theorem, it is often desirable to determine whether a given rule is admissible. The following result is used in Section 1.2 to show that a Poisson random variable is an admissible estimator for the parameter of its distribution.

Let X be a random variable with probability density $f_\theta(x)$, $\theta \in \Theta \subset (-\infty, \infty)$, and let $\phi(X)$ be any estimator of θ, with bias function $b_\phi(\theta) = E_\theta[\phi(X)] - \theta$. Then, under certain regularity conditions, the Cramér-Rao inequality states that the variance of $\phi(X)$ satisfies

$$(1.1.2) \qquad \sigma^2_\phi(\theta) \geq \frac{[1+b'_\phi(\theta)]^2}{E_\theta[(\frac{d}{d\theta} \log f_\theta(x)^2]} \quad , \quad \text{for each } \theta \in \Theta .$$

It is often desired to estimate θ subject to a loss function of the form $L(a,\theta) = \gamma(\theta)\,(a-\theta)^2$, where $0 < \gamma(\theta) < \infty$, for all $\theta \in \Theta$.

Let us first consider the squared error loss function, $L(a,\theta) = (a-\theta)^2$. Then

$$r_\phi(\theta) = E_\theta\,[(\phi(X) - \theta)^2 = b_\phi^2(\theta) + \sigma_\phi^2(\theta)$$

$$\geq b_\phi^2(\theta) + \frac{[1+b_\phi'(\theta)]^2}{E_\theta[(\frac{d}{d\theta}\log f_\theta(x))^2]} =: C_{b_\phi}(\theta) ,$$

where $C_b(\theta)$ is the lower bound on the risk function of any estimator with bias function $b(\theta)$. The following theorem, which is based on the idea of bounding risk functions, is due to Hodges and Lehmann (1950).

Theorem 1.1.5. If (1.1.2) is satisfied for all estimators $\phi(X)$ of θ, and $\phi^*(X)$ is an estimator which achieves the Cramér-Rao lower bound (that is, for $\phi^*(X)$, (1.1.2) becomes an equality), and if, for every bias function $b(\theta)$,

(1.1.3) $\quad C_b(\theta) \leq C_{b_{\phi^*}}(\theta)$ for all $\theta \in \text{Ⓗ}$ implies

$\quad b(\theta) = b_{\phi^*}(\theta)$ for all $\theta \in \text{Ⓗ}$,

then $\phi^*(X)$ is admissible subject to loss functions of the form $\gamma(\theta)(a-\theta)^2$, where $0 < \gamma(\theta) < \infty$ for all $\theta \in \text{Ⓗ}$.

Proof: Suppose that ϕ^* is inadmissible subject to squared error loss. Then there exists an estimator $\phi(X)$ with $r_\phi(\theta) \leq r_{\phi^*}(\theta)$ for all $\theta \in \text{Ⓗ}$, with strict inequality for some θ. Since $\phi^*(X)$ achieves the Cramér-Rao lower bound, $r_{\phi^*}(\theta) = C_{b_{\phi^*}}(\theta)$, for all $\theta \in \text{Ⓗ}$. Since the Cramér-Rao inequality is satisfied for all estimators, the mean squared error of $\phi(X)$ is bounded below as

$$r_\phi(\theta) \geq C_{b_\phi}(\theta) \quad \text{for all} \quad \theta \in \text{Ⓗ} .$$

Thus $C_{b_\phi}(\theta) \le C_{b_{\phi^*}}(\theta)$ for all $\theta \in \textcircled{H}$. By the final condition (1.1.3) of the theorem, $b_\phi(\theta) = b_{\phi^*}(\theta)$ for all $\theta \in \textcircled{H}$. But $C_b(\theta)$ depends on the estimator only through its bias function $b(\theta)$. Therefore $C_{b_\phi}(\theta) = C_{b_{\phi^*}}(\theta)$, for all $\theta \in \textcircled{H}$, so $r_\phi(\theta) \ge r_{\phi^*}(\theta)$, for all $\theta \in \textcircled{H}$. This contradicts the statement that there exists an estimator $\phi(X)$ with $r_\phi(\theta) \le r_{\phi^*}(\theta)$, for all $\theta \in \textcircled{H}$, and with strict inequality for some $\theta \in \textcircled{H}$. Therefore $\phi^*(X)$ is admissible subject to squared error loss. Since for each fixed θ, $\gamma(\theta)$ can be factored out of the expected loss, using the loss function $\gamma(\theta)(a-\theta)^2$ is equivalent to multiplying each of the risk functions by the same positive, finite constant. This does not affect the order relationship between the risk functions, so $\phi^*(X)$ is admissible under all loss functions of the form $L(a,\theta) = \gamma(\theta)(a-\theta)^2$, where $0 < \gamma(\theta) < \infty$ for all $\theta \in \textcircled{H}$. ‖

Unfortunately, Theorem 1.1.5 is not of wide use. Its application seems to be limited to exponential family problems where the parameter space is unrestricted, and even in such problems it will indicate the admissibility of only one estimator (and any equivalent to it), although there are usually an uncountable number of admissible estimators. Furthermore, verification of the conditions of the theorem may be somewhat difficult and time-consuming. This is illustrated in the next section, where Theorem 1.1.5 is used to prove that a Poisson random variable is an admissible estimator of the parameter of its distribution.

For exponential family problems, another method of proving an estimator admissible is given by the following theorem due to Karlin

(1958).

Theorem 1.1.6. (Karlin's Theorem) Let $f_\omega(x) = \beta(\omega)e^{x\omega}$ describe the density of a member of the exponential family with respect to a measure μ. If

(1.1.4) $\int_c^b \beta^{-2}(\omega)\, d\omega \to \infty$ as $b \to \bar{\omega}$, and

(1.1.5) $\int_a^c \beta^{-2}(\omega)\, d\omega \to \infty$ as $a \to \underline{\omega}$ where c is an interior point of the natural parameter space $\Omega = (\underline{\omega}, \bar{\omega})$, then $X/(\lambda+1)$ is an admissible estimator of $\theta(\omega) = E_\omega(X)$, subject to loss functions of the form $L(a,\theta) = \gamma(\theta)(a-\theta)^2$, where $0 < \gamma(\theta) < \infty$.

To apply Theorem 1.1.6 to the problem of estimating the mean of a normal distribution with variance known to be one (Example 1.1.1), we write the density of the normal distribution (with respect to Lebesgue measure) as

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} = K(x)\, e^{-\theta^2/2}\, e^{x\theta}, \quad X = \textcircled{H} = (-\infty, \infty),$$

so that the mean of X is in fact the natural parameter of this exponential family, and $\beta(\theta) = e^{-\theta^2/2}$. Conditions (1.1.4) and (1.1.5) become

$$\int_c^b e^{\lambda\theta^2/2}\, d\theta \to \infty \quad \text{as} \quad b \to \infty \quad \text{and}$$

$$\int_a^c e^{\lambda\theta^2/2}\, d\theta \to \infty \quad \text{as} \quad a \to -\infty,$$

which are satisfied for any $c \in (-\infty, \infty)$ by all $\lambda \geq 0$. Therefore all estimators of the form αX, where $0 \leq \alpha \leq 1$, are admissible subject to $L(a,\theta) = \gamma(\theta)(a-\theta)^2$. For an alternative proof of the admissibility

of X see Ferguson (1967) p.141.

Unfortunately, Karlin's Theorem also lacks wide applicability as it can only be used to determine the admissibility of estimators of the form αX, and many admissible estimators are not of this form.

By Sacks' (1963) result, we do know that admissible estimators are generalized Bayes, and this suggests looking within the class of generalized Bayes estimators for an admissible one. But, it is sometimes difficult to show that a given estimator is generalized Bayes, since a great number of prior distributions must be canvassed. In addition, this still leaves the problem of determining the admissibility of a generalized Bayes estimator.

In circumventing these difficulties, use may be made of a sufficient condition on the prior distribution Π for the admissibility of the generalized Bayes estimator with respect to Π, given by Zidek (1970). Since the generalized Bayes estimator with respect to Π is easily determined from Π, a characterization of priors leading to admissible estimators is quite useful. It might now be possible to reverse our usual process and to look within a class of admissible estimators for one which is otherwise reasonable. A method of Blyth (1951) can also be used in a wide variety of problems to determine the admissibility of a limit of Bayes estimators.

As a criterion for choosing decision rules, admissibility is rather weak. Many decision rules which seem ridiculous turn out to be admissible. For example, in the estimation of the mean of a normal distribution, the estimator 5, which totally ignores the sample evidence, is admissible because it is the only estimator with zero risk when $\theta = 5$. However, subject to squared error loss, 5 does very

poorly when θ is far away from 5. It would be desirable for an estimator to treat each of the unknown states of nature in a somewhat more equal manner. One criterion which does this is that of minimaxity.

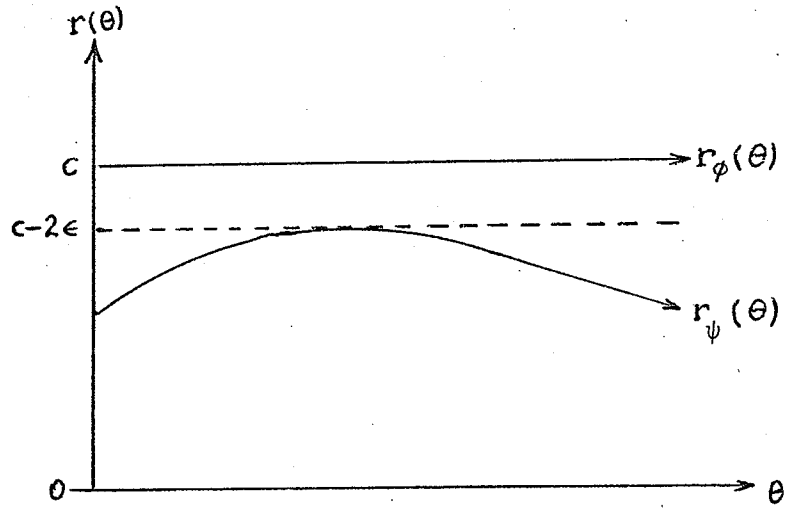
A decision rule ϕ is <u>minimax</u> if

$$\sup_{\theta \in \textcircled{H}} r_\phi(\theta) = \inf_{d \in \mathcal{D}} \sup_{\theta \in \textcircled{H}} r_d(\theta) .$$

Among admissible rules, a minimax rule tends to have the flattest risk function. A rule which has a constant risk function is termed an <u>equalizer</u> rule.

<u>Theorem 1.1.7.</u> If an equalizer rule, ϕ, is extended Bayes, then it is minimax.

<u>Proof:</u> Let $r_\phi(\theta) = c$, for all $\theta \in \textcircled{H}$. If ϕ is not minimax, then there exists a rule ψ whose maximum risk is strictly smaller than c,



say $\sup_{\theta \in \textcircled{H}} r_\psi(\theta) = c - 2\epsilon$, for some $\epsilon > 0$. Now, if ϕ is extended Bayes, then there exists a prior distribution Π_ϵ, with respect to which ϕ is ϵ-Bayes, that is,

$R_{\Pi_\epsilon}(\phi) \le R_{\Pi_\epsilon}(d) + \epsilon$, for all $d \in \mathcal{D}$. But

$$R_{\Pi_\epsilon}(\psi)+\epsilon = \int_{(H)} r_\psi(\theta)\, d\Pi_\epsilon(\theta)+\epsilon \le \int_{(H)} (c-2\epsilon)\, d\Pi_\epsilon(\theta)+\epsilon$$

$$= c-2\epsilon+\epsilon = c-\epsilon .$$

Thus $R_{\Pi_\epsilon}(\theta) = \int_{(H)} r_\phi(\theta)\, d\Pi_\epsilon(\theta) = c \le c - \epsilon$.

This is a contradiction, so ϕ must be minimax. ‖

In Example 1.1.1, $r_X(\theta) = \mathrm{Var}_\theta(X) = 1$, and X was shown to be extended Bayes, so by Theorem 1.1.7, X is minimax. The minimal sufficient statistic $\bar{X}$ for a random sample $X_1,\ldots,X_n$ from $N(\theta,1)$ has the normal distribution with mean θ and variance $1/n$, and so the above treatment of X applies without loss of generality to $\bar{X}$, as well.

As a criterion for choosing among decision rules, minimaxity is itself imcomplete, since one of two minimax estimators may be better than the other. Similarly, admissibility allows for ridiculous estimators. One reasonable method of choosing a decision rule might then be to combine the criteria of admissibility and minimaxity, that is, to search for an admissible minimax decision rule.

In the author's opinion, however, admissibility is a much more important property of a good decision rule than is minimaxity. Minimaxity seems too restrictive a criterion, especially considering its relative instability with respect to order-preserving fluctuations in the loss function. In the example treated above, X is admissible subject to all loss functions of the form $L(a,\theta) = \gamma(\theta)(a-\theta)^2$, where $0 < \gamma(\theta) < \infty$, while the minimaxity of X depends on $\gamma(\theta)$. Though minimaxity succeeds in eliminating many ridiculous admissible decision

rules, it also eliminates some which are quite reasonable, especially if the decision-maker's idea of the loss function is imprecise.

§1.2 Estimation of a Poisson Parameter

In this section, we consider some properties of a single Poisson random variable, X, as an estimator of the parameter θ of the Poisson distribution. The results presented here are not new (see, for example, Ferguson (1967) p.52, Hodges and Lehmann (1951), and Karlin (1958), but are included to provide illustration of the use of the decision theoretic concepts of the previous section, and as motivation for the results in Chapter III.

Since the sufficient statistic $T = \Sigma X_i$ based on a random sample from a Poisson distribution also possesses a Poisson distribution, we suppose, without loss of generality, that a single observation X will be taken from

$$f_\theta(x) = \theta^x e^{-\theta}/x! \ , \qquad x = 0,1,2,\ldots,$$

where $\theta \in \Theta = (0, \infty)$. An estimator of θ takes on a value in the action space $A = [0, \infty)$ and is subject to a penalty determined by the "normalized" squared error loss function, $L(a,\theta) = (a-\theta)^2/\theta$. As argued in the previous section, such a normalization does not affect the admissibility of estimators. However, when the loss is squared error, there is no minimax rule, since $r_\phi(\theta) \to \infty$ as $\theta \to \infty$ for all $\phi \in \mathcal{D}$. The normalization used here makes the risk function of X, the usual estimator (obtained by classical techniques such as method of moments, maximum likelihood, minimum variance unbiased estimation), constant. This implies that X is minimax for this problem. The use of $(a-\theta)^2/\theta$

as the loss function is perhaps also reasonable from the point of view that θ is the variance of X as well as the mean, so that an absolute error of one unit when θ is close to zero should be penalized more heavily than an error of the same magnitude when θ is large, say 100.

Theorem 1.2.1. In this problem, X is the limit of Bayes estimators with respect to $G[(n+1)/n, n]$ priors.

Proof: Consider prior Gamma densities,

$$\pi_{\alpha,\beta}(\theta) = \theta^{\alpha-1} e^{-\theta/\beta} \beta^{-\alpha} / \Gamma(\alpha) , \quad \theta > 0 .$$

The posterior density of Θ, given $X=x$, is

$$g(\theta|x) = \frac{f(x|\theta)\, \pi(\theta)}{\int_{(H)} f(x|\theta)\, d\Pi(\theta)} = \frac{\theta^{\alpha+x-1} e^{-\theta(\beta+1)/\beta}}{\int_0^\infty \theta^{\alpha+x-1} e^{-\theta(\beta+1)/\beta} d\theta}$$

$$= \theta^{\alpha+x-1} e^{-\theta(\beta+1)/\beta} [\beta/(\beta+1)]^{-(\alpha+x)} / \Gamma(\alpha+x) , \ 0 < \theta < \infty .$$

By Theorem 1.1.1, the Bayes estimator with respect to $\Pi = \Pi_{\alpha,\beta}$ is given by

$$\phi_\Pi(x) = \frac{\int_0^\infty g(\theta|x)\, d\theta}{\int_0^\infty \theta^{-1} g(\theta|x)\, d\theta} = \frac{\Gamma(\alpha+x)[\beta/(\beta+1)]^{\alpha+x}}{\Gamma(\alpha+x-1)[\beta/(\beta+1)]^{\alpha+x-1}} = \frac{\beta}{\beta+1} (x+\alpha-1) ,$$

providing $\alpha > 1$ and $R_\Pi(\phi_\Pi) < \infty$. And,

$$r_{\phi_\Pi}(\theta) = \theta^{-1}[\beta/(\beta+1)]^2 \operatorname{Var}_\theta(X+\alpha-1) + \theta^{-1}\{E_\theta[(X+\alpha-1)\beta/(\beta+1)]-\theta\}^2$$

$$= 1 + (\beta+1)^{-2} \{\theta - 2\alpha\beta - 1 + \beta^2(\alpha-1)^2/\theta\} .$$

But $\theta \sim G(\alpha,\beta)$, so $E[\theta] = \int_0^\infty \theta^\alpha e^{-\theta/\beta} \beta^{-\alpha}/\Gamma(\alpha)\, d\theta = \alpha\beta$,

$$E[\theta^{-1}] = \int_0^\infty \theta^{\alpha-2} e^{-\alpha/\beta} \beta^{-\alpha}/\Gamma(\alpha) = [(\alpha-1)\beta]^{-1} ,$$

and $R_\Pi(\phi_\Pi) = \beta/(\beta+1)$. Therefore, $(X+\alpha-1)\,\beta/(\beta+1)$ is Bayes with respect to $\Pi_{\alpha,\beta}$ for $\alpha > 1$, $\beta > 0$. Now define a sequence of priors $\{\Pi_n\}$ by $\Pi_n = \Pi_{\alpha,\beta}$, where $\alpha = (n+1)/n$, and $\beta=n$. Then the corresponding sequence of Bayes estimators is $\{(nX+1)/(n+1)\}$. Since

$$\lim_{n\to\infty} (nX+1)/(n+1) = X \ ,$$

X is the limit of Bayes estimators with respect to $G[(n+1)/n, n]$ priors. ‖

Theorem 1.2.2. In this problem, X is a generalized Bayes estimator with respect to Lebesgue prior measure, $d\Pi(\theta) = d\theta$.

