AN INVESTIGATION OF SOME ALTERNATIVE ESTIMATION PROCEDURES

by

Robert Loyal Davis

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Chairman, Dr. J. C. Arnold

Dr. B. Harshbarger

Dr. R. G. Krutchkoff

Dr. R. H. Myers

Mr. W. E. Pace

July, 1969

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author wishes to express his deep appreciation to Dr. Jesse C. Arnold for his guidance, patience and irrepressible good humor which made this dissertation as pleasant a task as possible.

Special thanks are also given to Dr. Richard G. Krutchkoff for his reading of the rough draft and his numerous helpful suggestions.

The author acknowledges the constructive discussions held with Dr. Raymond H. Myers and Dr. Clyde Y. Kramer.

Thanks also go to Dr. Boyd Harshbarger for his constant encouragement throughout the author's extensive graduate school career.

The author wishes to acknowledge the National Defense Education Act fellowship which provided funds from 1963 to 1966.

Appreciation is extended to for her conscientious typing of this paper.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The statistician often is confronted with the problem of deciding which of two or more estimators to use to estimate the parameter of interest. If he restricts himself to the class of unbiased estimators, then the best estimator he can obtain using squared error loss is that estimator based on the sufficient statistic. If, however, biased estimators are admitted, he frequently finds that some estimators have a mean squared error smaller than others but over only part of the parameter space.

For instance, suppose that one is only considering two estimators of θ, $\hat{\theta}_1(x)$ and $\hat{\theta}_2(x)$, neither of which has a uniformly lower mean squared error than the other. One reasonable approach might be to look at one estimator, say $\hat{\theta}_1$, and decide whether to use it or $\hat{\theta}_2$. This method is called a preliminary test procedure because a test is conducted on $\hat{\theta}_1$ before the selection of the final estimator. Here the preliminary test estimator might be

$$\begin{aligned} \hat{\theta} &= \hat{\theta}_1, \quad \text{if} \quad \hat{\theta}_1 \text{ is not rejected} \\ &= \hat{\theta}_2, \quad \text{if} \quad \hat{\theta}_1 \text{ is rejected.} \end{aligned}$$

It would then seem logical to attempt to select for each value of θ the estimator, $\hat{\theta}_1$ or $\hat{\theta}_2$, which has the smaller mean squared error at that θ value. The above estimator might then become

$$\hat{\theta} = \hat{\theta}_1, \quad \text{if } \hat{\theta}_1 \text{ falls in the } \theta \text{ space}$$

$$\text{where } MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$$

$$= \hat{\theta}_2 \text{ otherwise .}$$

In many estimation problems the experimenter possesses some information concerning the parameter space, knowledge which might enable him to estimate the parameter of interest more exactly. This is not to say that he knows enough to be able to place a prior distribution on the parameter. Rather, it may be that from past experience or personal intuition he feels that the parameter is in a certain region or it may be important to him to estimate precisely if the parameter does lie in a particular region.

An example of this situation would be the problem of estimating the mean, θ, and having the use of two estimators, the unbiased estimate $\bar{x}$ and a biased estimate, $k\bar{x}$, $0 < k < 1$. It is easily seen that $k\bar{x}$ has a lower mean squared error around $\theta = 0$. Thus, if there is reason to believe that θ is near 0 we should utilize this information and choose $k\bar{x}$ at least part of the time. A preliminary test estimator here would be

$$\hat{\theta} = \bar{x} \,, \quad \text{if} \quad \bar{x} \in R$$

$$= k\bar{x}, \quad \text{if} \quad \bar{x} \in \bar{R} \,,$$

where R is the region in which $\bar{x}$ is preferred to $k\bar{x}$ and $\bar{R}$ is the complement of R.

Another approach in attempting to improve the MSE when $k\bar{x}$ and $\bar{x}$ are considered would be to form a linear combination of the two estimators. Any number of weighting systems are available here.

A further application of the preliminary test procedure is its use in a test on two parameters, only one of which is of real interest, For instance, if both the mean and variance of a population were unknown, a preliminary test on the mean could yield information which would provide an improved estimate of the variance.

The aim of this dissertation is to investigate the properties of some preliminary tests and to show how they provide improved estimators when the experimenter has *a priori* information in the form of an initial guess concerning the parameter's location.

CHAPTER II

REVIEW OF LITERATURE

The body of literature on preliminary tests covers a broad spectrum of testing procedures. Most of the first papers are only remotely connected to the topics presented in this dissertation. However, we mention some of the important ones for their historical significance.