Proof: The posterior density of Θ, given $X=x$, is

$$g(\theta|x) = \theta^x e^{-\theta}/x! \ , \qquad \theta > 0 \ .$$

The generalized Bayes estimator with respect to Π minimizes and makes finite

$$\int_0^\infty [\phi(x) - \theta]^2 \, \theta^{x-1} e^{-\theta} \, d\theta, \quad \text{for each} \quad x \in X = \{0,1,2,\ldots\} \ .$$

$\phi_\Pi(x) = [\int_0^\infty \theta^x e^{-\theta} \, d\theta]/[\int_0^\infty \theta^{x-1} e^{-\theta} \, d\theta]$ achieves this by Theorem 1.1.1, provided both numerator and denominator are finite, that is, for $x = 1,2,\ldots$. When $x=0$, $\int_0^\infty [\phi(0) - \theta]^2 \, \theta^{-1} e^{-\theta} d\theta$ is made finite only for $\phi(0) = 0$ (and therefore minimized). Thus the estimator

$$\phi_\Pi(X) = \begin{cases} \Gamma(X+1)/\,\Gamma(X) & X = 1,2,\ldots \\ 0 & X = 0 \end{cases} = X$$

is generalized Bayes with respect to Lebesgue measure. It is not Bayes because

$$R_\Pi(X) = \int_0^\infty r_X(\theta) \, d\theta = \int_0^\infty 1 \, d\theta = \infty \ .$$ ‖

Remark: Later we will be interested in estimators which are generalized Bayes with respect to $\Pi_{\alpha,\beta}$ where $-1 < \alpha < \infty$, and $0 < \beta \leq \infty$. For $\alpha+x > 1$, Theorem 1.1.1 shows that

$$\phi_\Pi(x) = (x+\alpha-1)\,\beta/(\beta+1) \quad \text{minimizes and makes finite}$$

$$\int_0^\infty (\theta(x) - \theta)^2\, \theta^{\alpha+x-2}\, e^{-\theta(\beta+1)/\beta}\, d\theta \ ,$$

and $\phi(x) = 0$ does so for $\alpha+x \leq 1$. Thus, for all $\beta > 0$, $-1 < \alpha \leq 1$,

$$\phi_\Pi(X) = \max\{0, (X+\alpha-1)\beta/(\beta+1)\}$$

is generalized Bayes with respect to $\Pi = \Pi_{\alpha,\beta}$. Carrying out the same argument for $\beta = \infty$, the generalized Bayes estimator with respect to $d\Pi(\theta) = \theta^{\alpha-1}\, d\theta$ is

$$\phi_\Pi(X) = \max\{0, X+\alpha-1\} \ , \quad \text{for} \quad \alpha > -1 \ .$$

Theorem 1.2.3. In this problem X is extended Bayes.

Proof: With respect to $\Pi_{\alpha,\beta}$, where $\alpha > 1$, $\beta > 0$, the Bayes risk of the Bayes estimator $\phi(X) = (X+\alpha-1)\,\beta/(\beta+1)$ is

$$R_\Pi(\phi_\Pi) = \beta/(\beta+1) \ ,$$

and the Bayes risk of X is

$$R_\Pi(X) = \int_0^\infty \theta^{-1}\, E_\theta[(X-\theta)^2]\, d\Pi_{\alpha,\beta}(\theta) = \int_0^\infty 1 \ d\Pi_{\alpha,\beta}(\theta) = 1 \ .$$

For $0 < \epsilon < 1$, we can choose $\beta > 0$ as $\beta = (1-\epsilon)/\epsilon$, and any $\alpha > 1$ so that

$$R_\Pi(X) = 1 = R_\Pi(\theta_\Pi) + \epsilon = [(1-\epsilon)/\epsilon]/[1/\epsilon] + \epsilon = 1 \ ,$$

that is, X is ϵ-Bayes with respect to $\Pi_{\alpha,\beta}$. For $\epsilon \geq 1$, X is trivially ϵ-Bayes with respect to any prior distribution.

Therefore, X is extended Bayes. ‖

Theorem 1.2.4. In this problem, X is minimax.

Proof: Since $r_X(\theta) = \theta^{-1}E_\theta[(X-\theta)^2] = 1$,

X is an equalizer rule. By Theorem 1.2.3., X is extended Bayes, so we can apply Theorem 1.1.7 to conclude that X is minimax. ‖

Theorem 1.2.5. In this problem, X is admissible.

Proof: The proof consists of verifying that the conditions of Theorem 1.1.5 are satisfied in this particular problem. It must first be shown that Equation (1.1.2.), the Cramér-Rao inequality, is satisfied for all estimators $\phi(X)$, that is, that

$$\mathrm{Var}_\theta[\phi(X)] \geq [1+b'_\phi(\theta)]^2 / E_\theta[(\frac{d}{d\theta} \log f_\theta(X))^2]$$

where $b_\phi(\theta) = E_\theta[\phi(X)] - \theta$ is the bias function of ϕ

and $$E_\theta[(\frac{d}{d\theta} \log f_\theta(X))^2] = E_\theta[(\frac{d}{d\theta} (X \log\theta - \theta))^2] = E_\theta[(X/\theta-1)^2] = 1/\theta.$$

The inequality holds trivially when $\mathrm{Var}_\theta[\phi(X)]$ is infinite.
When $\mathrm{Var}_\theta[\phi(X)]$ exists,

$$E_\theta[\phi(X)] = \sum_{x=0}^{\infty} |\phi(x)\ \theta^x\ e^{-\theta}/x!|$$

becomes an absolutely convergent power series, and may be differentiated term by term. Thus

$$\frac{d}{d\theta}\ E_\theta[\phi(X)] = \sum_{x=0}^{\infty} \frac{\phi(x)}{x!}\ [\frac{d}{d\theta}\ \theta^x e^{-\theta}] = \sum_{x=0}^{\infty} \phi(x)[\frac{x}{\theta} - 1]\ \frac{\theta^x e^{-\theta}}{x!}$$

$$= E_\theta[\phi(X)(\frac{X}{\theta} - 1)] .$$

By definition of the bias function $E_\theta[\phi(X)] = \theta + b_\phi(\theta)$, so $\frac{d}{d\theta} E_\theta[\phi(X)] = 1 + b'_\phi(\theta)$. Now, since $E_\theta[X/\theta - 1] = 0$, $Cov_\theta[\phi(X), X/\theta - 1] = E_\theta[\phi(X)(X/\theta - 1)] = 1 + b'_\phi(\theta)$.

But $Cov^2_\theta[\phi(X), X/\theta - 1] \le Var_\theta[\phi(X)]\ Var_\theta[X/\theta - 1]$,

so $Var_\theta[\phi(X)] \ge Cov^2_\theta[\phi(X), X/\theta - 1]/Var_\theta[X/\theta - 1]$

$$= \theta[1 + b'_\phi(\theta)]^2 ,$$

for all $\theta \in \textcircled{H}$, and each $\forall\ \phi \in D$

Secondly, does X achieve the lower bound of (1.1.2)? Since X is unbiased, $b_X(\theta) = 0$, for each $\theta \in \textcircled{H}$. $Var_\theta[X] = \theta$ since X has the Poisson distribution, and $\theta[1 + b'_X(\theta)]^2 = \theta$, so the two sides of the inequality are equal and X achieves the Cramér-Rao lower bound. Finally, we must verify condition (1.1.3), that for every bias function $b(\theta)$, $C_b(\theta) = C_{b_X}(\theta)$, for all $\theta \in \textcircled{H}$, implies $b(\theta) = b_X(\theta)$ for all $\theta \in \textcircled{H}$, where $C_b(\theta)$ is the lower bound on the mean squared error of an estimator with bias function $b(\theta)$, that is, $C_b(\theta) = b^2(\theta) + \theta[1+b'(\theta)]^2$. Here, $C_{b_X}(\theta) = 0 + \theta = \theta$, so we must show that

$b^2(\theta) + \theta[1 + b'(\theta)]^2 \le \theta$, for all $\theta \in \textcircled{H}$, implies $b(\theta) = 0$, for all $\theta \in \textcircled{H}$. Suppose $b^2(\theta) + [1 + b'(\theta)]^2 \le \theta$ for all $\theta \in \textcircled{H}$. Then

1) $b^2(\theta) \le \theta$, so $|b(\theta)| \le \sqrt{\theta}$, and

2) $[1 + b'(\theta)]^2 \le 1 - b^2(\theta)/\theta \le 1 - b^2(\theta)/\theta + [b^2(\theta)/(2\theta)]^2$, and

therefore, $[1 + b'(\theta)]^2 \le [1 - b^2(\theta)/(2\theta)]^2$, with strict inequality whenever $b(\theta) \neq 0$. But since $|b(\theta)| \le \sqrt{\theta}$, we have $b^2(\theta)/(2\theta) \le \frac{1}{2}$,

and therefore $1 - b^2(\theta)/(2\theta)$ is the positive square root of $[1 - b^2(\theta)/(2\theta)]^2$. Thus, with strict inequality whenever $b(\theta) \neq 0$, we have $b'(\theta) \leq -b^2(\theta)/(2\theta)$, and so $b(\theta)$ is nonincreasing. Also, $\lim_{\theta \to 0^+} |b(\theta)| \leq \lim_{\theta \to 0^+} \sqrt{\theta} = 0$, so $b(\theta) \leq 0$, for all $\theta \in (0, \infty)$.

Now, suppose that there exists a θ_0 for which $b(\theta_0) < 0$. Then $b'(\theta_0) < -b^2(\theta_0)/(2\theta_0)$. Consider the function $C(\theta)$ defined by $C(\theta_0) = b(\theta_0)$, and $C'(\theta) = -C^2(\theta)/(2\theta)$. Now $b'(\theta_0) < -b^2(\theta_0)/(2\theta_0)$ $= -C^2(\theta_0)/(2\theta_0) = C'(\theta_0)$, so there exists $\theta' > \theta_0$ such that $b(\theta) < C(\theta)$, for every $\theta \in (\theta_0, \theta']$.

But $b'(\theta) - C'(\theta) < \frac{1}{2} [C^2(\theta) - b^2(\theta)]/\theta < 0$, for $\theta \in (\theta_0, \theta']$, so $b(\theta) - C(\theta)$is a strictly decreasing negative function on $(\theta_0, \theta']$.

Let $\theta^* = \sup \{\theta' | b(\theta) < C(\theta)$, for all $\theta \in (\theta_0, \theta')\}$.

If $b(\theta) > C(\theta)$ for some $\theta > \theta_0$, then $\theta^* < \infty$, and $b(\theta) - C(\theta)$ must increase to 0 at θ^* , so $b'(\theta) - C'(\theta)$ must be positive somewhere in (θ', θ^*) where $b(\theta) - C(\theta)$ is still negative. This is impossible because $b'(\theta) - C'(\theta)$ can be positive only if $C^2(\theta) - b^2(\theta) > 0$, that is $C(\theta) < b(\theta)$, which is false on (θ_0, θ^*) . Therefore no θ^* exists, and $b(\theta) \leq C(\theta)$, for all $\theta \in [\theta_0, \infty)$.

Now $C(\theta)$ is determined uniquely by solving the differential equation $C'(\theta) = -C^2(\theta)/(2\theta)$, and using the initial condition $C(\theta_0) = b(\theta_0)$:

$$\frac{dC(\theta)}{d\theta} = -\frac{C^2(\theta)}{2\theta} \text{ , so } \frac{1}{C^2(\theta)} dC(\theta) = \frac{1}{2\theta} d\theta \text{ , and}$$

$$\int -\frac{1}{C^2(\theta)}\, dC(\theta) = \int \frac{1}{2\theta}\, d\theta + K_{\theta_0} \quad ,$$

so the general solution is $1/C(\theta) = \frac{1}{2} \log \theta + K_{\theta_0}$. Using the initial condition $C(\theta_0) = b(\theta_0)$, $1/b(\theta_0) = \frac{1}{2} \log \theta_0 + K_{\theta_0}$, so $K_{\theta_0} = 1/b(\theta_0) - \frac{1}{2} \log \theta_0$.

Thus $C(\theta) = [\frac{1}{2} \log \theta + 1/b(\theta_0) - \frac{1}{2} \log \theta_0]^{-1} = [\frac{1}{2}\log(\theta/\theta_0)+1/b(\theta_0)]^{-1}$.

We have shown that $b(\theta) \le 0$, $|b(\theta)| \le \sqrt{\theta}$, and $b(\theta) \le C(\theta)$, for all $\theta \ge \theta_0$. Thus $b(\theta)$ is bounded between $-\sqrt{\theta}$ and $C(\theta)$, that is, $-\sqrt{\theta} \le b(\theta) \le C(\theta)$, for $\theta \ge \theta_0$. But $C(\theta)$ is everywhere continuous except at $t = \theta_0 e^{-2/b(\theta_0)}$, and $t \in (\theta_0, \infty)$, for all $\theta_0 > 0$. Now $\lim_{\theta \to t^-} C(\theta) = -\infty$ so $C(\theta)$ can be made arbitrarily small, while $-\sqrt{t} = -\sqrt{\theta_0 e^{-2/b(\theta_0)}}$ is finite for all $\theta_0 \in (0, \infty)$. Thus $C(t^-) < -\sqrt{t^-}$, contradicting the condition that $-\sqrt{\theta} \le b(\theta) \le C(\theta)$, so no θ_0 can exist for which $b(\theta_0) < 0$. Therefore $b(\theta) = 0$, for all $\theta \in \textcircled{H} = (0, \infty)$. Since the conditions of Theorem 1.1.5 are satisfied, X is admissible.

§1.3 Remarks on Improving Estimators and Admissibility in Problems with Restricted Parameter Spaces.

Since we will often wish to compare the performace of an estimator in a problem where restrictions have been placed on the parameter space to its performance in the corresponding problem without the restrictions, we will use the following notation:

$\textcircled{H}^U$ denotes the original unrestricted parameter space,

Θ^R the restricted parameter space,

P^U the problem with the unrestricted parameter space, and

P^R the corresponding problem with the restricted parameter space.

In many problems with unrestricted parameter spaces, classical techniques such as minimum variance unbiased estimation and invariance lead to estimators which have desirable properties. When restrictions are placed on the parameter space in such problems, the estimators provided by these techniques are no longer optimal, and there is no natural way to find an optimal estimator (except perhaps by using the maximum likelihood approach, to which we will return). We now have little choice other than to judge estimators on the ordering of their risk functions and to use admissibility and minimaxity as optimality criteria for estimation.

The use of admissibility as a criterion for choosing "good" estimators has been heavily criticized on the following grounds.

1) Admissible estimators may be difficult to characterize, and their form is often too complex to warrant ease of evaluation.
2) Many intuitively reasonable estimators obtained by classical techniques are inadmissible.
3) The gains made by using admissible estimators instead of these intuitively reasonable estimators are not appreciable in many problems.
4) Admissibility depends on the choice of loss function.
5) Many ridiculous estimators are admissible.

The first two of these criticisms are levelled at the form of

admissible estimators and are irrelevant once it has been decided to compare estimators objectively on the basis of their risk functions. Some classically obtained estimators perform well in P^U's, however they are often far from being admissible in the corresponding P^R's. In the problems of interest in this thesis, therefore, considerable gains can be made by using admissible estimators. In answer to the fourth criticism, as we have seen, admissibility is independent of the normalization used in the loss function. For example, if the loss function is known to be of the form $L(a,\theta) = \gamma(\theta)\ W(a,\theta)$, where W is known, and $0 < \gamma(\theta) < \infty$ is unknown, then an estimator will either be admissible subject to each of these loss functions, or inadmissible for each $\gamma(\theta)$. Though admissibility may depend on $W(a,\theta)$, it does not depend on $\gamma(\theta)$, and thus allows a great deal of flexibility in the choice of loss function. Furthermore, admissibility results concerning commonly used estimators tend to be robust against changes in $W(a,\theta)$ as well, and the same method may be used to provide an improvement over the inadmissible estimator under consideration. For example, in certain problems (e.g. Brewster & Zidek (1974)) the "usual" estimator ϕ_0 will be generalized Bayes with respect to a prior Π, but an improvement will be obtained by using the generalized Bayes estimator with respect to a prior Π^*. It will often be the case in these problems that the same prior Π^* can be used to improve ϕ_0 for a large class of loss functions. To this extent, admissibility does not depend on the choice of the loss function, although this choice often determines the estimator used. In some problems (see, for example, Section 2.1), even the estimator which improves ϕ_0 does not depend on the loss function. To dispose of the fifth criticism, we may combine admissibility with a

criterion such as minimaxity, so that those ridiculous estimators which favour some states of nature and ignore the others completely may be eliminated. However, minimaxity is not independent of order-preserving fluctuations in $L(a,\theta)$ (that is, changes in $\gamma(\theta)$), and so the combined criteria may be said to depend too much on the choice of loss function.

In this thesis, we are primarily interested in finding "reasonable" estimators for problems involving restricted parameter spaces. Since most classical techniques fail to work in $\mathcal{P}^R$, we will use the following approach, which is geared towards satisfying the criteria of admissibility and minimaxity.

Let us start with an estimator ϕ_0 which possesses certain desirable properties in $\mathcal{P}^U$. In $\mathcal{P}^R$, of course, ϕ_0 will characteristically have the undesirable property of taking on values outside of Θ^R. However, it may be possible to perceive an estimator ϕ^* which is better than ϕ_0 in $\mathcal{P}^R$, and which also possesses desirable properties in $\mathcal{P}^R$ (such as admissibility and minimaxity). In fact, if ϕ_0 is minimax in $\mathcal{P}^U$, then it may happen that ϕ^* will be minimax in $\mathcal{P}^R$ if and only if ϕ^* is no worse than ϕ_0 in $\mathcal{P}^R$. This occurs in the Poisson problem of Section 3.2 and the normal example of Sections 2.3 and 2.4, both of which are exponential family problems where Θ^R is unbounded in one direction. In such a situation, a "reasonable" estimator might be one which is better than ϕ_0 and which is admissible.

It should be observed that in $\mathcal{P}^R$'s where the loss function is of the form $\gamma(\theta)\, W(a,\theta)$, where $\gamma(\theta)$ is unspecified, the criterion of minimaxity is of no use. Here, in particular, it seems reasonable to search for an estimator which is admissible and improves some "optimal"

estimator in P^U, as these combined criteria are independent of fluctuations in $\gamma(\theta)$.

Chapter II of this thesis will be devoted to the presentation and discussion of techniques for improving estimators. It should be kept in mind that we are actually looking for admissible improvements, and in many problems this means that the improvements must be, in some sense, "smooth", because admissible estimators must be generalized Bayes. Maximum likelihood estimators are not smooth enough to be admissible in many problems with restricted parameter spaces (see, for example, Kale (1970)). Some of the techniques to be discussed are introduced with this smoothness property in mind.

Although the techniques in Chapter II are generally applicable, it should be remembered that the improvement which can be made over classical methods of estimation is much more significant in P^R than in P^U. This is a further justification of the use of the admissibility criterion for estimation in P^R. Chapters III and IV discuss the improvements which can be made in estimation of a restricted Poisson parameter, and in estimation of ordered binomial parameters, respectively. These are two problems occuring in practice which have restricted parameter spaces for which we shall try to find reasonable estimators.

CHAPTER II

Improving Estimators

§2.0. Introduction and Summary

The techniques presented in this chapter for improving estimators are quite general, in that they may be applied no matter what family of underlying distributions is considered. However, they may succeed in providing a better estimator only if certain conditions are met by the structures of the sample space, the parameter space, and the loss function. Some of them are particularly suited to problems where the parameter space is restricted. The five techniques considered here are:

1. Projection onto the closed convex hull of the parameter space.
2. Optimizing within a subclass of the available decision rules.
3. Searching among a complete class.
4. Step function approach — an extension to #2.
5. Combining estimators.