The early investigations on preliminary test procedures assumed that two samples were taken, each sample providing an estimate of an unknown parameter. The preliminary test was used as an aid for deciding whether to use just one of the estimates or to pool the two estimates in some fashion. For example, Bancroft (1944), in one of the earliest papers on preliminary tests, considered estimating σ_1^2, under the assumptions that s_1^2 and s_2^2 are independent estimators of σ_1^2 and σ_2^2, that $n_1 s_1^2/\sigma_1^2$ and $n_2 s_2^2/\sigma_2^2$ are independent χ^2's and that $\sigma_2^2 \leq \sigma_1^2$. His rule of procedure was to use

$$\begin{aligned} \hat{\sigma}_1^2 &= (n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2), && \text{if} \quad F = s_1^2/s_2^2 < \lambda \\ &= s_1^2 , && \text{if} \quad F \geq \lambda . \end{aligned}$$

Mosteller (1948) and Kitagawa (1963) investigated the problem of estimating, on the basis of samples from normally distributed variates x_1 and x_2, the mean of x_1. Kitagawa proposed a preliminary test estimator using a pooled t-test as the indicator function. This estimator was defined as

$$\begin{aligned} \hat{\mu}_1 &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}, && \text{if} \quad |t| < t_{n_1 + n_2 - 2} (\alpha) \\ &= \bar{x}_1 , && \text{if} \quad |t| \geq t_{n_1 + n_2 - 2} (\alpha) . \end{aligned}$$

This procedure for "sometimes pooling" estimates of means was also considered by Bennett (1952, 1956) and Asano (1960a, 1960b, 1961). Asano investigated the problem of having estimators based on two samples, S_1 providing stable information, S_2 unstable information but more closely related to the matter at hand. His final estimator was given as

$$\hat{\mu} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}, \quad \text{if} \quad s_2^2 \geq c$$

$$= \bar{x}_2 \quad , \qquad \text{if} \quad s_2^2 < c ,$$

where c is some prescribed constant. In other words, if the unstable sample seems to be especially variable, then the pooled estimator is used. But if S_2 is judged to be sufficiently stable, its mean is used as the final estimator.

More recent literature centers on the use of preliminary tests for situations in which only one sample has been taken, but more than one estimator is considered. Here it is a question of choosing between two or more estimators, each based on the sample obtained.

Arnold (1967) investigated some properties of a general preliminary test (abbreviated PT hereafter) estimator defined as

$$\text{(2.1)} \qquad \hat{\theta}p = \hat{\theta}_1, \quad \text{if} \quad \hat{\theta}_1 \varepsilon R$$

$$= \hat{\theta}_2, \quad \text{if} \quad \hat{\theta}_1 \varepsilon \bar{R} ,$$

where R is the θ space such that $MSE(\hat{\theta}_1|\theta) \leq MSE(\hat{\theta}_2|\theta)$ and $\bar{R}$ is the complement of R. He found that the PT estimation procedure provides a good compromise estimator, one whose MSE is always reasonably close to the MSE of the better of $\hat{\theta}_1$ and $\hat{\theta}_2$ in R and $\bar{R}$.

He also used the Rao-Blackwell theorem to show that if one of the estimators used in the PT procedure, say $\hat{\theta}_1$, is a sufficient statistic, then a uniformly better estimator than (2.1) is

$$(2.2) \qquad \tilde{\theta}_P = \hat{\theta}_1 , \qquad \text{if} \quad \hat{\theta}_1 \varepsilon R$$

$$= E(\hat{\theta}_2|\hat{\theta}_1), \qquad \text{if} \quad \hat{\theta}_1 \varepsilon \bar{R} ,$$

where $R = \{\theta|MSE(\hat{\theta}_1|\theta) \leq MSE(\hat{\theta}_2|\theta)\}$. This procedure was generalized to the k-parameter estimation problem.

Thompson (1968) employed a so-called "shrinkage technique" for estimating the mean. He compared $MSE(c\bar{x})$ with $MSE(\bar{x})$ for $c = \bar{x}^2/(\bar{x}^2 + \sigma^2/n)$ and $c = \bar{x}^2/(\bar{x}^2 + ks^2/n)$ and found that, not unexpectedly, $c\bar{x}$ performed better around the "natural origin", μ_o, $\bar{x}$ was better in the middle ranges of

$$\left|\frac{\mu - \mu_o}{\sigma_{\bar{x}}}\right| ,$$

and $\bar{x}$ and $c\bar{x}$ performed approximately the same for large values of

$$\left|\frac{\mu - \mu_o}{\sigma_{\bar{x}}}\right| .$$

Thompson also examined the use of a preliminary t-test to decide whether to choose c = 0 or c = 1; i.e.,

$$\hat{c} = 0, \quad \text{if} \quad \left|\frac{\bar{x}}{s\sqrt{n}}\right| \leq t_{\alpha, n-1}$$

$$= 1, \quad \text{if} \quad \left|\frac{\bar{x}}{s\sqrt{n}}\right| > t_{\alpha,\, n-1} \,.$$

Alam and Thompson (1968) assumed a prior distribution on θ and used the PT to test the marginal distribution of x. That is, if $f(x|\theta) = N(\theta, \sigma^2)$ and *a priori* $g(\theta) = N(\theta_o, \tau^2)$, then the marginal distribution of x is $N(\theta_o, \sigma^2 + \tau^2)$ and the marginal distribution corresponding to $\tau \to \infty$ (the uniform distribution) is tested against that distribution corresponding to some fixed finite τ. Their estimator was defined as

$$\hat{\theta} = \theta_0 + \lambda(x - \theta_o), \quad \text{for} \quad |x - \theta_o| \leq c\sigma$$

$$= x, \qquad\qquad \text{for} \quad |x - \theta_o| > c\sigma \,,$$

where $\lambda = \tau^2/(\sigma^2 + \tau^2)$ and $c > 0$. Here $\theta_o + \lambda(x - \theta_o)$ is the Bayes estimate of θ for finite τ^2 and x is the Bayes estimate of θ for uniform prior. Alam and Thompson showed that although $\hat{\theta}$ is not a Bayes estimate it also does not have unbounded risk as the Bayes estimates do.

A two-parameter PT problem was investigated by Kale (1966). He assumed a random sample of size two, (z_1, z_2), from the population whose probability density function is

$$f(z, \theta, \sigma) = \frac{1}{\sigma} \exp\left[-\left(\frac{z - \theta/2}{\sigma}\right)\right], \quad z \geq \theta,$$

$$-\infty < \theta < \infty$$

$$\sigma > 0 \,,$$

He supposes that σ is the parameter of interest but it must be estimated in the presence of the nuisance parameter θ. Then, if $z_{(1)}$ and $z_{(2)}$

represent the order statistics for a sample of size two, $\hat{\theta} = 2z_{(1)}$ and $z_{(1)} + z_{(2)}$ are jointly complete and sufficient for θ and σ, while $\hat{\sigma} = z_{(2)} - z_{(1)} = (z_{(1)} + z_{(2)}) - 2z_{(1)}$ is the uniformly minimum variance estimator of σ. Kale's estimator for σ is given by

$$\tilde{\sigma} = \hat{\sigma}, \quad \text{if} \quad \hat{\theta} < 0, \quad \text{or if} \quad \hat{\theta} > 0 \quad \text{and} \quad \hat{\theta} \geq \hat{\sigma}$$

$$= \frac{\hat{\sigma} + \hat{\theta}}{2}, \quad \text{if} \quad \hat{\theta} > 0 \quad \text{and } \hat{\theta} < \hat{\sigma} .$$

Note that both R and $\bar{R}$ are two-dimensional here. The mean squared error of $\tilde{\sigma}$ is uniformly smaller than that of $\hat{\sigma}$. Hence, Kale's PT estimator causes the uniformly minimum variance unbiased estimator to be inadmissible with respect to squared error loss functions.

CHAPTER III

A PT ESTIMATOR USING $\bar{x}$ AND $k\bar{x}$

In this chapter we give attention to the PT estimator which chooses between $\bar{x}$ and $k\bar{x}$. First we shall look at the bias and mean squared error functions for this estimator and finally some admissibility considerations for the estimator are examined.

We assume that x is a random variable with mean θ and variance σ^2 on which we take a random sample of size n. The mean θ is to be estimated.

As mentioned in the Introduction, if we feel that the parameter is near some "natural origin", θ_o, we should seek an estimator which will perform well in the region around θ_o. Without loss of generality we suppose that $\theta_o = 0$.

Next we consider the MSE functions for $\bar{x}$ and $k\bar{x}$, where

(1) $$\mathrm{MSE}(\bar{x}|\theta) = \mathrm{var}(\bar{x}|\theta) = \sigma_{\bar{x}}^2$$

(2) $$\mathrm{MSE}(k\bar{x}|\theta) = \mathrm{var}(k\bar{x}|\theta) + \mathrm{Bias}^2\ (k\bar{x}|\theta)$$

$$= k^2\ \sigma_{\bar{x}}^2 + \theta^2\ (1 - k)^2\ .$$

It is easily seen that $\mathrm{MSE}(k\bar{x}|\theta) > \mathrm{MSE}(\bar{x}|\theta)$ uniformly if $k > 1$, and hence, only values of k between 0 and 1 will be considered.