The first of these will improve any estimator which takes on values outside the convex hull of the parameter space, and is therefore quite applicable in $\mathcal{P}^R$, the restricted parameter space problem, for improving estimators which are admissible in $\mathcal{P}^U$, the unrestricted problem. The second method restricts the class of available decision rules to a small enough subclass which contains the original estimator. If, for each parameter value, the optimal choice of the estimators in this subclass is not the original estimator, then, under certain condi-

tions, we may find an estimator in this subclass which is better than the original estimator. This technique was introduced in its simplest form by Goodman (1953), and extended by Blight (1971). The form used here is discussed in Brewster and Zidek (1974) and Arnold (unpublished). The method has applications in both P^R and P^U.

The third technique consists of searching through a complete class or through a subclass of the class of admissible estimators for an estimator which improves the original estimator, and is admissible. This approach is used in recognition of the possibility that an estimator, which improves the original estimator, may itself be improved. The eventual goal of improving estimators is to obtain an estimator which cannot itself be improved, that is, an admissible estimator. This technique makes use of the admissibility and complete class theorems discussed in Section 1.1.

The fourth technique was introduced by Brewster and Zidek (1974) and consists of constructing a sequence of step-function estimators, each of which improves the original estimator, and then showing that the limit of this sequence also improves the original estimator. This technique is valuable in that it provides an estimator which is better than the original estimator and has the additional property that it is "smooth" — a property required of admissible estimators in many problems. Maximum likelihood estimation for P^R is often achieved by projecting onto the boundary of $\textcircled{H}^R$, and this projection is not "smooth", so the first method does not always provide admissible estimators. For example, Sackrowitz and Strawderman (1974) have proved that, for estimating ordered binomial parameters when the sample sizes are not very small, the maximum likelihood estimator, which is a pro-

jection of the MLE for P^U onto the boundary of the convex parameter space $\Theta^R = \{0 \leq \theta_1 \leq \ldots \leq \theta_m \leq 1\}$, is inadmissible, subject to many loss functions, including sum of squared errors. Their proof is based on certain consistency requirements of admissible estimators on the boundary of Θ^R, and is analogous to the claim that the MLE does not approach the boundary $\hat{\theta}_i = \hat{\theta}_j$ in a smooth enough fashion to be admissible.

By recursively applying the technique of optimizing within subclasses of $\mathcal{D}$, we obtain, in the limit, an estimator which is smooth, and beats the original estimator. This approach will sometimes provide a method of constructing admissible estimators. This step-function approach has been used by Brewster and Zidek (1974) and Brewster (1974) to obtain smooth improvements on equivariant estimators which are also generalized Bayes and which turn out to be admissible.

The first four techniques provide improvement with the view of eventually obtaining an admissible estimator. If improvement is desired in reducing the maximum risk of an estimator, we can use the fifth technique. For any two estimators, this technique will provide an estimator whose risk function is bounded above by a convex combination of the individual risk functions. If each of the two estimators improves an original estimator, the fifth technique may be used to provide an estimator which comes closer to being an equalizer estimator than the other estimators, and which therefore becomes closer to being a minimax estimator while still improving the original estimator.

§2.1 Projection onto the Closed Convex Hull of the Parameter Space

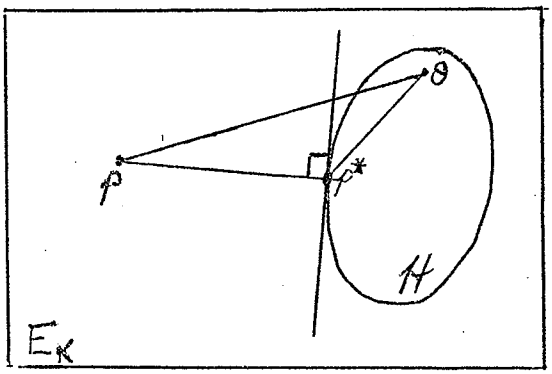
Definition: The closed convex hull of a set S is the closure of the intersection of all convex sets containing S (that is, the closed convex hull of S is the smallest closed convex set containing S).

If an estimator takes on values outside the closed convex hull, H, of the parameter space, $\textcircled{H}$, then for each value p outside H, there may be a value p^* on the boundary of H which is closer to each $\theta \in \textcircled{H}$ than is p. If this is the case, then the estimator constructed by replacing the values that the original estimator takes on outside H by their corresponding points on the boundary of H may have uniformly smaller risk than the original estimator. One important case where this works is given by the following theorem.

Theorem 2.1. Suppose that the action space, A, and the parameter space, $\textcircled{H}$, are subspaces of k-dimensional Euclidean space, E_k. Consider loss functions, $L(a,\theta)$, which are strictly increasing functions of $||a - \theta||$ in a, for each given $\theta \in \textcircled{H}$. If an estimator, $\phi(X)$, takes on values outside H with positive probability for each $\theta \in \textcircled{H}$, then there exists a projection onto H, $\phi^*(X)$, which is uniformly better than $\phi(X)$.

Proof: Let p be any point outside H. Choose p^* as the value of a which minimizes the distance $||p - a||$ over all $a \in H$. Then, by the Separating Hyperplane Theorem, there exists a k-1 dimensional hyperplane passing through p^*, which separates p from H, and is orthogonal to the line joining p and p^*. For values $\phi(X) = p$ which are outside H, define the projected estimator $\phi^*(X)$ to be the correspondingly choosen p^*. Otherwise $\phi^*(X) = \phi(X)$. Now, for each

$\theta \in H$, the triangle formed by the points θ, p, p^* must have an angle of at least 90° at p^*, so the side opposite p^* is the longest, and thus $||p - \theta|| > ||p^* - \theta||$. Since the loss function $L(a,\theta)$ is assumed to be strictly increasing in $||a - \theta||$ (for given θ), we have that $L(p,\theta) > L(p^*,\theta)$, for each $\theta \in H$. Therefore the projected estimator $\phi^*(X)$ has smaller loss than $\phi(X)$ whenever they are unequal, that is, whenever $\phi(X)$ is outside H . Thus, for every $\theta \in \textcircled{H} \subset H$,



$$E_\theta[(\phi(X),\theta)|\phi(X) \notin H\ ] > E_\theta[L(\phi^*(X),\theta)|\phi(X) \notin H\ ] .$$

Since $\phi^*(X) = \phi(X)$ for $\phi(X) \in H$, we have that

$$\begin{aligned} r_\phi(\theta) &= E_\theta[L(\phi(X),\theta)|\phi(X) \notin H]\ P_\theta(\phi(X) \notin H) \\ &\qquad + E_\theta[L(\phi(X),\theta)|\theta(X) \in H]\ P_\theta(\phi(X) \in H) \\ &> E_\theta[L(\phi^*(X),\theta)|\phi(X) \notin H]\ P_\theta(\phi(X) \notin H) \\ &\qquad + E_\theta(L(\phi^*(X),\theta)|\phi(X) \in H]\ P_\theta(\phi(X) \in H) \\ &= r_{\phi^*}(\theta) , \qquad \forall\ \theta \in \textcircled{H} , \text{ so } \phi^* \text{ uniformly improves } \phi . \| \end{aligned}$$

Example 2.1.1.. (Proportion of Fraternal Twins)

Assume, for the sake of simplicity, that the sexes of a pair of twins have a trinomial distribution with the following cell probabilities:

MM	MF	FF	
¼	½	¼	if the twins are fraternal,
½	0	½	if the twins are identical.

From a sample of N pairs of twins, we observe N_1, N_2, N_3 pairs of twins in the MM, MF, FF classes, respectively, but when both of a pair of twins have the same sex, we cannot tell (from the data) whether they are identical or fraternal twins.

To estimate the proportion $\theta \in \textcircled{H} = [0,1]$ of fraternal twins among all twins, using the normalized squared-error loss function $L(a,\theta) = (a-\theta)^2/[\theta(1-\theta)]$, we notice that N_2 is a sufficient statistic and has the binomial distribution with parameters N and $\theta/2$. Since the series

$$E_\theta[g(N_2)] = \sum_{n_2=0}^{N} \binom{N}{n_2} \left(\frac{\theta}{2}\right)^{n_2} \left(1 - \frac{\theta}{2}\right)^{N-n_2} g(n_2)$$

$$= (2-\theta)^N \sum_{n_2=0}^{N} \binom{N}{n_2} [\theta/(2-\theta)]^{n_2} g(n_2)$$

is $(2-\theta)^N$ times a polynomial in $\theta/(2-\theta)$, it cannot be zero for each $\theta \in \textcircled{H}$ unless $g(n_2)$ is identically zero for $n_2 = 0,\ldots,N$. Thus N_2 is a complete sufficient statistic for θ, and the unbiased estimator $2N_2/N$ is the uniformly minimum variance unbiased estimator. But $2N_2/N$ may be larger than 1 and thus outside the closed convex hull of the parameter space, $H = [0,1]$. The conditions of Theorem 2.1 are satisfied, so the estimator

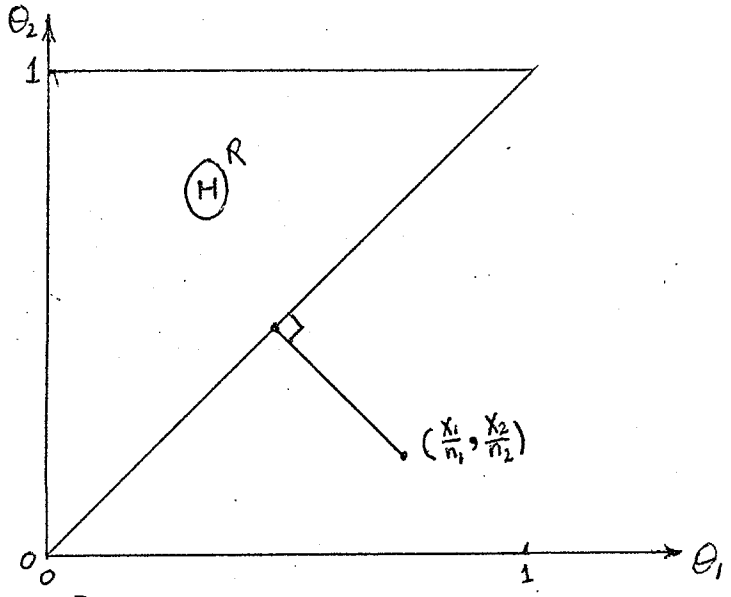
$$\phi^*(N_2) = \begin{cases} 2N_2/N & N_2 \le N/2 \\ 1 & N_2 > N_2 \end{cases}$$

is uniformly better than $2N_2/N$.

Example 2.1.2. (Ordered Binomial Parameters)

In bioassay, the probability of a particular response, for

example, death, cure, remission, etc. is monotonic in the dosage level of the treatment administered. In reliability, the probability that a processed product is defective decreases when an improvement has been made in the process. In these kinds of problems, we often wish to estimate a number of binomial probabilities subject to certain order restrictions. Here, we consider estimation of 2 ordered binomial parameters, $\textcircled{H}^R = \{(\theta_1,\theta_2) | 0 < \theta_1 < \theta_2 < 1\}$, on the basis of two independent random variables $X_1 \sim B(n_1,\theta_1)$, $X_2 \sim B(n_2,\theta_2)$. $(X_1/n_1, X_2/n_2)$ is both the UMVUE for (θ_1,θ_2) and the MLE for (θ_1,θ_2) in the absence of the order restriction. This estimator takes



on values outside $\textcircled{H}^R$ whenever $n_2X_1 > n_1X_2$. The loss function, $L(a,\theta) = ||a - \theta||^2 = (a_1-\theta_1)^2 + (a_2-\theta_2)^2$ satisfies the condition for Theorem 2.1, and the convex hull $\textcircled{H}^R$ is $H = \{(\theta_1,\theta_2) | 0 \le \theta_1 \le \theta_2 \le 1\}$. Since points on the circle $(X_1/n_1 - \theta_1)^2 + (X_2/n_2 - \theta_2)^2 = L$ are equidistant from $(X_1/n_1, X_2/n_2)$, the closest point to $(X_1/n_1, X_2/n_2)$ in H is its projection onto the line $\theta_1 = \theta_2$, $(\frac{1}{2}X_1/n_1 + \frac{1}{2}X_2/n_2 , \frac{1}{2}X_1/n_1 + \frac{1}{2}X_2/n_2)$. Applying Theorem 2.1, the estima-

tor

$$\phi^*(X_1,X_2) = \begin{cases} (X_1/n_1, X_2/n_2) & n_2X_1 \le n_1X_2 \\ (\tfrac{1}{2}X_1/n_1 + \tfrac{1}{2}X_2/n_2,\ \tfrac{1}{2}X_1/n_1 + \tfrac{1}{2}X_2/n_2) & n_2X_1 > n_1X_2 \end{cases}$$

uniformly improves $(X_1/n_1, X_2/n_2)$.

It should be noted that, unless $n_1 = n_2$, $\phi^*(X_1,X_2)$ is not the restricted maximum likelihood estimator for this problem. The likelihood is maximized by projecting instead to $([X_1+X_2]/[n_1+n_2]\ ,\ [X_1+X_2]/[n_1+n_2])$, whenever $n_2X_1 > n_1X_2$. We will return to this problem in Chapter IV, where we will show that the maximum likelihood estimator does not always improve $(X_1/n_1, X_2/n_2)$.

§2.2. Optimizing within a Subclass of $\mathcal{D}$

Consider a subclass, $D \subset \mathcal{D}$, of the decision rules whose elements $d_c : c \in \mathbb{R}^1$ can be indexed by a single real-valued parameter c. In an attempt to improve $d_{c0} \in D$, we first find the optimal estimator in D for each $\theta \in \textcircled{H}$, that is, for each $\theta \in \textcircled{H}$, we choose d_{c_θ} to minimize $r_{d_c}(\theta)$ for variations in c . It may happen that the optimal estimator is the same for each θ , that is, $c_\theta \equiv c^*$. If $c^* \neq c^0$, then d_{c^*} will obviously improve d_{c0} . This is the basic technique discussed by Goodman (1953). [See also Example 2.2.1.] .

Or, it may happen that the optimal estimators depend on θ , and yet the c_θ's are bounded away from c^0 by c^* , that is, either $c_\theta \ge c^* > c^0$, or $c_\theta \le c^* < c^0$. Then it seems reasonable to use d_{c^*} instead of d_{c0} , and this is justified by Theorem 2.2.1. [See also, Brewster and Zidek (1974), Arnold (unpublished), and Example 2.2.2.] .

Definition: A function $h : \mathbb{R}^1 \to \mathbb{R}^1$ is called strictly bowl-shaped in u if there exists u_0 such that h decreases in u for $u \le u_0$ and h increases in u for $u \ge u_0$.

Theorem 2.2.1. If $r_{d_c}(\theta)$ is strictly bowl-shaped in c for each $\theta \in \Theta$, and if $\{c_\theta | \theta \in \Theta\}$ is bounded away from c^0 by c^*, then $d_{c^*}(X)$ uniformly improves $d_{c^0}(X)$.

Proof: Assume without loss of generality that $c_\theta \le c^* < c^0$, for all $\theta \in \Theta$. For each fixed $\theta \in \Theta$, c_θ minimizes $r_{d_c}(\theta)$ for variations in c, and the strictly bowl-shaped nature of $r_{d_c}(\theta)$, as a function of c, allows us to conclude that

$$r_{d_{c^*}}(\theta) < r_{d_{c^0}}(\theta), \text{ for all } \theta \in \Theta. \quad \|$$

Example 2.2.1. Estimation of the Variance of a Normal Distribution

$X_1, \ldots, X_n$ form a random sample of size n from a normal population with unknown mean μ and unknown variance τ. We wish to estimate τ using the squared-error loss function, $L(a,\tau) = (a-\tau)^2$. Consider the restricted class of estimators

$$D = \{d_c(X_1, \ldots, X_n) | d_c(X_1, \ldots, X_n) = cT, T = \sum_{i=1}^{n} (X_i - \bar{X})^2 \}.$$

Both the maximum likelihood estimator ($c^0 = 1/n$), and the uniformly minimum-variance unbiased estimator ($c' = 1/(n-1)$), are in this class. For each $\tau \in \mathcal{T} = (0, \infty)$,

$$r_{d_c}(\tau) = E_\tau[(cT - \tau)^2] = c^2 E_\tau(T^2) - 2c\tau E_\tau(T) + \tau^2$$

is an upward-facing parabola in c. This satisfies the bowl-shaped requirement. Since T/τ has the χ^2-distribution on $n-1$ degrees of

freedom, $E_\tau(T) = (n-1)\tau$, $E_\tau(T^2) = (n-1)(n+1)\tau^2$. Thus $r_{d_c}(\tau)$ is minimized in c for

$$c_\theta = \frac{\tau E_\tau(T)}{E_\tau(T^2)} = \frac{1}{n+1} \quad , \text{ for each } \tau \in \mathcal{T} .$$

Since both $c^0 = 1/n$, $c' = 1/(n-1) > c^* = 1/(n+1) \geq c_\theta$, for all $\theta \in \Theta$, $T/(n+1)$ is uniformly better than both T/n and $T/(n-1)$.

Example 2.2.2. — Estimation of a Poisson parameter

Suppose X is a random variable from the Poisson distribution with parameter $\theta \in \Theta = (0, \infty)$. We wish to estimate θ under normalized squared-error loss, $L(a,\theta) = (a-\theta)^2/\theta$.

(a) Consider estimators in the class $D = \{d_c(X) = cX, \ c > 0\}$. For each $\theta \in \Theta$, $r_{d_c}(\theta) = \theta^{-1}E_\theta[(cX - \theta)^2] = c^2E_\theta(X^2)/\theta - 2cE_\theta(X) + \theta$,

which is an upward-facing parabola in c, (this satisfies the bowl-shaped requirement) and is minimized by $c_\theta = \theta E_\theta(X)/E_\theta(X^2) = \theta/(1+\theta)$. As $\theta \to \infty$, c_θ gets arbitrarily close to $c^0 = 1$, and therefore c_θ cannot be uniformly bounded away from 1, so Theorem 2.2.1 may not be applied. This is hardly surprising, as X is admissible in this problem (see Section 1.2) .

Now consider the restricted parameter space $\Theta^R = (0,B]$. This may arise in practice if a change occurs in a Poisson process which decreases θ , the mean rate at which events occur. If it was known that the rate was in fact B, before the process was changed, then the parameter space has been restricted to $\Theta^R = (0,B]$.

Since $c_\theta = \theta/(1+\theta)$ is an increasing function of θ , the c_θ's

are now bounded above by $B/(1+B)$ which is smaller than 1, and we can apply Theorem 2.2.1 to conclude that $BX/(1+B)$ uniformly improves X. It should be noted that had the parameter space been restricted to $[B, \infty)$, then c_θ would not have been uniformly bounded away from $c^0 = 1$, and Theorem 2.2.1 could not have been applied.

(b) Now, for fixed k, consider estimators in the class $D_k = \{d_c(X) | c > 0\}$, where

$$d_c(X) = \begin{cases} X + c & X \le k \\ X & X > k \end{cases} .$$ For each $\theta \in \Theta = (0, \infty)$,

$$r_{d_c}(\theta) = \theta^{-1} E_\theta\{[d_c(X) - \theta]^2\}$$

$$= \theta^{-1} E_\theta[(X+c-\theta)^2 | X \le k] \, P_\theta(X \le k) + \theta^{-1} E[(X-\theta)^2 | X > k] \, P_\theta(X > k),$$

which is an upward-facing parabola in c with minimum $c_\theta = \theta - E_\theta[X | X \le k]$.

Although $c_\theta > 0$, c_θ gets arbitrarily close to $c^0 = 0$ (which provides the usual estimator X), so c_θ cannot be uniformly bounded away from 0, and X cannot be improved using Theorem 2.2.1. (Again, of course, this isn't surprising since X was shown to be admissible in Section 1.2.).

Now consider the restricted parameter space $\Theta^R = [B, \infty)$. It will be proved in Section 3.2 that the c_θ's are now bounded below by $c^* = B - E_B[X | X \le k]$, which is larger than $c^0 = 0$, so the conditions of Theorem 2.2.1 are satisfied, and the estimator

$$d_{c^*}(X) = \begin{cases} X + B - E_B[X | X \le k] & X \le k \\ X & X > k \end{cases}$$

uniformly improves X. This result will be used repeatedly in Section

3.2 to obtain a "smooth" estimator for θ which takes on all of its values in $\mathcal{H}^R$.

Here, $d_{c^*}(0) = B - E_B[X|X \le k] < B$, so d_{c^*} does not take on all of its values in $\mathcal{H}^R = [B, \infty)$.

A Modification

Although we cannot bound c_θ uniformly away from c^0 in some problems, we may still be able to improve an estimator using an ad-hoc modification of the above procedure. This modification is unreliable in that it is difficult to predict whether or not it will work in a given problem. It consists of substituting d_{c_θ} for θ in the formula for d_{c_θ}, and then solving for c_θ (which will be independent of θ). [This modification was suggested by Arnold (1973).]

Example 2.2.3. Suppose $\underline{X} = (X_1,\ldots,X_n)$ forms a random sample from a geometric distribution with mean $\theta \in \mathcal{H} = (1, \infty)$. We wish to estimate θ under squared-error loss, $L(a,\theta) = (a-\theta)^2$. The usual estimator, $\bar{X}$, is uniformly minimum-variance unbiased. Consider estimators in the class $D = \{d_c(\underline{X}) = c\bar{X} \mid c > 0\}$. For each θ in $\mathcal{H}$, $r_{d_c}(\theta) = E_\theta[(c\bar{X} - \theta)^2] = c^2E(\bar{X}^2) - 2c\theta E(\bar{X}) + \theta^2$ is strictly bowl-shaped in c, and the best choice of c is

$$c_\theta = \frac{\theta E_\theta(\bar{X})}{E_\theta(\bar{X}^2)} = \frac{\theta^2}{\frac{\theta(\theta-1)}{n} + \theta^2} = \frac{n\theta}{n\theta+\theta-1} .$$ This

cannot be uniformly bounded away from $c^0 = 1$, for as $\theta \to 1$, $c_\theta \to 1$. Thus, Theorem 2.1 cannot be applied. Suppose, however, that $c_\theta\bar{X}$ is is substituted for θ, in the expression for c_θ. This gives the equation $c_\theta = nc_\theta\bar{X}/[(n+1)c_\theta\bar{X}-1]$ which we solve for c_θ, yielding

$c_\theta = (n\bar{X}+1)/[\bar{X}(n+1)]$, and therefore the estimator $(n\bar{X}+1)/(n+1)$. This estimator is not in the class D, but its mean-squared-error is

$$r_{\frac{n\bar{X}+1}{n+1}}(\theta) = \mathrm{Var}\left(\frac{n\bar{X}}{n+1}\right) + \left\{ E\left[\frac{n\bar{X}+1}{n+1}\right] - \theta \right\}^2$$

$$= \left(\frac{n}{n+1}\right)^2 \frac{\theta(\theta-1)}{n} + \left[\frac{n\theta+1 - (n+1)\theta}{n+1}\right]^2$$

$$= \frac{n\theta(\theta-1) + (\theta-1)^2}{(n+1)^2} < \frac{n\theta(\theta-1) + \theta(\theta-1)}{(n+1)^2} = \frac{\theta(\theta-1)}{n+1}$$

$$< \frac{\theta(\theta-1)}{n} = r_{\bar{X}}(\theta) ,$$

uniformly smaller than that of $\bar{X}$, so the estimator produced by this modification, $(n\bar{X}+1)/(n+1)$, is uniformly better than $\bar{X}$.

§2.3. Searching among a complete class

The improvements made by projecting onto the convex hull of Ⓗ , or optimizing within a subclass D of $\mathcal{D}$, though in some cases considerable, do not provide admissible estimators. It should be remembered that we are searching for an estimator which is reasonable, that is, improves some given estimator and is admissible. Such an estimator (or its equivalent) must belong to an essentially complete class. If we search through an essentially complete class until we find an estimator which improves the given estimator, and show that the estimator we find is admissible, then it must be a reasonable estimator. This search may take a long time however, especially if the original estimator is close to being admissible, or if the essentially complete class is large.

Example 2.3. Let X be normally distributed with unknown mean θ, and variance known to be 1. It was shown in Chapter I that, subject to squared error loss, X was an admissible minimax estimator for θ. Suppose it is known that $\theta \geq B$. Then X is no longer admissible, for X takes values outside $H = [B, \infty)$, and we can apply Theorem 2.1 to conclude that the projection

$$\psi(X) = \begin{cases} X & X \geq B \\ B & X < B \end{cases} \quad \text{uniformly improves } X .$$

However, it has been shown by Kale (1970) that, in continuous exponential family problems, generalized Bayes estimators are analytic functions of X. Here, $\psi(X)$ is not differentiable at $X = B$, and therefore not analytic. In this problem, by Sacks (1963) result, the class of generalized Bayes estimators is essentially complete, so $\psi(X)$ is inadmissible.

In this problem X is still minimax. To prove this, we make use of an argument similar to the one used in the proof of Theorem 1.1.5. The risk function of any estimator $d(X)$ with bias function $b(\theta)$ is bounded below (using the Cramér-Rao inequality) as

$$r_d(\theta) \geq b^2(\theta) + [1 + b'(\theta)]^2 / E_\theta[\frac{d}{d\theta} \log f_\theta(X)]^2 .$$

(See Hodges and Lehmann (1950) for proof of regularity conditions.)

But $E_\theta[\frac{d}{d\theta} \log e^{-\frac{1}{2}(X-\theta)^2}]^2 = E_\theta[\frac{d}{d\theta}(-\frac{1}{2}(X-\theta)^2)]^2 = E_\theta[(X-\theta)^2] = 1$,

so $r_d(\theta) \geq b^2(\theta) + [1+b'(\theta)]^2$.

Now, if X is not minimax, then there exists an estimator d with $r_d(\theta) \leq \sup_{\theta \in H} r_X(\theta) - \epsilon$, for some $0 < \epsilon < 1$. But $r_X(\theta) = E_\theta[(X-\theta)^2] = 1$,

so $1 - \epsilon \geq r_d(\theta) \geq b^2(\theta) + [1 + b'(\theta)]^2$, for all $\theta \geq B$.

This implies that

$$[1 + b'(\theta)]^2 \leq 1 - \epsilon \ , \quad \text{so} \quad b'(\theta) \leq \sqrt{1 - \epsilon} - 1 < 0 \ ,$$

so $b(\theta) \to -\infty$ as $\theta \to \infty$ which contradicts $b^2(\theta) \leq 1 - \epsilon$. (Lehmann (1950)) . Thus X is minimax, and any estimator which improves X is also minimax. It is of interest to note, that, in this problem, an estimator will be reasonable (that is, an admissible improvement of X) if and only if it is admissible and minimax.

We now search through the class of generalized Bayes estimators for an admissible minimax estimator. As X is generalized Bayes with respect to Lebesgue prior measure in $\mathcal{P}^U$ (see Section 1.1), it might be reasonable to consider the generalized Bayes estimator with respect to Lebesgue prior measure on $[B, \infty)$, which is given by

$$\phi(X) = \frac{\int_B^\infty \theta \, e^{-\frac{1}{2}(X-\theta)^2} d\theta}{\int_B^\infty e^{-\frac{1}{2}(X-\theta)^2} d\theta} = X + \frac{e^{-\frac{1}{2}(X-B)^2}}{\int_B^\infty e^{-\frac{1}{2}(X-\theta)^2} d\theta} \ .$$

This is the estimator shown to be admissible by Theorem 1 of Katz (1961). (The proof involves showing that $\phi(X)$ is a limit of Bayes rules with respect to exponential prior densities,

$$\pi_\sigma(\theta) = \frac{e^{-\theta/\sigma}}{\int_B^\infty e^{-\theta/\sigma} \, d\theta} \ , \quad \theta \geq B \ ,$$

and applying the method of Blyth (1951)). Katz also shows that $\phi(X)$ is minimax. Therefore $\phi(X)$ is an admissible improvement of X. Subject to loss functions of the form $\gamma(\theta)(a-\theta)^2$, $\phi(X)$, though not generally minimax, is reasonable for all $\gamma(\theta)$.

Unfortunately, it turns out that in many problems, the prior which leads to an admissible minimax estimator is much more complicated in form and thus much more difficult to find by trial and error. Thus the hit or miss method of searching the class of generalized Bayes estimators may be inadequate. In such a situation it would be desirable to be able to construct an admissible estimator which improves the original estimator. A method of construction which improves the original estimator and sometimes provides an estimator which is admissible is the topic of Section 2.4.

§2.4. Constructing Reasonable Estimators — A Step-function approach

Return to the problem of estimating θ, subject to squared error loss, when $X \sim N(\theta,1)$ and it is known that $\theta \geq B$. Now fix $k_1 \in \mathbb{R}$ and consider estimators of the form

$$\begin{cases} X + c & X \leq k_1 \\ X & X > k_1 \end{cases},$$

where $c \in \mathbb{R}$. The risk function of such an estimator is

$$E_\theta[(X + c - \theta)^2 | X \leq k_1]\, P_\theta(X \leq k_1) + E_\theta[(X-\theta)^2 | X > k_1] P_\theta(X > k_1),$$

which is an upward-facing parabola in c and is minimized for given θ by $c_{k_1,\theta} = \theta - E_\theta[X | X \leq k_1]$. Since $E_\theta[X | X \leq k_1] < E_\theta(X) = \theta$, $c_{k_1,\theta}$ is strictly positive. Furthermore,

$$c_{k_1,\theta} = \theta - \left[\int_{-\infty}^{k_1} x\, e^{-\frac{1}{2}(x-\theta)^2} dx\right] \Big/ \left[\int_{-\infty}^{k_1} e^{-\frac{1}{2}(x-\theta)^2} dx\right]$$

$$= - \left[\int_{-\infty}^{k_1-\theta} ze^{-z^2} \, dz \right] \Big/ \left[\int_{-\infty}^{k_1-\theta} e^{-z^2} \, dz \right] = - E[Z | Z \le k_1 - \theta]$$

where $Z \sim N(0,1)$, so that $c_{k_1,\theta}$ is a strictly increasing function of θ , and is minimized by $c_{k_1,B} = B - E_B[X | X \le k_1]$. Therefore,

$$c_{k_1,\theta} \ge c_{k_1,B} > 0 , \qquad \text{for all} \quad \theta \ge B ,$$

and by Theorem 2.2.1 the estimator

$$d_1(X) = \begin{cases} X + c_{k_1,B} & X \le k_1 \\ X & X > k_1 \end{cases}$$

uniformly improves X .

Since $c_{k,\theta}$ is a strictly decreasing function of k, for each $\theta \in \text{Ⓗ}$, we may repeat the above argument to conclude that for fixed $k_2 < k_1$, the estimator

$$d_2(X) = \begin{cases} X + c_{k_2,B} & X \le k_2 \\ X + c_{k_1,B} & k_2 < X \le k_1 \\ X & X > k_1 , \end{cases}$$

uniformly improves $d_1(X)$, where $c_{k_2,B}$ bounds the $c_{k_2,\theta}$'s away from $c_{k_1,B}$. We may continue in this manner indefinitely, and for the partition $P = \{\ldots,k_n \ldots,k_3,k_2,k_1\}$ we obtain the sequence of estimators $\{d_n(X)\}$. We would like to interchange limit and integral (expectation) and conclude that $\lim_{n\to\infty} d_n(X)$ is better than all d_n's (and therefore also better than X) . Since $(d_n(x) - \theta)^2 \ge 0$, an

application of Fatou's Lemma yields

$$r_{\lim_{n\to\infty} d_n}(\theta) = E_\theta[(\lim_{n\to\infty} d_n(X) - \theta)^2]$$

$$\leq \lim_{n\to\infty} E_\theta[(d_n(X) - \theta)^2] = \lim_{n\to\infty} r_{d_n}(\theta) \quad ,$$

justifying this conclusion. [Fatou's Lemma is stated in the Appendix.] However, this result holds for any partition $P_j = \{\ldots,k_{nj},\ldots,k_{3j}, k_{2j}, k_{1j}\}$. The finer the partition becomes, and the larger k_{1j} is, the smoother the resultant estimator will be. By taking a sequence of partitions $\{P_j\}$ such that $\lim_{j\to\infty} k_{1j} = \infty$, $\lim_{n\to\infty} k_{nj} = -\infty$, for all j, and $\lim_{j\to\infty} \max_i (k_{ij} - k_{i+1,j}) = 0$, we obtain a sequence of estimators $\{\phi_j(X) = \lim_{n\to\infty} d_{nj}(X)\}$, each improving X, which converges pointwise to the estimator given by

$$\phi(x) = x + c_{x,B} = x + B - E_B[X|X \leq x]$$

$$= x + B - \left[\int_{-\infty}^{x} u\, e^{-\frac{1}{2}(u-B)^2}\, du\right] \quad \left[\int_{-\infty}^{x} e^{-\frac{1}{2}(u-B)^2}\, du\right]$$

$$= x - \left[\int_{-\infty}^{x} (u-B)\, e^{-\frac{1}{2}(u-B)^2}\, du\right] \quad \left[\int_{-\infty}^{x} e^{-\frac{1}{2}(u-B)^2}\, du\right] \quad .$$

By another application of Fatou's Lemma, we can conclude that

$$r_\phi(\theta) = E_\theta[(\lim_{j\to\infty} \phi_j(X) - \theta)^2]$$

$$\leq \liminf_{j\to\infty} E_\theta[(\phi_j(X) - \theta)^2]$$

$$= \liminf_{j\to\infty} r_{\phi_j}(\theta) \leq r_X(\theta) = 1 \ ,$$

that is, $\phi(X)$ is as good as X . We cannot make the stronger

statement that $\phi(X)$ improves X, since ϕ_{j+1} is not necessarily better than ϕ_j. Here, X and $\phi(X)$ are quite different in form, and it would be surprising if they were equivalent. We will see below that $\phi(X)$ does, in fact, improve X.

Substitution of u by $x - \theta + B$ leads to

$$\phi(x) = x - \left[\int_B^\infty (x-\theta)\, e^{-\frac{1}{2}(x-\theta)^2} d\theta\right] \Big/ \left[\int_B^\infty e^{-\frac{1}{2}(x-\theta)^2} d\theta\right]$$

$$= \left[\int_B^\infty \theta\, e^{-\frac{1}{2}(x-\theta)^2} d\theta\right] \Big/ \left[\int_B^\infty e^{-\frac{1}{2}(x-\theta)^2} d\theta\right] ,$$

defining the same estimator which was found to be an admissible and minimax improvement of X in Section 2.3. Here we have given an alternative demonstration of the minimaxity of $\phi(X)$, while at the same time constructing $\phi(X)$ as a "smooth" improvement to X rather than guessing it as in Section 2.3

Brewster and Zidek (1974) and Brewster (1974) have used the above method to obtain a sequence of step-function estimators whose limit improves an equivariant estimator, and in some problems can be shown to be admissible. It is used in Chapter III to improve upon a Poisson random variable X as an estimator of the parameter θ of the distribution of X, when the parameter space is restricted by either an upper or lower bound.

§2.5. Combining Estimators

In our pursuit of reasonable estimators, we may encounter two estimators, $\phi(X)$ and $\psi(X)$, neither of which need improve the original estimator $d(X)$, such that some convex combination of their

risk functions is nowhere above the risk of the original estimator. That is, there may exist $c \in (0,1)$ such that

$$cr_{\phi}(\theta) + (1 - c)\, r_{\psi}(\theta) \leq r_{d}(\theta) , \quad \text{for all} \quad \theta \in \Theta .$$

In such a situation, we may apply the following theorem to conclude that the same convex combination of the estimators, $c\phi(X) + (1 - c)\, \psi(X)$, is as good as $d(X)$.

Theorem 2.5.1. Subject to a loss function $L(a,\theta)$ which is convex in a for each fixed θ, consider the estimator defined as $\xi(X) = c\phi(X) + (1 - c)\, \psi(X)$, for some $c \in (0,1)$. Then

$$r_{\xi}(\theta) \leq cr_{\phi}(\theta) + (1 - c)\, r_{\psi}(\theta), \quad \text{for all} \quad \theta \in \Theta .$$

Proof: For each $x \in \mathcal{X}$, convexity implies that $L[\xi(X),\theta] \leq cL[\phi(X),\theta] + (1-c)\, L[\psi(X),\theta]$. The conclusion follows from taking the expected value of both sides for each given $\theta \in \Theta$. ||

When the estimators $\phi(X)$ and $\psi(X)$ are such that $\phi(X)$ performs well on the set of θ's for which $\psi(X)$ performs poorly, and vice versa, the combined estimator $\xi(X) = c\phi(X) + (1 - c)\, \psi(X)$ has a flatter risk function than both $\phi(X)$ and $\psi(X)$ (See Figure 3.4, for example.).

Since there will often be one value of c for which $r_{\xi}(\theta)$ is flattest, the choice of c should depend on the risk functions $r_{\phi}(\theta)$ and $r_{\psi}(\theta)$. It seems reasonable to weight the estimators according to their relative merits. A recursive procedure which does this and provides two new estimators, $\phi^{*}(X)$ and $\psi^{*}(X)$, which may be combined

again, is the following:

Define

$$\phi^*(x) = \{r_\phi[\psi(x)]\ \phi(x) + r_\phi[\phi(x)]\ \psi(x)\} / \{r_\phi[\phi(x)] + r_\phi[\psi(x)]\} ,$$

and

$$\psi^*(x) = \{r_\psi[\psi(x)]\ \phi(x) + r_\psi[\phi(x)]\ \psi(x)\} / \{r_\psi[\phi(x)] + r_\psi[\psi(x)]\} .$$

Let $\phi(X) = \phi^*(X)$ and $\psi(X) = \psi^*(X)$ and repeat until ϕ^* and ψ^* are identical or "close enough" to each other. Such a procedure may be quite time-consuming without the use of a computer, and there is no guarantee that it will provide an estimator with flatter risk. As an exception, consider the following simple example.