The points of intersection of $\mathrm{MSE}(\bar{x}|\theta)$ and $\mathrm{MSE}(k\bar{x}|\theta)$ are found to be at

$$\theta = \pm \sqrt{1 - k^2}\ \sigma_{\bar{x}}/(1 - k)$$

so that

$$MSE(k\bar{x}|\theta) < MSE(\bar{x}|\theta)$$

for

$$-\sqrt{1-k^2}\,\sigma_{\bar{x}}/(1-k) < \theta < \sqrt{1-k^2}\,\sigma_{\bar{x}}/(1-k)\ .$$

The larger the value of k ($0 < k < 1$), the wider the θ region in which $k\bar{x}$ is superior to $\bar{x}$, and $k\bar{x}$ always has a smaller MSE than $\bar{x}$ at $\theta = 0$ for $0 < k \leq 1$. Of course, for values of θ sufficiently distant from $\theta = 0$, the $MSE(k\bar{x}|\theta)$ function increases rapidly. Rather than use either $k\bar{x}$ or $\bar{x}$ all the time we propose the following PT estimator:

$$\begin{aligned} \hat{\theta}p &= \bar{x}, \quad \text{if} \quad \bar{x} \in R \\ &= k\bar{x}, \quad \text{if} \quad \bar{x} \in \bar{R}\ , \end{aligned} \tag{3.1}$$

where $0 < k \leq 1$, R is defined as the region of θ in which $MSE(\bar{x}|\theta) < MSE(k\bar{x}|\theta)$, and $\bar{R}$ is the complement of R. Hence, our procedure is to draw a sample of size n and compute $\bar{x}$. If $\bar{x}$ is contained in the region where $\bar{x}$ has smaller MSE than $k\bar{x}$, we use $\bar{x}$ as the estimator. Otherwise, we use $k\bar{x}$.

To see how k might be chosen, we consider an example problem. Suppose that we take a sample of size 10 on x, a random variable with unknown mean θ and known variance 50. In addition, suppose that our guesstimate of θ lies between 20 and 30. Scaling these θ values so that $\theta_o = 0$, we say that we are confident that θ lies between -5 and 5. Then

$$MSE(\bar{x}) = \sigma_{\bar{x}}^2 = 5$$

and

$$MSE(k\bar{x}) = 5k^2 + \theta^2 (1 - k)^2 .$$

We wish to find the smallest k such that $k\bar{x}$ will have lower MSE than $\bar{x}$ over the region in which we expect θ to lie. θ^2 attains a maximum of 25 over this region. Then

$$MSE(\bar{x}) \geq MSE(k\bar{x})$$

implies

$$5 \geq 5k^2 + 25(1 - k)^2 .$$

This inequality holds for $2/3 \leq k \leq 1$. Then $k = 2/3$ is the smallest value of k consistent with our objectives. Nevertheless, in order to keep away from the unbounded risk which one encounters when using $k\bar{x}$, we consider $\hat{\theta}p$ in hopes that because it uses $\bar{x}$ part of the time it will not have unbounded risk.

3.1 Bias and Mean Squared Error of $\hat{\theta}p$

We will now consider some statistical properties of $\hat{\theta}p$. The bias function for $\hat{\theta}p$ is given by

$$B(\hat{\theta}p|\theta) = E(\hat{\theta}p|\theta) - \theta .$$

But

$$E(\hat{\theta}p|\theta) = E_R(\bar{x}|\theta) + E_{\bar{R}}(k\bar{x}|\theta)$$

where

$$E_R(\bar{x}|\theta) = \int_R \bar{x}\ f(\bar{x})\ d\bar{x} .$$

Also,

$$\theta = E_R(\bar{x}|\theta) + E_{\bar{R}}(\bar{x}|\theta) ,$$

and hence

$$B(\hat{\theta}p|\theta) = (k - 1) E_{\bar{R}}(\bar{x}|\theta) .$$

Here we assume that either by some of the central limit theorem considerations, or because the x's themselves are normally distributed, $\bar{x}$ is distributed $N(\theta, \sigma^2/n) = N(\theta, \sigma_{\bar{x}}^2)$. Then

$$(3.1.1) \qquad B(\hat{\theta}p|\theta) = (k - 1) \int_a^b \frac{\bar{x}}{\sqrt{2\pi}\,\sigma_{\bar{x}}} \epsilon^{-\frac{(\bar{x} - \theta)^2}{2\sigma_{\bar{x}}^2}} d\bar{x} ,$$

where $b = (1 - k^2)^{1/2} \sigma_{\bar{x}}/(1 - k)$ and $a = -(1 - k^2)^{1/2} \sigma_{\bar{x}}/(1 - k)$. Let $z = (\bar{x} - \theta)/\sigma_{\bar{x}}$, $\eta(\cdot)$ and $N(\cdot)$ represent the density and cumulative density for the standard normal, and then (3.1.1) can be written as