Example 2.5.1. Suppose $X \sim N(\theta,1)$ and θ is to be estimated subject to squared error loss. Consider the estimators $\psi(X) = X-1$ and $\phi(X) = X+1$ with risk functions $r_\psi(\theta) = r_\phi(\theta) = 2$.
The above technique gives

$$\phi^*(X) = \{2(X-1) + 2(X+1)\}/\ \{2 + 2\}\ = X\ ,\ \text{and}$$

$$\psi^*(X) = \{2(X-1) + 2(X+1)\}/\ \{2 + 2\}\ = X\ ,$$

so "convergence" occurs after one iteration. The resulting estimator X has smaller maximum risk than both $\phi(X)$ and $\psi(X)$.

CHAPTER III

Application of Techniques for Improving Estimators to the Poisson Estimation Problem when the Parameter Space is Truncated.

§3.1. Estimation of a Poisson Parameter Bounded Above

Let X be distributed as

$$f_\theta(x) = \theta^x e^{-\theta}/x! \quad , \quad x \in X = \{0,1,2,\ldots\} \ ,$$

for some θ known to be in the interval $\Theta^R = (0, B]$. The problem is to estimate θ subject to normalized squared error loss, $L(a,\theta) = \theta^{-1}(a-\theta)^2$. This problem arises when a change in a Poisson process causes the known rate B to decrease to some unknown value θ , which is to be estimated.

In $\mathcal{P}^U$, the same problem without the restriction placed on the parameter space, the estimator X is minimax (Theorem 1.2.4) and admissible (Theorem 1.2.5). It is also the uniformly minimum variance unbiased estimator, the method of moments estimator, and the maximum likelihood estimator. Thus we consider X to be optimal in $\mathcal{P}^U$.

When the restriction $\theta \le B$ is imposed on the parameter space, the estimator X will take on values outside $H = [0, B]$, the closed convex hull of Θ^R, and by Theorem 2.1.1 we can conclude that the projection

$$X^T = \begin{cases} X & \text{if} \quad X \le B \\ B & \text{if} \quad X > B \end{cases}$$

uniformly improves X. This projection is also the estimator provided by the method of maximum likelihood and performs extremely well for

large values of θ . (See Figure 3.1.)

We now try to improve X using the method of Section 2.2. Consider estimators of the form cX. In Example 2.2.2(a), we saw that

$$r_{cX}(\theta) = \theta^{-1}E_\theta[(cX - \theta)^2] = c^2\theta^{-1} E_\theta(X^2) - 2cE_\theta(X) + \theta .$$

As a function of c, $r_{cX}(\theta)$ is an upward-facing parabola, and is minimized (for each given θ) at its vertex

$$c_\theta = \theta E_\theta(X)/E_\theta(X^2) = \theta/(1+\theta) .$$

This best choice of c for each given θ is a strictly increasing function of θ, and therefore bounded above by $B/(1+B) < 1$. Applying Theorem 2.2 we conclude that the estimator

$$BX/(1+B)$$

uniformly improves X .

In this problem, we are searching for reasonable estimators, that is, estimators which improve X and are admissible. Are the estimators $BX/(1+B)$ and X^T, both of which improve X, admissible? It is clear that $BX/(1+B)$ is inadmissible, for it still takes on values outside $\mathbb{H}^R$. It is uniformly improved (using the method of Section 2.1) by the estimator

$$[BX/(B+1)]^T = \begin{cases} BX/(B+1) & X \le B + 1 \\ B & X > B + 1 \end{cases} .$$

We do not know whether X^T, the maximum likelihood estimator, is admissible. The results of Kale (1970) for the continuous exponential family problem, and Sackrowitz and Strawderman (1974) for the discrete problem of estimating ordered binomial parameters, seem to

suggest that this projection into $\mathcal{H}^R$ is not "smooth" enough to be (generalized) Bayes or admissible. For example, X^T estimates θ by the same value whether $X=B$ or $X=100B$.

It may be noticed that Sacks' (1963) result concerning the essential completeness of the class of generalized Bayes estimators does not apply in the Poisson estimation problem where the loss function is of the form $\gamma(\theta)\,(a-\theta)^2$, since his second condition on the loss function is violated. However, we might still expect that admissible estimators should behave like generalized Bayes estimators. And, as the results cited above seem to indicate, generalized Bayes estimators should be smooth, so we might expect that X^T will turn out to be inadmissible.

It may be observed that (generalized) Bayes estimators with respect to truncated Gamma and Beta priors project into $\mathcal{H}^R$ in a much smoother manner than X^T. Most of these, however, do not improve X. One which does is

$$\psi(X) = \int_0^B \theta^X e^{-\theta}\, d\theta \;/\; \int_0^B \theta^{X-1} e^{-\theta}\, d\theta \quad ,$$

the generalized Bayes estimator with respect to the rectangular prior on $[0, B]$. Although we have not proved that $\psi(X)$ improves X, this fact is apparent from Figure 3.1. In fact, $\psi(X)$ turns out to be admissible in this problem by a result of Katz (1961).

In view of the Normal example considered in Sections 2.3 and 2.4, we might expect $\psi(X)$ to be obtainable via a step-function approach. We will now use such an approach to recursively improve X. Although the estimator obtained is not $\psi(X)$, it is smooth and has other desirable properties.

Now, consider estimators of the form

$$d_k(X) = \begin{cases} cX & X < k \\ c_k X & X \geq k \end{cases} .$$

These will improve cX if the $c_{k,\theta}$'s (the best choices of c_k for each given $\theta \in \Theta^R$) are bounded away from c by c_k. Following the method of Section 2.4,

$$c_{k,\theta} = \theta E_\theta[X|X \geq k]/E_\theta[X^2|X \geq k]$$

$$= \theta \left[\sum_{x=k}^{\infty} x\, \theta^x\, e^{-\theta}/x!\right] \Big/ \left[\sum_{x=k}^{\infty} x^2\, \theta^x\, e^{-\theta}/x!\right]$$

$$= \theta^2 \left[\sum_{(k-1)^+}^{\infty} \theta^x/x!\right] \Big/ \left[\theta^2 \sum_{(k-2)^+}^{\infty} \theta^x/x! + \theta \sum_{(k-1)^+}^{\infty} \theta^x/x!\right] ,$$

where $(a)^+ = \max\{0,a\}$.

We now prove two useful lemmae concerning the behaviour of $c_{k,\theta}$.

Lemma 3.3.1. For each $k \in X$, $c_{k,\theta}$ is a strictly increasing function of θ .

Proof: $$c_{k,\theta}^{-1} = \begin{cases} (\theta+1)/\theta + [\theta^{k-2}/(k-2)!]/\sum_{x=k-1}^{\infty} \theta^x/x! & k = 2,3,\ldots \\ (\theta+1)/\theta & k = 0,1. \end{cases}$$

For $k = 2,3,\ldots$

$$c_{k,\theta}^{-1} = (\theta+1)/\theta + \left[(k-2)! \sum_{k-1}^{\infty} \theta^{x-k+2}/x!\right]^{-1}$$

$$= (\theta+1)/\theta + \left[(k-1)! \sum_{x=0}^{\infty} \theta^{x+1}/(x+k)!\right]^{-1}$$

and $\frac{d}{d\theta} c_{k,\theta}^{-1} = -\theta^{-2} - (k-1)! \left[\frac{d}{d\theta} \sum_{x=0}^{\infty} \theta^{x+1}/(x+k)!\right] \Big/ \left[(k-1)! \sum_{x=0}^{\infty} \theta^{x+1}/(x+k)!\right]^2$.

Now, $\sum_{x=0}^{\infty} \theta^{x+1}/(x+k)! < \sum_{x=0}^{\infty} \theta^{x+1}/(x+1)! < \sum_{x=0}^{\infty} \theta^{x}/x! = e^{\theta}$, $-\infty < \theta < \infty$,

so $\sum_{x=0}^{\infty} \theta^{x+1}/(x+k)!$ can be differentiated term by term on

$\Theta^R \subset (-\infty, \infty)$, and

$$\frac{d}{d\theta} c_{k,\theta}^{-1} = -\theta^{-2} - (k-1)! \left[\sum_{x=0}^{\infty} (x+1)\theta^{x}/(x+k)!\right] \Big/ \left[(k-1)! \sum_{x=0}^{\infty} \theta^{x+1}/(x+k)!\right]^2$$

$$< 0 , \quad \text{for all } \theta \in (0,B] .$$

For $k=0$ or 1, $\frac{d}{d\theta} c_{k,\theta}^{-1} = -\theta^{-2} < 0$, for all $\theta \in (0,B]$.

Thus $c_{k,\theta}^{-1}$ is strictly decreasing and $c_{k,\theta}$ is strictly increasing in θ . ‖

Lemma 3.1.2. For each fixed $\theta \in \Theta^R$, $c_{k,\theta}$ is strictly decreasing in k (for $k = 1, 2, \ldots$) .

Proof:

$$c_{k,\theta}^{-1} = (\theta+1)/\theta + \left[(k-1)! \sum_{x=0}^{\infty} \theta^{x+1}/(x+k)!\right]^{-1}$$

$$= (\theta+1)/\theta + \left[\frac{\theta}{k-1} + \frac{\theta^2}{(k-1)k} + \frac{\theta^3}{(k-1)k(k+1)} + \ldots\right]^{-1}$$

is strictly increasing in k and therefore, $c_{k,\theta}$ is strictly decreasing in k (for $k = 1,2,3,\ldots$) . ‖

Now, by Lemma 3.1.1, the $c_{k,\theta}$'s are bounded above by $c_{k,B} < 1$. We can thus conclude that

$$d_1(X) = \begin{cases} X & X < 1 \\ c_{1,B}X & X \geq 1 \end{cases} \quad \text{uniformly improves } X \text{ where}$$

$c_{1,B} = B/(B+1)$.

Also, $d_2(X) = \begin{cases} X & X < 1 \\ c_{1,B}X & X = 1 \\ c_{2,B}X & X \geq 2 \end{cases}$ uniformly improves $d_1(X)$, since

$$c_{2,\theta} \leq c_{2,B} < c_{1,B} \quad \text{for all } \theta \in \textcircled{H}.$$

We may continue in this manner, since $c_{k,\theta} \leq c_{k,B} < c_{k-1,B}$, for all $\theta \in \textcircled{H}$, and conclude that

$$d_k(X) = \begin{cases} X & X = 0 \\ c_{1,B}X & X = 1 \\ c_{2,B}X & X = 2 \\ \vdots & \\ c_{k-1,B}X & X=k-1 \\ c_{k,B}X & X \geq k \end{cases} \quad \text{uniformly improves } d_{k-1}(X) = \begin{cases} X & X = 0 \\ c_{1,B}X & X = 1 \\ c_{2,B}X & X = 2 \\ \vdots & \\ c_{k-2,B}X & X=k-2 \\ c_{k-1,B}X & X \geq k-1 \end{cases}$$

for each $k = 1,2,3,\ldots$. We would now like to conclude that $\phi(X)$, the limit of the sequence $\{d_k(k)\}_{k=1}^{\infty}$, uniformly improves each $d_k(X)$. This is a problem in interchange of limit and expection, and we will use the Bounded Convergence Theorem to show that this conclusion is, in fact, justified. Fatou's Lemma could be used instead, as in Section

2.4, but the Bounded Convergence Theorem yields a slightly stronger result.

Theorem 3.1.1. (Bounded Convergence Theorem): If $f_n(X) \xrightarrow{a.s.} f(X)$, and there exists a function $g(X)$ such that $P[|f_n(X)| \le g(X)] = 1$, and $E[g(X)] < \infty$, then $\lim_{n\to\infty} E[f_n(X)]$ exists and equals $E[\lim_{n\to\infty} f_n(X)]$.

Here $d_k(X) \xrightarrow{a.s.} \phi(X)$, and $[d_k(X)]^2 \xrightarrow{a.s.} [\phi(X)]^2$, and for all $x \in \mathcal{X}$, $d_k(x) \le x$, and $[d_k(x)]^2 \le x^2$. X and X^2 have finite expectations θ and $\theta+\theta^2$, respectively (for all $\theta \in \Theta^R$). Therefore, by the Bounded Convergence Theorem,

$$\lim_{k\to\infty} r_{d_k}(\theta) = \lim_{k\to\infty} E_\theta\{[d_k(X)]^2\}/\theta - 2 \lim_{k\to\infty} E_\theta[d_k(X)] + \theta$$

$$= E_\theta\{[\phi(X)]^2\}/\theta - 2 E_\theta[\phi(X)] + \theta = r_\phi(\theta) .$$

We have now proved the following theorem.

Theorem 3.1.2. The constructed estimator $\phi(X)$ uniformly improves each of the estimators $d_k(X)$ and thereby uniformly improves X as well.

By the relationship between the distribution function of a Poisson random variable and the incomplete gamma integral (see Appendix), we may write $\phi(X)$ as

$$\phi(x) = c_{x,B}x = xB[1-F_B(x-2)]/[B(1-F_B(x-3)) + 1-F_B(x-2)]$$

$$= \frac{\left[xB \int_0^B \theta^{x-2} e^{-\theta} d\theta\right] / (x-2)!}{\left[B \int_0^B \theta^{x-3} e^{-\theta} d\theta\right] /(x-3)! + \left[\int_0^B \theta^{x-2} e^{-\theta} d\theta\right] /(x-2)!}$$

We have been unable to show that $\phi(X)$ is generalized Bayes or admissible, although its form resembles that of the (generalized) Bayes estimators with respect to a prior of the form $\Pi(\theta) = \theta^{\alpha}$, for then

$$d_{\Pi}(X) = \left[\int_0^B \theta^{x+\alpha-1} e^{-\theta} d\theta\right] / \left[\int_0^B \theta^{x+ -2} e^{-\theta} d\theta\right] .$$

However, $\phi(X)$ is "smooth", and as the following theorem shows, $\phi(X)$ takes on all its values inside $[0,B)$, so it is still possible that $\phi(X)$ is generalized Bayes and possibly a candidate for admissibility.

Theorem 3.1.3. For each $x \in X = \{0,1,2,...\}$, $0 \le \phi(x) < B$.

Proof:

$$\phi(x) = xc_{x,B} = xBE_B[X|X \ge x]/E_B(X^2|X \ge x] \ge 0 \text{ , for all } x \in X .$$

Since $Var_B[X|X \ge x] = E_B[X^2|X \ge x] - \{E_B[X|X \ge x]\}^2 > 0$,

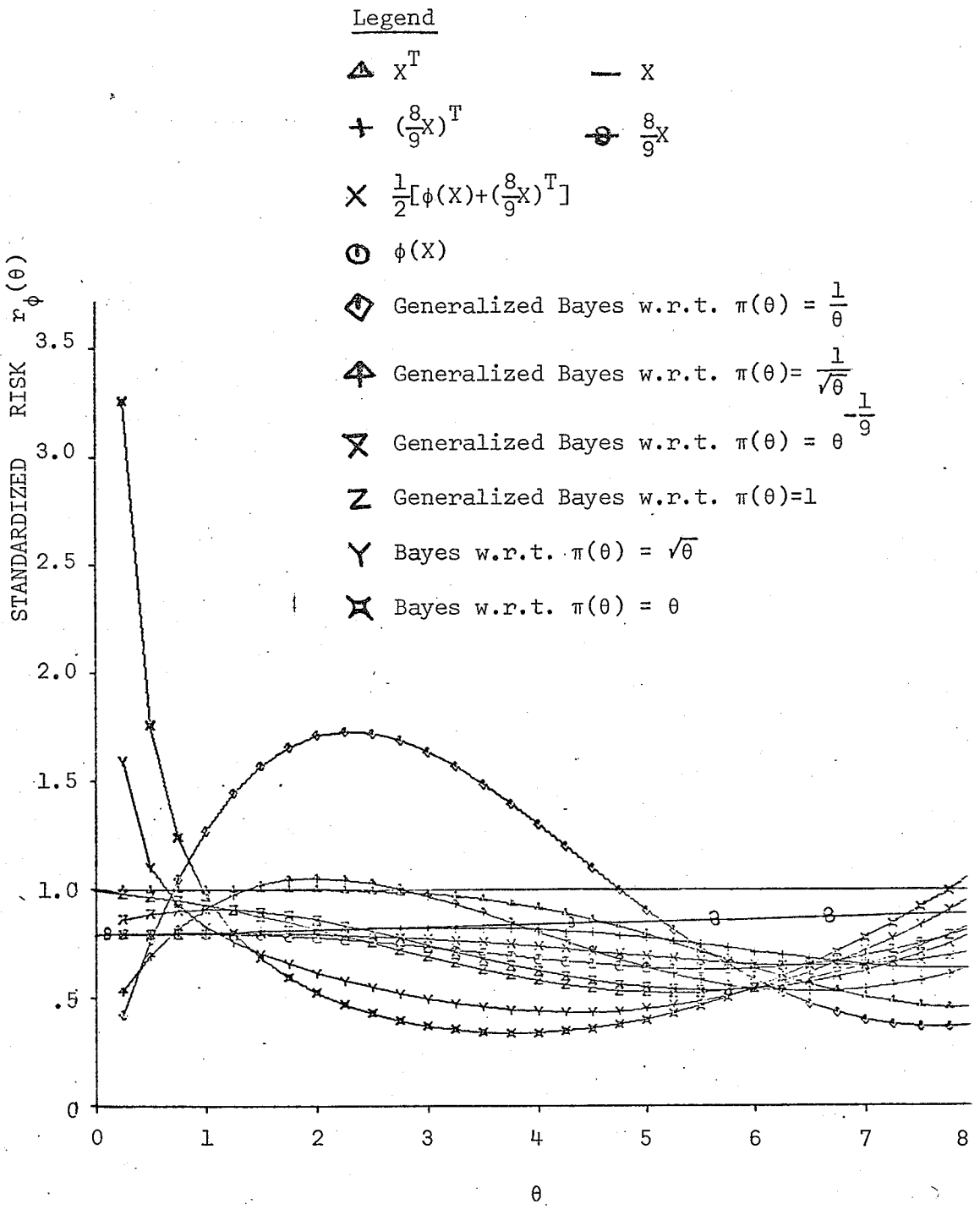
$$E_B[X^2|X \ge x] > \{E_B[X|X \ge x]\}^2$$

$$> xE_B[X|X \ge x] ,$$

so $\phi(x) < B$, for all $x \in X$. ‖

For the problem in which B is known to be 8, various estimators were programmed and their risk functions evaluated and plotted using the IBM 370/158 computer and its Calcomp plotter at the

Figure 3.1 Risk functions of various estimators of a Poisson parameter θ bounded above by B=8.



University of Manitoba. The estimators considered were

$$X\ ,\ \frac{8}{9}X\ ,\ X^T = \begin{cases} X & X \le 8 \\ 8 & X > 8 \end{cases}\ ,\quad \left(\frac{8}{9}X\right)^T = \begin{cases} 8X/9 & X \le 9 \\ 8 & X > 9 \end{cases}\ ,\quad \phi(X)\ ,$$

and (generalized) Bayes estimators with respect to priors of the forms $\theta^{\alpha-1}$, $\theta^{\alpha-1} e^{-\theta/\beta}$, $\theta^{\alpha-1} (1-\theta)^{\beta-1}$, for various (α,β)'s . The (generalized) Bayes estimators were calculated using a Gaussian quadrature technique (which will work exactly for polynomials of degree 63 or less) to evaluate the incomplete integrals. Plots of the risk functions of these estimators are presented here to illustrate the results of this section.

From Figure 3.1, we see that the estimator provided by the method of Section 2.2, $8X/9$, substantially improves X at each $\theta \in (0,8]$. The projection provided by the method of Section 2.1, X^T (which is also the maximum likelihood estimator), improves X only slightly for $\theta \le 3$, but performs extremely well for larger θ:

$$r_{X^T}(\theta) \le \tfrac{1}{2}\, r_X(\theta) \quad \text{for} \quad \theta > 7\ .$$

A similar situation may be observed in the relationship between $8X/9$ and its projection $(8X/9)^T$.

The estimator $\phi(X)$, provided by the step-function approach of Section 2.4, also substantially improves X. In fact, we knew that $\phi(X)$ would even improve $8X/9$, as $8X/9$ was the first step in the construction of $\phi(X)$, and this improvement is significant over a wide range of parameter values.

Furthermore, the risk function of $\phi(X)$ is relatively flat and

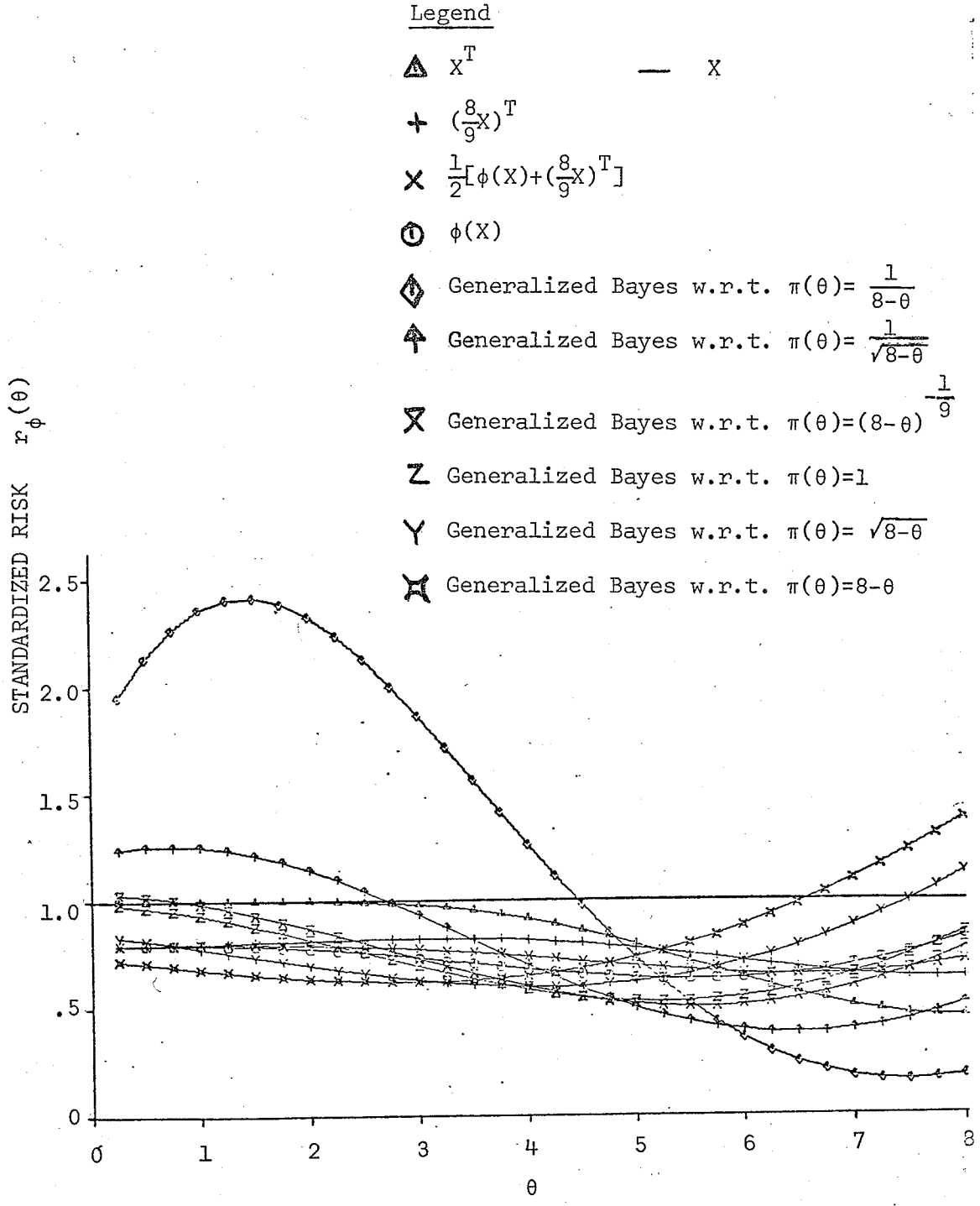
is maximized at $\theta=8$. On the other hand, the risk function of X^T is substantially smaller than that of $\phi(X)$ at $\theta=8$, and is maximized as θ approaches 0 . Since both estimators uniformly improve X , it follows from Theorem 2.5.1 that any convex combination $\alpha\phi(X) + (1-\alpha) X^T$ will uniformly improve X , and there will exist a range of α's which yield smaller maximum risk than $\phi(X)$. From Figure 3.3, this range appears to be $0 < \alpha < .2$, and the choice of α which minimizes the maximum risk is approximately .05 .

The iterative technique of Section 2.5 is illustrated for $\phi(X)$ and $(8X/9)^T$ in Figure 3.4 . It may be noticed that the frist iteration has smaller maximum risk than the estimators to which $\phi(X)$ and $(8X/9)^T$ converge. Thus the iterative technique does not converge to an estimator which is minimax, although it does flatten out the risk to some extent.

Of the (generalized) Bayes estimators with respect to $\pi(\theta) = \theta^{\alpha-1}$, it may be observed from Figure 3.1 that they improve X only for values of α larger than about .6 and smaller than or equal to 1. Since these estimators must have $\alpha > 1$ to be Bayes, there are no Bayes estimators with respect to prior densities of the form $\theta^{\alpha-1}$ which improve X . From Figure 3.2 we see that generalized Bayes estimators with respect to priors of the form $(8-\theta)^{\beta-1}$ improve X only for $1 \leq \beta \leq 1.3$ (approximately).

To conclude, some estimators which improve X uniformly in this problem are X^T, $\phi(X)$, and $\psi(X)$, the generalized Bayes estimator with respect to constant prior density on $(0,B]$. $\psi(X)$ is known to be admissible by Katz' (1961) result, and X^T is easy to calculate and has smaller risk for large θ . However, $\phi(X)$ has the flattest risk

Figure 3.2 Risk functions of various estimators of a Poisson parameter θ bounded above by B=8



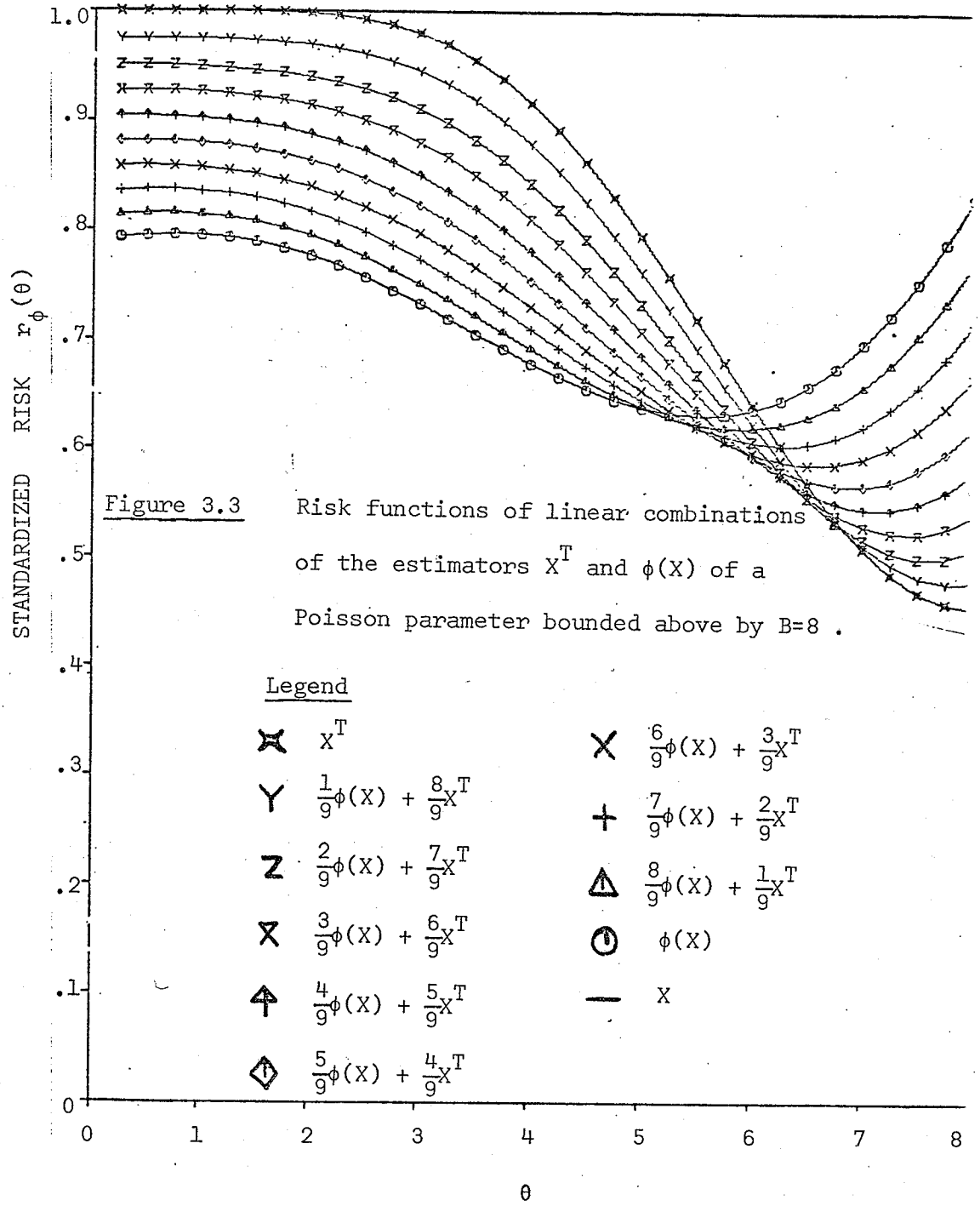


Figure 3.3 Risk functions of linear combinations of the estimators X^T and $\phi(X)$ of a Poisson parameter bounded above by B=8 .

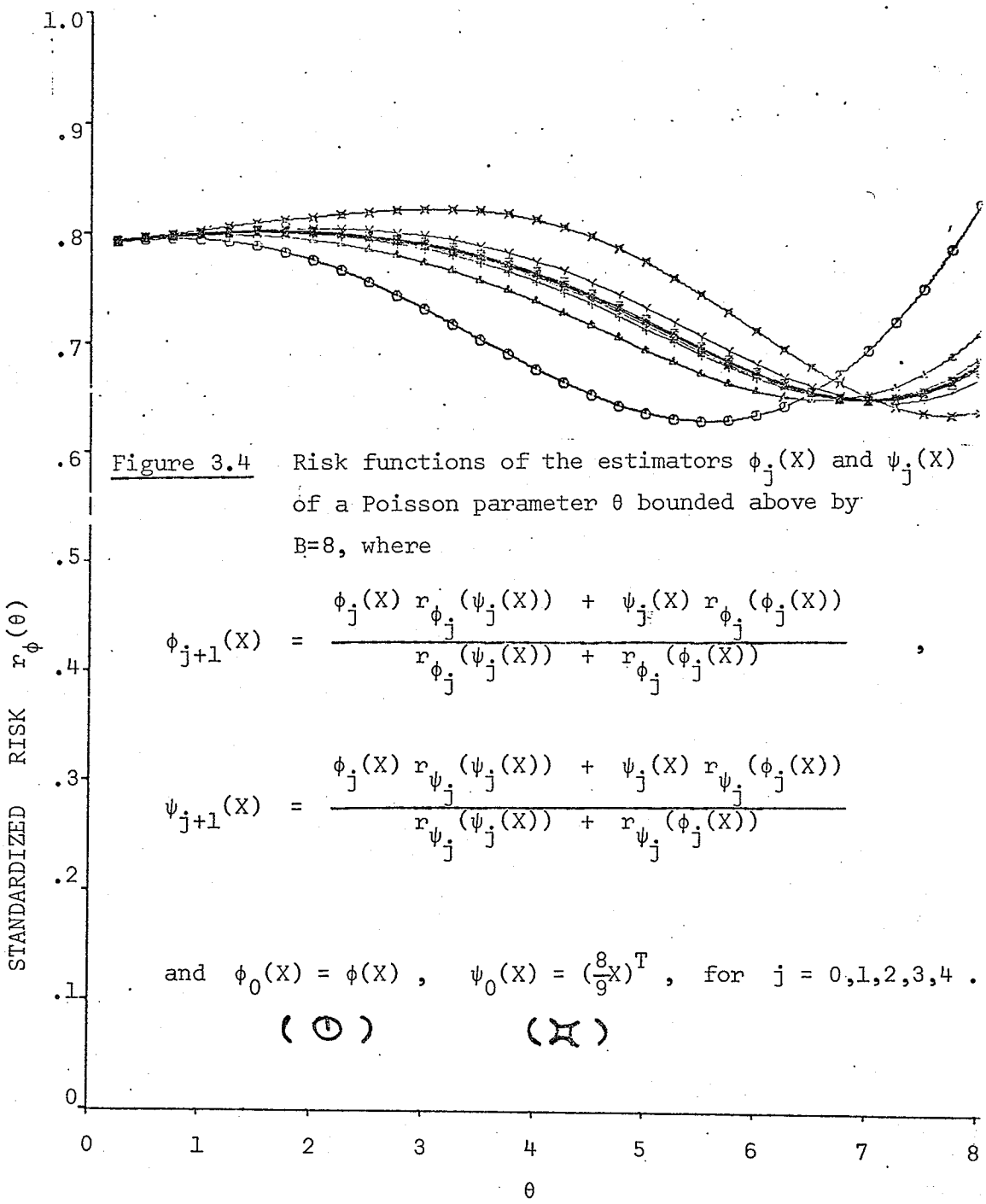


Figure 3.4 Risk functions of the estimators $\phi_j(X)$ and $\psi_j(X)$ of a Poisson parameter θ bounded above by B=8, where

$$\phi_{j+1}(X) = \frac{\phi_j(X)\, r_{\phi_j}(\psi_j(X)) + \psi_j(X)\, r_{\phi_j}(\phi_j(X))}{r_{\phi_j}(\psi_j(X)) + r_{\phi_j}(\phi_j(X))} ,$$

$$\psi_{j+1}(X) = \frac{\phi_j(X)\, r_{\psi_j}(\psi_j(X)) + \psi_j(X)\, r_{\psi_j}(\phi_j(X))}{r_{\psi_j}(\psi_j(X)) + r_{\psi_j}(\phi_j(X))}$$

and $\phi_0(X) = \phi(X)$, $\psi_0(X) = (\frac{8}{9}X)^T$, for $j = 0,1,2,3,4$.

(⊙) (⋈)

and has much smaller risk than X^T for $\theta < 6$ (see Figure 3.3). In a situation where θ is known to be smaller than 8, prior feeling for θ may concentrate more on intermediate values of θ than on values only slightly smaller than 8. In such cases there is further justification for using an estimator such as $\phi(X)$, which projects smoothly into $\mathbb{H}^R$ and improves X. And, it may be that $\phi(X)$ is admissible, although we have not been able to prove this result.

§3.2. Estimation of a Poisson Parameter Bounded Below.

Consider the problem of the previous section with the parameter space restricted instead to $\mathbb{H}^R = [B, \infty)$. Again, the "optimal" estimator in $\mathcal{P}^U$, X, takes on values outside $\mathbb{H}^R$ with positive probability for each $\theta \in \mathbb{H}^R$, so by Theorem 2.1.1 it is uniformly improved by

$$X^T = \begin{cases} X & X \geq B \\ B & X < B \end{cases},$$

which is also the maximum likelihood estimator for $\mathcal{P}^R$. However, the method of Section 2.2 fails here if used directly, for $c_\theta = \theta/(1+\theta)$ cannot be bounded away from 1 uniformly on $\mathbb{H}^R$. Let us attempt a step-function approach analogous to the one used in the previous section. The choice of c which minimizes the risk of the estimator

$$d_k(X) = \begin{cases} cX & X \leq k \\ X & X > k \end{cases},$$

when the parameter is θ, is

$$c_{k,\theta} = \theta E_\theta[X|X \leq k]/E_\theta[X^2|X \leq k] .$$

$$= \theta F_\theta(k-1)/[F_\theta(k-1) + \theta F_\theta(k-2)] \ .$$

Therefore $c_{k,\theta}^{-1} = (\theta+1)/\theta - [1 + \frac{k-1}{\theta} + \frac{(k-1)(k-2)}{\theta^2} + \frac{(k-1)!}{\theta^{k-1}}]^{-1}$.

The series decreases in θ for fixed k, and increases in k for fixed θ, and the left term is a decreasing function of θ, so $c_{k,\theta}$ increases in θ for fixed k, and decreases in k for fixed θ.

Consider

$$d_1(X) = \begin{cases} c_{1,B} & X \le 1 \\ X & X > 1 \end{cases} .$$

The $c_{1,\theta}$'s are bounded below by

$$c_{1,B} = B\, E_\theta[X|X = 0 \text{ or } 1]/E_\theta[X^2|X = 0 \text{ or } 1] = B \ .$$

$d_1(X)$ will improve X uniformly if and only if $B > 1$, so we must assume $B > 1$ to continue. Then the $c_{1,\theta}$'s are bounded away from 1 by B, and

$$\phi_1(X) = \begin{cases} B & X \le 1 \\ X & X > 1 \end{cases} \quad \text{uniformly improves } X \ .$$

Now consider

$$d_2(X) = \begin{cases} c_{2,B}X & X \le 2 \\ X & X > 2 \end{cases} \ . \quad \text{If, and only if,}$$

$c_{2,B} > 1$, $d_2(X)$ uniformly improves X and may itself be improved by

$$\phi_2(X) = \begin{cases} B & X = 0 \\ BX & X = 1 \\ c_{2,B}X & X = 2 \\ X & X > 2 \end{cases} \ . \quad \text{Unfortunately, this}$$

process may not be repeated indefinitely, for $\lim_{k\to\infty} c_{k,B} = B/(B+1)$.