$$(k - 1) \int_{\frac{a - \theta}{\sigma_{\bar{x}}}}^{\frac{b - \theta}{\sigma_{\bar{x}}}} (z\sigma_{\bar{x}} + \theta)\, \eta(z)\, dz$$

$$= (k - 1)\left[\sigma_{\bar{x}} \int_{\frac{a-\theta}{\sigma_{\bar{x}}}}^{\frac{b-\theta}{\sigma_{\bar{x}}}} z\, \eta(z)\, dz + \theta \int_{\frac{a-\theta}{\sigma_{\bar{x}}}}^{\frac{b-\theta}{\sigma_{\bar{x}}}} \eta(z)\, dz\right]$$

$$= (k - 1)\left[\sigma_{\bar{x}}\left\{- \eta\left(\frac{b-\theta}{\sigma_{\bar{x}}}\right) + \eta\left(\frac{a-\theta}{\sigma_{\bar{x}}}\right)\right\} + \theta\left\{N\left(\frac{b-\theta}{\sigma_{\bar{x}}}\right) - N\left(\frac{a-\theta}{\sigma_{\bar{x}}}\right)\right\}\right].$$

Thus, the bias function for $\hat{\theta}p$ becomes

(3.1.2) $$B(\hat{\theta}p|\theta) = (1 - k)\left[\sigma_{\bar{x}}\left\{- \eta\left(\frac{a-\theta}{\sigma_{\bar{x}}}\right) + \eta\left(\frac{b-\theta}{\sigma_{\bar{x}}}\right)\right\} + \theta\left\{N\left(\frac{a-\theta}{\sigma_{\bar{x}}}\right) - N\left(\frac{b-\theta}{\sigma_{\bar{x}}}\right)\right\}\right].$$

For comparison with other estimators we let $c_a = (\theta - a)/\sigma_{\bar{x}}$ and $c_b = (\theta - b)/\sigma_{\bar{x}}$. Using the relationships $\eta(-t) = \eta(t)$, $N(-t) = 1 - N(t)$, and the fact that $a = - b$ we have

(3.1.3) $$B(\hat{\theta}p|\theta) = (1 - k)\ [\sigma_{\bar{x}}\{-\ n(-c_a) + n(-c_b)\} + \theta\{N(-c_a) - N(c_b)\}]$$

$$= (1 - k)\ [\sigma_{\bar{x}}\{-\ n(c_a) + n(c_a)\} + \theta\{N(c_a) - N(c_a)\}]\ .$$

It is easily seen that the bias of $\hat{\theta}p$ is negative when θ is positive, positive for negative θ, and zero when $\theta = 0$. Also, we see that the bias curve is symmetric with respect to the origin. The bias function is plotted for $\theta > 0$, $k = .5$, and $\sigma_{\bar{x}} = 1$ in Figure 2 in Chapter IV.

We now examine the mean squared error of $\hat{\theta}p$, which is

$$MSE(\hat{\theta}p|\theta) = \int_R (\bar{x} - \theta)^2\ f(\bar{x}|\theta)\ d\bar{x} + \int_{\bar{R}} (k\bar{x} - \theta)^2\ f(\bar{x}|\theta)\ d\bar{x}\ .$$

If we write the variance of $\bar{x}$ as

$$\int_R (\bar{x} - \theta)^2\ f(\bar{x}|\theta)\ d\bar{x} + \int_{\bar{R}} (\bar{x} - \theta)^2\ f(\bar{x}|\theta)\ d\bar{x}\ ,$$

then the

$$MSE(\hat{\theta}p|\theta) = \sigma_{\bar{x}}^2 - \int_{\bar{R}} \{(1 - k^2)\ \bar{x}^2 - 2(1 - k)\ \theta\bar{x}\}\ f(\bar{x}|\theta)\ d\bar{x}\ .$$

Utilizing the fact that $\int t^2\ n(t)\ dt = -\ t\ n(t) + N(t)$ and making the same $z = (\bar{x} - \theta)/\sigma_{\bar{x}}$ transformation, we find that

(3.1.4) $$MSE(\hat{\theta}p|\theta) = \sigma_{\bar{x}}^2 + [-(1 - k^2)\ (\sigma_{\bar{x}}^2 + \theta^2) + 2(1 - k)\ \theta^2]$$

$$[N(c_a) - N(c_b)] + [(1 - k^2)\, \sigma_{\bar{x}}^2\, c_a - 2k(1 - k)\, \sigma_{\bar{x}}\, \theta]\, \eta(c_a)$$

$$+ [-(1 - k^2)\, \sigma_{\bar{x}}^2\, c_b + 2k(1 - k)\, \sigma_{\bar{x}}\, \theta]\, \eta(c_b)\ .$$

The MSE$(\hat{\theta}p|\theta)$ function is plotted in Figure 3.1 for $k = .5$, $\sigma_{\bar{x}} = 1$, and for $b = a = \sqrt{3} = 1.732$. It is symmetric about the MSE-axis (the line $\theta = 0$). The MSE functions for $\bar{x}$ and $k\bar{x}$ are also shown in Figure 1. We note that $\hat{\theta}p$ has a smaller MSE than $\bar{x}$ in the region $|\theta| \leq 1$ and has a higher MSE than either $\bar{x}$ or $k\bar{x}$ for approximately $1 < |\theta| < 2$. When $|\theta| > 2$, the MSE$(\hat{\theta}p|\theta)$ approaches $\sigma_{\bar{x}}^2$ while the MSE$(k\bar{x}|\theta)$ diverges.

In addition, Figure 1 shows the MSE$(\hat{\theta}p|\theta)$ when a and b are fixed at $\pm\ (2\sqrt{3} - 1)$ and $\pm\ 1$, while holding $k = .5$. From the figure it is seen that, not surprisingly, the farther away from 0 that a and b are moved the more pronounced the difference between MSE$(\hat{\theta}p|\theta)$ and MSE$(\bar{x}|\theta)$.

There is a rather large region where $\hat{\theta}p$ performs appreciably worse than $\bar{x}$. In Chapter IV we consider an estimator which attempts to improve upon $\hat{\theta}p$ in this region. First, however, we shall consider some admissibility aspects of $\hat{\theta}p$.

3.2 Admissibility of the PT Estimator

We now discuss the question of the admissibility of the PT estimator defined as

$$\hat{\theta}p = \bar{x}, \quad \text{if} \quad \bar{x} \in R$$

$$= k\bar{x}, \quad \text{if} \quad \bar{x} \in \bar{R}\ ,$$

and we will say that $\hat{\theta}p$ is admissible universally if there is no estimator ϕ for the parameter θ such that