Let $k^* = \max\{k | c_{k,B} > 1\}$. Then we may repeat the process until we obtain the estimator

$$\phi_{k^*}(X) = \begin{cases} B & X = 0 \\ BX & X = 1 \\ c_{2,B}X & X = 2 \\ \vdots & \\ c_{k^*}(X) & X = k^* \\ X & X > k^* \end{cases} ,$$

which uniformly improves X. However we may go no further since the $c_{k^*+1,\theta}$'s are not bounded away from 1 by $c_{k^*+1,B}$. It is not known whether the risk functions of $\phi_1(X), \phi_2(X), \ldots, \phi_{k^*}(X)$ are ordered, but $\phi_{k^*}(X)$ is the "smoothest" of these estimators.

As an alternative approach, consider estimators of the form

$$\begin{cases} X + a_k & X < k \\ X & X \geq k \end{cases} .$$

The best choice of a_k for each given θ is found by minimizing

$$r_a(\theta) = E_\theta[(X+a-\theta)^2 | X < k] P_\theta[X < k]/\theta + E_\theta[(X-\theta)^2 | X \geq k] P_\theta[X \geq k]/\theta ,$$

and is $a_{k,\theta} = \theta - E_\theta[X|X < k]$, since $E[(X-m)^2]$ is minimized for $m = E(X)$.

$$\text{Now } a_{k,\theta} = \theta - \left[\sum_{x=0}^{k-1} x\,\theta^x/x!\right] \Big/ \left[\sum_{x=0}^{k-1} \theta^x/x!\right]$$

$$= \theta - \left[\theta \sum_{x=1}^{k-1} \theta^{x-1}/(x-1)!\right] \Big/ \left[\sum_{x=0}^{k-1} \theta^x/x!\right]$$

$$= \theta\left[\sum_{x=0}^{k-1} \theta^x/x! - \sum_{x=0}^{k-2} \theta^x/x!\right] \Big/ \left[\sum_{x=0}^{k-1} \theta^x/x!\right]$$

$$= \left[\theta^k\, e^{-\theta}/(k-1)!\right] \Big/ \left[\sum_{x=0}^{k-1} \theta^x\, e^{-\theta}/x!\right]$$

$$= \theta f_\theta(k-1)/F_\theta(k-1) > 0 \text{ , for all } k \in X,\ \theta \in [B,\infty) .$$

$$= \left[\frac{1}{\theta} + \frac{(k-1)}{\theta^2} + \frac{(k-1)(k-2)}{\theta^3} + \ldots + \frac{(k-1)!}{\theta^k}\right]^{-1} ,$$

and thus $a_{k,\theta}$ increases in θ for fixed k, and decreases in k for fixed θ. Thus, for each fixed k, $a_{k,\theta}$ is bounded below by $a_{k,B} > 0$, so Theorem 2.2.1 may be applied to conclude that

$$d_{kl}(X) = \begin{cases} X + a_{k,B} & X < k \\ X & X \geq k \end{cases}$$

uniformly improves X. Now consider estimators of the form

$$\begin{cases} X + a_{k-1} & X < k-1 \\ X + a_{k,B} & X = k-1 \\ X & X \geq k \end{cases} .$$

The $a_{k-1,\theta}$'s are bounded below by $a_{k-1,B} > a_{k,B}$, and thus

Theorem 2.2.1 may be applied to conclude that

$$d_{k2}(X) = \begin{cases} X + a_{k-1,B} & X < k-1 \\ X + a_{k,B} & X = k-1 \\ X & X \geq k \end{cases} \quad \text{uniformly improves } d_{k1}(X).$$

This process may be repeated k times until we obtain the estimator

$$\phi_k(X) = d_{kk}(X) = \begin{cases} X + a_{1,B} & X = 0 \\ X + a_{2,B} & X = 1 \\ \vdots & \\ X + a_{k,B} & X = k-1 \\ X & X \geq k \end{cases},$$

which takes on all its values in $[B, \infty)$, and which uniformly improves $d_{kj}(X)$ for each $j = 1,2,\ldots,k-1$, and thus uniformly improves X. But this holds for each $k \in \mathcal{X}$, so each of the estimators $\phi_1(X), \phi_2(X),\ldots,$ uniformly improves X though they are not necessarily ordered. Consider the limit of this sequence, $\phi(X) = \lim_{k\to\infty} \phi_k(X) = X+a_{X+1,B}$, which is smooth. Now, for each $\theta \in [B, \infty)$, we have that

$$[\phi_k(x) - \theta]^2/\theta \geq 0, \quad \text{for all } k \in \mathcal{X}, \ x \in \mathcal{X}, \text{ and an}$$

application of Fatou's Lemma yields

$$r_X(\theta) \geq \liminf_{k\to\infty} r_{\phi_k}(\theta) = \liminf_{k\to\infty} \int_{\mathcal{X}} [\phi_k(x) - \theta]^2/\theta \ dF_\theta(x)$$

$$\geq \int_{\mathcal{X}} \liminf_{k\to\infty} [\phi_k(x) - \theta]^2/\theta \ dF_\theta(x)$$

$$= \int_{\mathcal{X}} [\phi(x) - \theta]^2/\theta \ dF_\theta(x) = r_\phi(\theta).$$

We have thus proved the following theorem.

Theorem 3.2.1. The constructed estimator $\phi(X) = X+a_{X+1,B}$ is as good as X .

It is of interest to note that the estimator which adds $a_{X,B}$ to X instead of $a_{X+1,B}$ may be written as

$$\psi(X) = X+a_{X,B} = X + [B^X e^{-B}/(X-1)!]/[\sum_{j=0}^{X-1} B^j e^{-B}/j!]$$

$$= X + [B^X e^{-B}/(X-1)!] \,/\, [\int_B^\infty \theta^{X-1} e^{-\theta}\, d\theta/(X-1)!] ,$$

by the relationship between the Poisson distribution function and the incomplete Gamma integral (see Appendix) .

Thus $\psi(X) = X + B^X e^{-B} / \int_B^\infty \theta^{X-1} e^{-\theta}\, d\theta$ is the estimator shown by Katz (1961) to be an admissible limit of Bayes estimators. Katz points out that all estimators in this problem are minimax (subject to squared error loss). We are interested in minimaxity subject to normalized squared error loss. The estimator $\psi(X)$ may also be written as

$$\psi(X) = X + [B^X e^{-B}/(X-1)!]/[\sum_{j=0}^{X-1} B^j e^{-B}/j!]$$

$$= X\,[\sum_{j=0}^{X} B^j e^{-B}/j!]/[\sum_{j=0}^{X-1} B^j e^{-B}/j!]$$

$$= X\,[\int_0^\infty \theta^X e^{-\theta}\, d\theta/X!]/[\int_B^\infty \theta^{X-1} e^{-\theta}\, d\theta/(X-1)!]$$

$$= \int_0^\infty \theta^X e^{-\theta}\, d\theta / \int_B^\infty \theta^{X-1} e^{-\theta}\, d\theta \text{ , the generalized Bayes}$$

estimator with respect to Lebesgue prior measure on $[B, \infty)$. Since X is an equalizer estimator, we can conclude from the following theorem that, in this problem, an estimator is minimax if and only if it is as good as X.

Theorem 3.2.2. In this problem, although we have shown that X is inadmissible, X is minimax.

Proof: If X is not minimax, then there exists an estimator $d(X)$ such that $r_d(\theta) \leq \sup_{\theta \in \Theta} r_X(\theta) - \epsilon = 1 - \epsilon$, for some $0 < \epsilon < 1$.

But $r_d(\theta) = \theta^{-1}[b_d^2(\theta) + \sigma_d^2(\theta)]$

$$\geq \theta^{-1}\{b_d^2(\theta) + [1+b'(\theta)]^2 / E_\theta[(\frac{d}{d\theta} \log f_\theta(X))^2]\} ,$$

by the Cramér-Rao inequality (the regularity conditions were shown to be satisfied in Section 1.2), and since $E_\theta[(\frac{d}{d\theta} \log f_\theta(X))^2] = \theta^{-1}$, we have that

(3.2.1) $b^2(\theta)/\theta + [1+b'(\theta)]^2 \leq 1 - \epsilon$, for $\theta \in [B, \infty)$.

Thus $1+b'(\theta) \leq |1+b'(\theta)| \leq \sqrt{1 - \epsilon - b^2(\theta)/\theta}$,

and so $b'(\theta) \leq -c$, for all $\theta \geq B$, where $c = 1 - \sqrt{1 - \epsilon} > 0$.

But Equation (3.2.1) implies that $b^2(\theta)/\theta \leq 1 - \epsilon$, so $b(\theta)$ is absolutely dominated as

$$|b(\theta)| \leq \sqrt{\theta(1 - \epsilon)} ,$$

and thus $b'(\theta) \geq -\frac{1}{2}\sqrt{(1-\epsilon)/\theta}$, for all $\theta > B$. Therefore, $-\frac{1}{2}\sqrt{(1-\epsilon)/\theta} \leq b'(\theta) \leq -c$, for all $\theta > B$. But this is impossible, since, for sufficiently large (finite) θ , the lower bound becomes greater than the upper bound. Therefore X is minimax. ||

Now, since $\phi(X)$ has been constructed to be as good as X , $\phi(X)$ must also be minimax. Figures 3.5 and 3.6 illustrate the risk functions of $X, X^T, \phi(X)$ and generalized Bayes estimators with respect to prior densities of the forms $\theta^{\alpha-1}$, $(\theta-B)^{\alpha-1}$, for the problem of this section with $B=1$. Although from our construction of $\phi(X)$ we may conclude only that $\phi(X)$ is as good as X , we see in Figures 3.5 and 3.6 that (at least for $B=1$) $\phi(X)$ substantially improves X. Furthermore, $\psi(X)$, the estimator shown by Katz to be admissible, does not improve X and is not minimax, and neither do the generalized Bayes estimators of the forms examined. Finally, although we have not been able to show that $\phi(X)$ is admissible (or even generalized Bayes), we may prefer its use to that of X^T on the grounds that $\phi(X)$ is smooth, while X^T is not, and except for $\theta < 1.6$, $r_\phi(\theta)$ is substantially smaller than $r_{X^T}(\theta)$.

Figure 3.5 Risk functions of various estimators of a Poisson parameter θ bounded below by B=1.

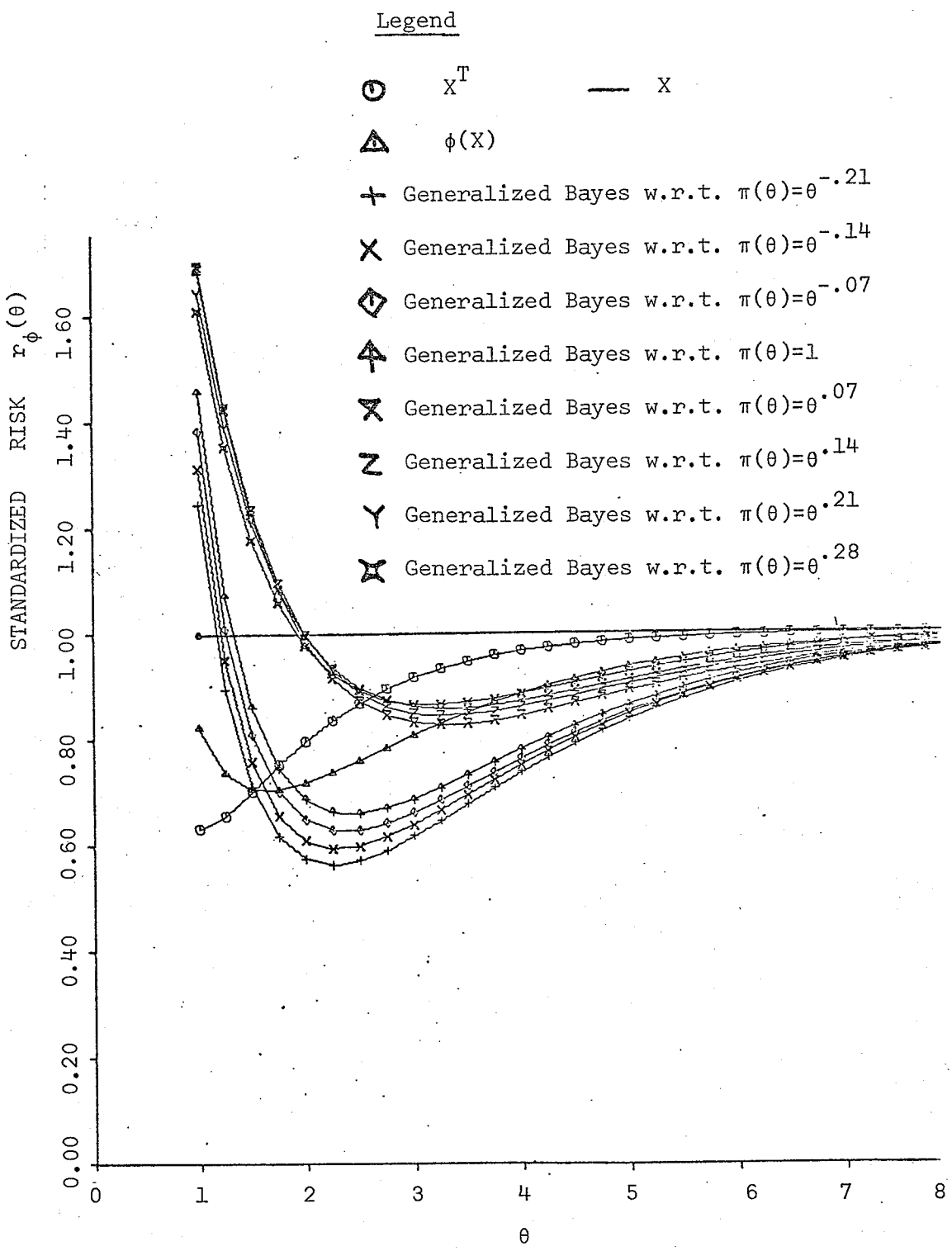
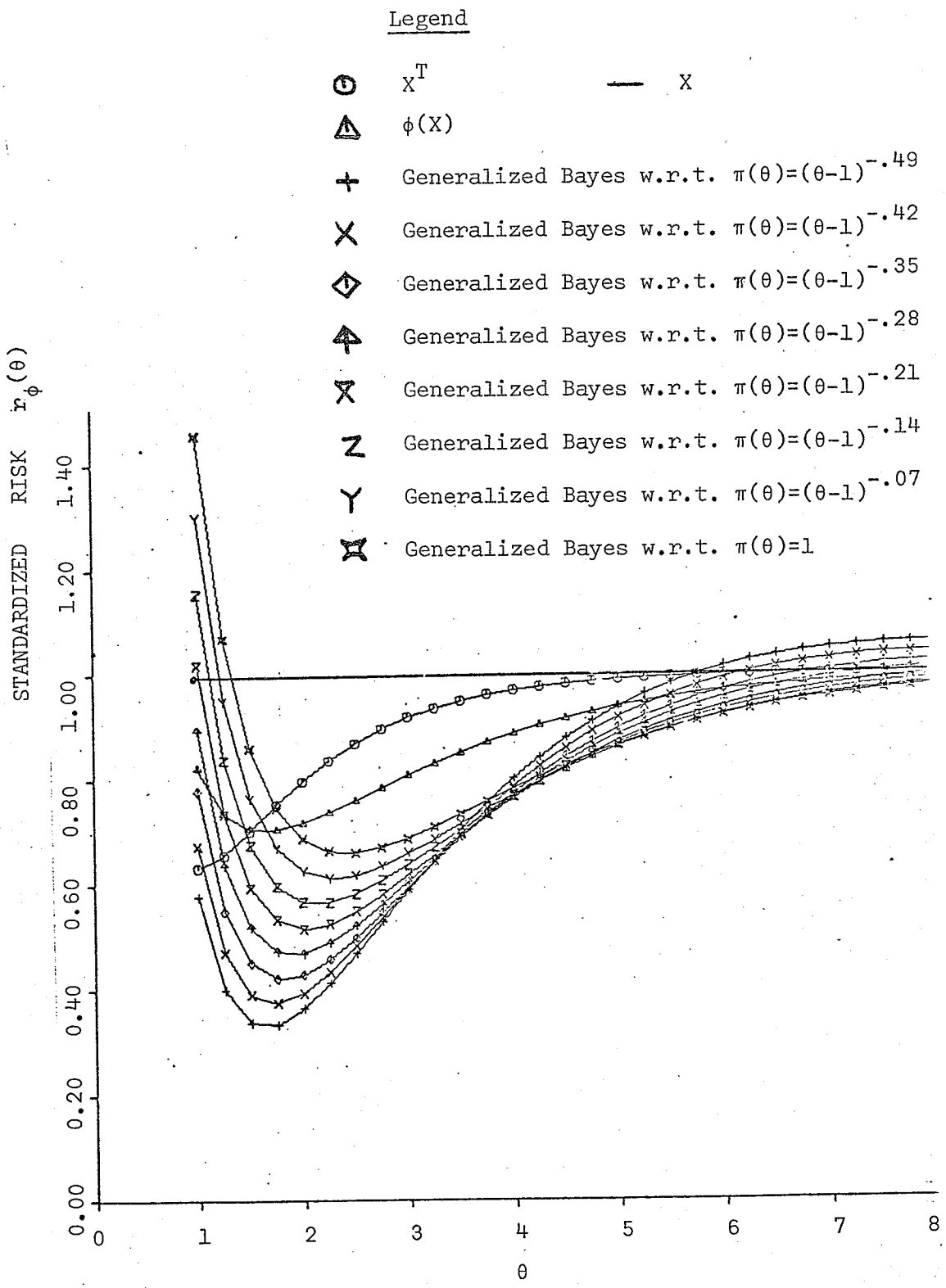


Figure 3.6 Risk functions of various estimators of a Poisson parameter θ bounded below by B=1 .



CHAPTER IV

Remarks on the Problem of Estimating Ordered Binomial Propensities

In Chapter III, we applied the techniques of Chapter II to the problem of estimating a single Poisson parameter bounded above or below. We now comment on the application of one of these techniques to the more complex problem of estimating ordered binomial propensities.

It is often desired to simultaneously estimate the propensities

$$\underline{p} = (p_1, p_2, \ldots, p_m) \in \textcircled{H} = [0, 1]^m ,$$

to succeed, for each of a number, m, of independent binomial experiments respectively based on $n_1, n_2, \ldots, n_m$ trials, on which $\underline{X} = (X_1, X_2, \ldots, X_m)$ successes are observed. In this situation, $\phi^U(\underline{X}) = \underline{f} = (f_1, f_2, \ldots, f_m)$, the generalization to m dimensions of the sample proportion of successes, $f_i = X_i/n_i$, is an intuitively appealing estimator with some desirable properties.