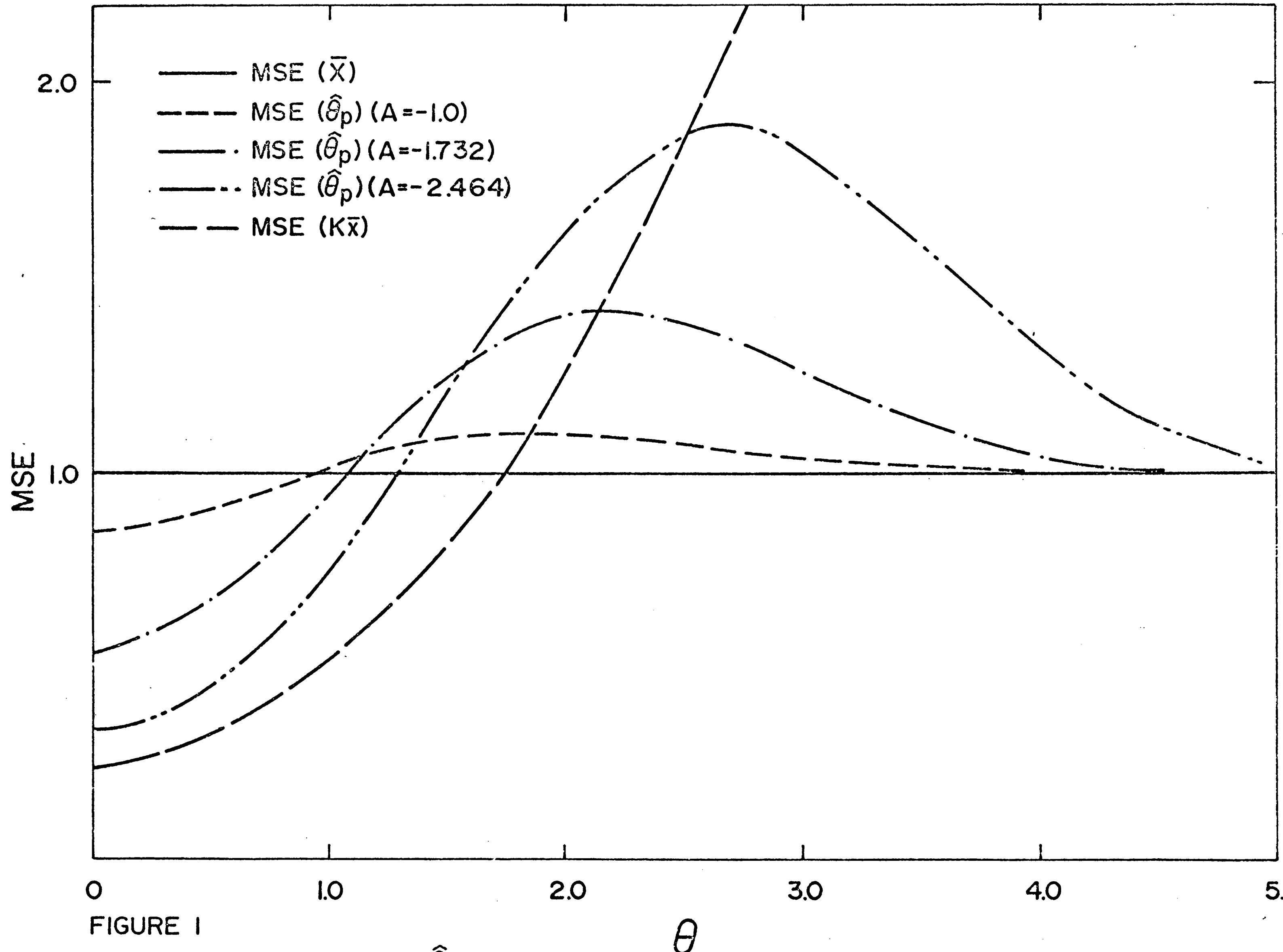

FIGURE 1

MSE OF PT ESTIMATOR $\hat{\theta}_p$ $\sigma_{\bar{x}} = 1$, K=.5

$$E(\phi - \theta)^2 \leq E(\hat{\theta}p - \theta)^2 ,$$

for all θ, with strict inequality for some θ.

Sacks (1963) showed that for squared error loss and an exponentially distributed random variable (x) any admissible estimate is a Bayes estimate. Cohen (1965) pointed out that any Bayes estimate is either an analytic function of x or is equivalent to an analytic function. Thus, if an estimator is not analytic nor is equivalent to an analytic function, it cannot be admissible. Looking at $\hat{\theta}p$, we see that at the boundary of R and $\bar{R}$ (where we change from $\bar{x}$ to $k\bar{x}$ or vice versa, say at $\bar{x} = \bar{x}_o$) the derivative of $\hat{\theta}p$ with respect to $\bar{x}$ does not exist. This becomes obvious when we see that the limit of $\hat{\theta}p$ as $\bar{x}$ approaches $\bar{x}_o$ from the R side is $\bar{x}_o$ whereas the limit as $\bar{x}$ approaches $\bar{x}_o$ from the $\bar{R}$ side is $k\bar{x}_o$. Therefore, the derivative does not exist at $\bar{x}_o$, $\hat{\theta}p$ is neither analytic nor equivalent to an analytic function, and hence, is inadmissible. This shows that there exists an estimator which uniformly improves upon $\hat{\theta}p$ in terms of squared error risk. Finding that estimator is another matter. Alam and Thompson (1968) investigated a quite similar problem but could find no such estimator.

Of greater practical importance is the question of whether $\hat{\theta}p$ is admissible with respect to $\bar{x}$, the UMVUE of θ. This can be easily shown by considering $MSE(\bar{x}|\theta)$ and $MSE(\hat{\theta}p|\theta)$ when $\theta = 0$. Then

$$MSE(\bar{x}|\theta = 0) = \int_R \bar{x}^2 f(\bar{x})\, d\bar{x} + \int_{\bar{R}} \bar{x}^2 f(\bar{x})\, d\bar{x}$$

and

$$MSE(\hat{\theta}p|\theta = 0) = \int_R \bar{x}^2 f(\bar{x})\, d\bar{x} + k^2 \int_{\bar{R}} \bar{x}^2 f(\bar{x})\, d\bar{x} \ .$$

Since k is between 0 and 1,

$$k^2 \int_{\bar{R}} \bar{x}^2 f(\bar{x})\, d\bar{x} < \int_{\bar{R}} \bar{x}^2 f(\bar{x})\, d\bar{x} \ ,$$

and

$$MSE(\hat{\theta}p|\theta = 0) < MSE(\bar{x}|\theta = 0) \ .$$

Thus $\hat{\theta}p$ cannot be inadmissible with respect to $\bar{x}$.