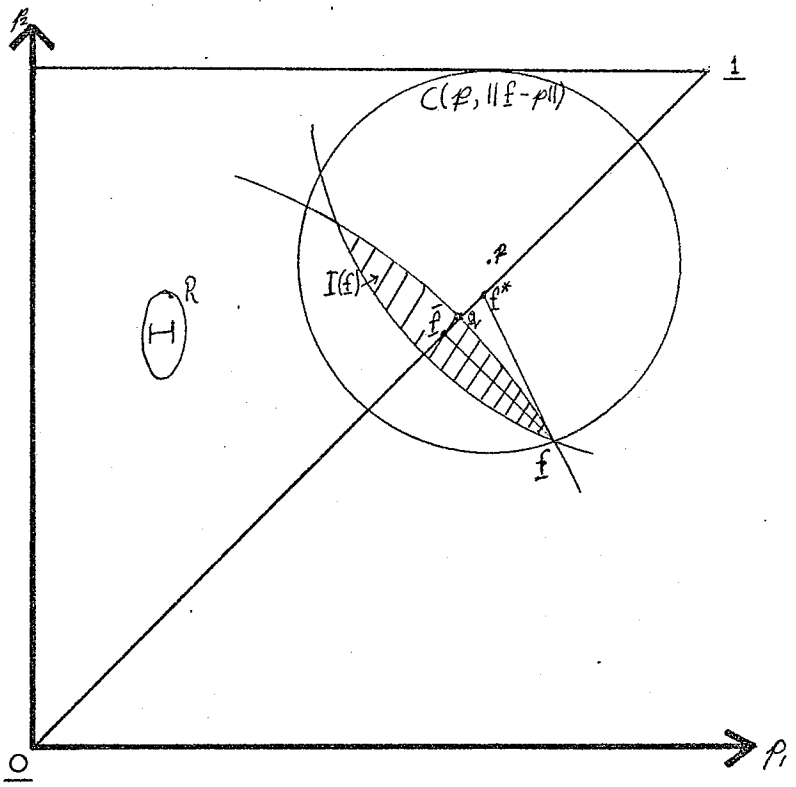
In a bioassay problem, it may be desired to estimate propensities to respond for a given set of levels of stimulus, where it is known that propensity to respond is an increasing (decreasing) function of the level of the stimulus. Similarly, the propensity for a manufacturing process to produce defective items increases as its components deteriorate or decreases as successive improvements are made in the process, and it may be desired to simultaneously estimate the propensities to produce defective items at each of m known levels of deterioration, or after each of m successive improvements are made.

For these examples, the above formulation is not quite adequate,

for it does not take into account a set of order restrictions on the parameters: $p_{i_1} \leq p_{i_2} \leq \ldots \leq p_{i_m}$ where $(i_1,i_2,\ldots,i_m)$ is a permutation of $(1,2,\ldots,m)$. By relabelling the trials, this restriction can be expressed without loss of generality as $p_1 \leq p_2 \leq \ldots \leq p_m$, and the restricted parameter space denoted as $\Theta^R = \{p_1,\ldots,p_m | 0 \leq p_1 \leq p_2 \leq \ldots \leq p_m \leq 1\}$. Though still the uniformly minimum-variance unbiased estimator, ϕ^U is no longer the maximum-likelihood estimator in $\mathcal{P}^R$, nor does it have the properties which made its use appealing in $\mathcal{P}^U$. For example, it may now take on values outside Θ^R. Since Θ^R is closed and convex, it is its own convex hull, so we can apply Theorem 2.1.1 and conclude that there is a projection of ϕ^U onto the boundary of Θ^R which uniformly improves ϕ^U. This is hard to visualize, and so we will deal with the simplified case of estimation of two ordered binomial propensities:

$X_1 \sim B(n_1,p_1)$, $X_2 \sim B(n_2,p_2)$, $0 \leq p_1 \leq p_2 \leq 1$. The problem is to estimate $\underline{p} = (p_1,p_2)$ subject to a loss function $L(\underline{a},\underline{p})$ which is an increasing function of $||\underline{a} - \underline{p}||^2 = (a_1 - p_1)^2 + (a_2 - p_2)^2$.

Now let $C(\underline{p}, ||\underline{f} - \underline{p}||)$ denote the interior of the circle with centre $\underline{p} \in \mathbb{H}^R$ and radius $||\underline{f} - \underline{p}||$.



Theorem 4.1.1. Let $I(\underline{f}) = C(\underline{0}, ||\underline{f} - \underline{0}||) \cap C(\underline{1}, ||\underline{f} - \underline{1}||)$, and let $\underline{f}$ be fixed, and outside $\mathbb{H}^R$. Then, for every $\underline{p} \in \mathbb{H}^R$, and for every $\underline{a} \in \mathbb{H}^R$, $L(\underline{a},\underline{p}) < L(\underline{f},\underline{p})$.

Proof: Since $C(\underline{p}, ||\underline{f} - \underline{p}||) = \{\underline{a} \mid ||\underline{a} - \underline{p}|| < ||\underline{f} - \underline{p}||\}$, we have $\bigcap_{\underline{p} \in \mathbb{H}^R} C(\underline{p}, ||\underline{f} - \underline{p}||) = \{\underline{a} \mid ||\underline{a} - \underline{p}|| < ||\underline{f} - \underline{p}||$ for all $\underline{p} \in \mathbb{H}^R\}$.

Clearly, $\bigcap_{\underline{p} \in \mathbb{H}^R} C(\underline{p}, ||\underline{f} - \underline{p}||) \subset I(\underline{f})$. Now, let $\underline{a} \in I(\underline{f})$.

The perpendicular bisector of the line $\overline{\underline{a}\ \underline{f}}$, which by construction lies outside $\mathbb{H}^R$, separates those points which are closer to $\underline{a}$ from those which are closer to $\underline{f}$. Thus there exists no $\underline{p} \in \mathbb{H}^R$ such that

$||\underline{a} - \underline{p}|| \geq ||\underline{f} - \underline{p}||$. That is, for every $\underline{p} \in \mathbb{H}^R$, $C(\underline{p}, ||\underline{f} - \underline{p}||)$ contains $\underline{a}$, so we have shown that

$$I(\underline{f}) = \bigcap_{\underline{p} \in \mathbb{H}^R} C(\underline{p}, ||\underline{f} - \underline{p}||) = \{\underline{a} \mid ||\underline{a} - \underline{p}|| < ||\underline{f} - \underline{p}||, \text{ for all } \underline{p} \in \mathbb{H}^R\} .$$

Since $L(\underline{a},\underline{p}) = ||\underline{a} - \underline{p}||^2$ is a strictly increasing function of $||\underline{a} - \underline{p}||$,

$$I(\underline{f}) = \{\underline{a} \mid L(\underline{a},\underline{p}) < L(\underline{f},\underline{p}) \quad \text{for all } \underline{p} \in \mathbb{H}^R\} . \qquad \|$$

Theorem 4.1.2. The estimator provided by orthogonally projecting the values which $\phi^U(X_1,X_2) = \underline{f}$ takes on outside $\mathbb{H}^R$ onto the boundary of $\mathbb{H}^R$,

$$\phi^R(X_1,X_2) = \begin{cases} \overline{\underline{f}} = (\tfrac{1}{2}[f_1 + f_2] , \tfrac{1}{2}[f_1 + f_2]) & f_1 > f_2 \\ \underline{f} = (f_1, f_2) & f_1 \leq f_2 , \end{cases}$$

uniformly improves ϕ^U .

Proof: As in Theorem 2.1.1, the result is obtained directly from Theorem 4.1.1 by taking expectations. We need only verify that $\overline{\underline{f}} \in I(\underline{f})$, for any $\underline{f} \notin \mathbb{H}^R$.

But, for any $\underline{f} \notin \mathbb{H}^R$, the points $\underline{0}$, $\underline{f}$, $\overline{\underline{f}}$ are the vertices of a right triangle with hypotenuse $\overline{\underline{0}\,\underline{f}}$ (see Figure 4.1) . Therefore $||\overline{\underline{f}} - \underline{0}|| < ||\underline{f} - \underline{0}||$, and $\overline{\underline{f}} \in C(\underline{0}, ||\underline{f} - \underline{0}||)$.

Similarly, $\underline{1}$, $\underline{f}$, $\overline{\underline{f}}$ are the vertices of a right triangle with hypotenuse $\overline{\underline{1}\,\underline{f}}$, so $\overline{\underline{f}} \in C(\underline{1}, ||\underline{f} - \underline{1}||)$. Thus $\overline{\underline{f}} \in I(\underline{f})$. $\|$

When $n_1 = n_2$, Katz (1963) has shown that ϕ^U is inadmissible when different conditions are placed on the loss function.

When estimates from independent samples are combined, a weighted mean is usually taken, where the weights are proportional to the relative numbers of observations, or amounts of information, of the two samples. The projection provided by ϕ^R does not vary with the sizes of the sample. In this sense, an intuitively more reasonable projection is given by the maximum likelihood estimator,

$$\phi^{ML} = \begin{cases} \phi^U & f_1 \le f_2 \\ ([X_1+X_2]/[n_1+n_2] \ , \ [X_1+X_2]/[n_1+n_2]) & f_1 > f_2 \ . \end{cases}$$

In fact, if $p_1 = p_2 = p$, the most reasonable estimator of the common p is $[X_1+X_2]/(n_1+n_2)$ and uniformly improves $\underline{f}$. In view of the result of Theorem 4.2, that ϕ^R uniformly improves ϕ^U, the following result is rather surprising.

Theorem 4.3. If $2 \le n_2 < n_1(\sqrt{2} - 1)$ or $2 \le n_1 < n_2(\sqrt{2} - 1)$, then ϕ^{ML} does not improve ϕ^U.

Proof: Without loss of generality, take $n_2 < n_1$. If we can show that the projection provided by ϕ^{ML}, $\underline{f}^* = ([n_1f_1 + n_2f_2]/[n_1+n_2] \ , [n_1f_1 + n_2f_2]/[n_1+n_2])$, lies outside $\overline{C(\underline{0}, ||\underline{f} - \underline{0}||)}$ for each $\underline{f} \in \mathbb{H}^R$, then there exists $\underline{p} = (p_1, p_2)$ sufficiently close to $\underline{0}$ and with $p_1 > 0$, such that f^* is outside of $C(\underline{p}, ||\underline{f} - \underline{p}||$, for all $\underline{f} \notin \mathbb{H}^R$. (This follows since $\underline{f}$ can assume only a finite number of values. We need to select such a $\underline{p}$ instead of $\underline{0}$, since otherwise, $\underline{f}$

would lie in $\mathbb{H}^R$ with probability one.) Thus ϕ^U will have smaller loss than ϕ^{ML} at this $\underline{p}$ and at each $\underline{f} \notin \mathbb{H}^R$. Now ϕ^U and ϕ^{ML}, are identical when $\underline{f} \notin \mathbb{H}^R$, so ϕ^U would have strictly smaller risk than ϕ^{ML}, for this $\underline{p}$, and therefore ϕ^{ML} could not improve ϕ^U.

We must now show that for $n_2 < n_1(\sqrt{2} - 1)$, $\underline{f}^*$ lies outside $C(\underline{0}, ||\underline{f} - \underline{0}||)$ for each $\underline{f} \in \mathbb{H}^R$. Note that the line joining f to f^* has constant slope, equal to $-n_1/n_2$. Let $\underline{q} = \underline{q}(\underline{f}) \in \mathbb{H}^R$ be the point where the circle $C(\underline{0}, ||\underline{f} - \underline{0}||)$ intersects the line $p_1 = p_2$. Now, the slope of $\overline{\underline{q}\,\underline{f}}$ is an increasing function of f_2 for fixed f_1 and is constant in f_1 when $f_2 = 0$. Thus if $\underline{f}^* \in C(\underline{0}, ||\underline{f} - \underline{0}||)$ for any $\underline{f}$ such that $f_2 = 0$, then $\underline{f}^*$ will be strictly outside the circle for all $f_1 > f_2 > 0$. We require $n_2 \geq 2$ so that $P\{f_1 > f_2 > 0\} > 0$. Now, $\underline{f}^* \notin C(\underline{0}, ||f - 0||)$ for all $f_2 = 0$ if $||f^* - 0||^2 \geq f_1^2$ that is, if $2[n_1 f_1/(n_1+n_2)]^2 \geq f_1^2$, which occurs whenever $n_2 \leq n_1(\sqrt{2} - 1)$. But $\sqrt{2} - 1$ is irrational, so f^* will lie outside $\overline{C(\underline{0}, ||\underline{f} - \underline{0}||)}$ for all $\underline{f} \notin \mathbb{H}^R$ whenever $2 \leq n_2 < n_1(\sqrt{2} - 1)$, and the proof is complete.

For $n_2 > n_1$, the above argument may be repeated using $C(\underline{1}, ||\underline{f} - \underline{1}||)$. ‖

We have shown that ϕ^{ML} does not improve ϕ^U because it has larger risk at the constructed $\underline{p}$. On the other hand, ϕ^{ML} considerably inproves ϕ^U for parameter values far from this constructed $\underline{p}$ (that is, close to $\underline{1}$ when $n_1 > n_2$).

Actually, Sackrowitz and Strawderman (1974) have shown that ϕ^{ML} is inadmissible, subject to squared error loss, when $n_1 + n_2 \geq 7$

and $n_1 \geq 2$, $n_2 \geq 2$. (It is admissible otherwise.) However, they did not produce an estimator which improves the maximum likelihood estimator, but rather they demonstrated the inadmissibility of ϕ^{ML} using its non-Bayes nature.

The problem of constructing a reasonable estimator for ordered binomial propensities is thus an interesting problem. One might hope that some of the techniques in Chapter II might be useful to this end.

APPENDIX

Some Useful Results

1. A Relationship Between the Distribution Function of a Poisson Random Variable and the Incomplete Gamma Integral

Theorem: Suppose $X \sim P(\theta)$, $T_k \sim G(k+1,1)$.

$$\text{Then } P(X \le k) = \sum_{x=0}^{k} \frac{\theta^x e^{-\theta}}{x!} = \int_{\theta}^{\infty} \frac{t^k e^{-t}}{k!}\, dt = P(T_k > \theta) .$$

Proof: (By Induction). Suppose $k = 0$. $P(X \le 0) = \sum_{x=0}^{0} \frac{\theta^x e^{-\theta}}{x!} = e^{-\theta}$.

$$P(T_0 > \theta) = \int_{\theta}^{\infty} \frac{t^0 e^{-t}}{0!}\, dt = \int_{\theta}^{\infty} e^{-t}\, dt = -e^{-t}\Big|_{\theta}^{\infty} = e^{-\theta} = P(X \le 0) \quad (\surd)$$

Suppose the relationship holds for some $k - 1$. i.e.

$$\sum_{x=0}^{k-1} \frac{\theta^x e^{-\theta}}{x!} = \int_{\theta}^{\infty} \frac{t^{k-1} e^{-t}}{(k-1)!}\, dt .$$

$$\int_{\theta}^{\infty} \frac{t^{k-1} e^{-t}}{(k-1)!}\, dt = \frac{t^k e^{-t}}{k!}\Big|_{\theta}^{\infty} + \int_{\theta}^{\infty} \frac{t^k}{k!} e^{-t}\, dt$$

$$= -\frac{\theta^k e^{-\theta}}{k!} + \int_{\theta}^{\infty} \frac{t^k}{k!} e^{-t}\, dt .$$

(Using Integration by Parts with $u = e^{-t}$, $dv = \frac{t^{k-1}}{(k-1)!}\, dt$, $du = -e^{-t}dt$, $v = t^k/k!$)

So, $\sum_{x=0}^{k-1} \frac{\theta^x e^{-\theta}}{x!} + \frac{\theta^k e^{-\theta}}{k!} = \sum_{x=0}^{k} \frac{\theta^x}{x!} e^{-\theta} = \int_{\theta}^{\infty} \frac{t^k e^{-t}}{k!}\, dt$, and the theorem is true by induction.

2. Fatou's Lemma: Let $f_1, f_2, \ldots$ be a sequence of nonnegative μ-measurable functions defined on χ. Then

$$\liminf_{n\to\infty} \int_{\chi} f_n(x)\, d\mu(x) \ge \int_{\chi} \liminf_{n\to\infty} f_n(x)\, d\mu(x) .$$

BIBLIOGRAPHY

Arnold, B.C. (1973). Two modifications of Goodman's technique for improving estimators (unpublished).

Ash, R.D. (1972). Real Analysis and Probability. Academic Press, New York.

Barlow, Bartholomew, Bremner, Brunk (1972). Statistical Inference Under Order Restrictions — the theory and application of isotonic regression. Wiley, New York.

Blight, B.J.N. (1971). Some general results on reduced mean-square-error estimation. Amer. Statist. 25 #3, 24-25.

Blyth, C.R. (1951). On minimax statistical decision procedures and their admissibility. Ann. Math. Statist. 22, 22-42.

Blyth, C.R. (1974). Necessary and sufficient conditions for inequalities of Cramér-Rao type. Ann. Statist. 2, 464-473.

Blyth, C.R. and D.M. Roberts (1972). On inequalities of Cramér-Rao type and admissibility proofs. Proc. Sixth Berkeley Symp. Math. Statist. Prob. 1, 17-30. Univ. of California Press.

Brewster, J.F. (1974). Alternative estimators for the scale parameter of the exponential distribution with unknown location. Ann. Statist. 2, 553-557.

Brewster, J.F., and J.V. Zidek (1974). Improving on equivariant estimators. Ann. Statist. 2, 21-38.

Ferguson, T.S. (1967). Mathematical Statistics: a decision-theoretic approach. Academic Press, New York.

Goodman, L. (1953). A simple method for improving some estimators. Ann. Math. Statist. 24, 114-117.

Hodges, J.L. Jr., and E.L. Lehmann (1950). Some applications of the Cramér-Rao inequality. Proc. Second Berkeley Symp. Math. Statist. Prob. 1, 13-22. Univ. of California Press.

Johnson, B.M. (1971). On the admissible estimators for certain fixed sample binomial problems. Ann. Math. Statist. 42, 1579-1587.

Kale, B.K. (1970). Inadmissibility of the maximum likelihood estimator in the presence of prior information. Canadian Mathematical Bulletin 13 #3, 391-393.

Karlin, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist. 29, 406-436.

Katz, M.W. (1961). Admissible and minimax estimators of parameters in truncated spaces. Ann. Math. Statist. 32, 136-142.

Katz, M.W. (1963). Estimating ordered probabilities. Ann. Math. Statist. 34, 967-972.

Lehmann, E.L. (1950). Notes on the Theory of Estimation. Univ. of Calif.

Sackrowitz, H., and W. Strawderman (1974). On the admissibility of the M.L.E. for ordered binomial parameters. Ann. Statist. 2, 822-828.

Sacks, J. (1963). Generalized Bayes solutions in estimation problems. Ann. Math. Statist. 34, 751-768.

Stone, M. (1966). Generalized Bayes decision functions, admissibility and the exponential family. Technical Report No. 74, University of Wisconsin, Department of Statistics.

Zacks, S. (1971). The Theory of Statistical Inference. Wiley, New York.

Zidek, J.V. (1970). Sufficient conditions for the admissibility under squared error loss of formal Bayes estimators. Ann. Math. Statist. 41, 446-456.