CHAPTER IV

A WEIGHTED ESTIMATOR USING $\bar{x}$ AND $k\bar{x}$

In order to improve upon the poor performance of $\hat{\theta}p$ in the areas around the points of intersection (a,b) of MSE($\bar{x}$) with MSE($k\bar{x}$), we consider the weighted estimator

$$\hat{\theta} = \bar{x}\, g(\bar{x}) + k\bar{x}[1 - g(\bar{x})] \ .$$

We would like for $\hat{\theta}$ to have the property that, for a given value of $\bar{x}$, $\bar{x}$ or $k\bar{x}$ is given greater weight depending on which is more likely to have the smaller MSE. Conceivably, there are many such weighting functions $g(\bar{x})$ which will provide this property. Consider the weighting function:

$$g(\bar{x}) = \int_{\frac{\bar{x} - a}{\sigma_{\bar{x}}}}^{\infty} \eta(z) + \int_{-\infty}^{\frac{\bar{x} - b}{\sigma_{\bar{x}}}} \eta(z)\, dz$$

where

$$z = \frac{\bar{x} - \theta}{\sigma_{\bar{x}}} ,$$

and $\eta(z)$ is the standard normal density. Note also that

$$1 - g(\bar{x}) = \int_{\frac{\bar{x} - b}{\sigma_{\bar{x}}}}^{\frac{\bar{x} - a}{\sigma_{\bar{x}}}} \eta(z)\, dz \ .$$

We would prefer to look at $g(\bar{x})$ as a weighting function rather than as a fiducial probability. However, it does seem to make sense intuitively from a fiducial probability point of view; i.e.,

$$\hat{\theta} = k\bar{x}\ P[(a < \theta < b)|\bar{x}] + \bar{x}\{P[(\theta < a)|\bar{x}] + P[(\theta > b)|\bar{x}]\}$$

$$= k\bar{x}\ P\left[\left(\frac{\bar{x} - a}{\sigma_{\bar{x}}} > \frac{\bar{x} - \theta}{\sigma_{\bar{x}}} > \frac{\bar{x} - b}{\sigma_{\bar{x}}}\right)\middle|\bar{x}\right]$$

$$+ \bar{x}\left\{P\left[\left(\frac{\bar{x} - a}{\sigma_{\bar{x}}} < \frac{\bar{x} - \theta}{\sigma_{\bar{x}}}\right)\middle|\bar{x}\right] + P\left[\left(\frac{\bar{x} - b}{\sigma_{\bar{x}}} > \frac{\bar{x} - \theta}{\sigma_{\bar{x}}}\right)\middle|\bar{x}\right]\right\}.$$

To see that $g(\bar{x})$ fits the concept of fiducial probability one need only look at the logical development of a fiducial probability statement as shown in Rao (1965).

We shall assume that $\bar{x}$ is normally distributed with unknown mean θ, known variance $\sigma_{\bar{x}}^2$. Then

$$z = \frac{\bar{x} - \theta}{\sigma_{\bar{x}}} \cap N(0,1)$$

and

$$\frac{\bar{x} - a}{\sigma_{\bar{x}}} = \frac{\bar{x} - \theta}{\sigma_{\bar{x}}} + \frac{\theta - a}{\sigma_{\bar{x}}} = z + c_a\ ,$$

where $c_a = (\theta - a)/\sigma_{\bar{x}}$ as in Chapter III. Thus,

$$\hat{\theta} = k(z\sigma_{\bar{x}} + \theta)\ [N(z + c_a) - N(z + c_b)]$$

$$+ (z\sigma_{\bar{x}} + \theta)\ [1 - N(z + c_a) + N(z + c_b)]$$

$$= (z\sigma_{\bar{x}} + \theta) + (1 - k)\ (z\sigma_{\bar{x}} + \theta)\ [N(z + c_b) - N(z + c_a)]\ .$$

At this point it becomes obvious that in order to compute the bias and MSE of $\hat{\theta}$ we will need to be able to evaluate the expected values of such quantities as $N(z + c)$, $zN(z + c)$, etc. The expected values which can be expressed in closed form will be presented here in the form of a theorem.

Theorem 4.1: If z is a standard normal random variable, then

(i) $$E\{N(az)\} = \frac{1}{2}$$

(ii) $$E\{N(az + c)\} = N(c/\sqrt{a^2 + 1}\ )$$

(iii) $$E\{zN(z + c)\} = \frac{e^{-\frac{c^2}{4}}}{2\sqrt{\pi}}$$

(iv) $$E\{z^2\ N(z + c)\} = N\left(\frac{c}{\sqrt{2}}\right) - \frac{ce^{-\frac{c^2}{4}}}{4\sqrt{\pi}}$$

where a and c are constants.

Proof:

(i) Let $$f(a) = E\{N(az)\} = \int_{-\infty}^{\infty} N(az)\ n(z)\ dz\ .$$

Using the differentiation with respect to a parameter method of integral evaluation described in Widder (1961) and the fact that

$$\frac{d}{da} N(az) = z\, n(az) \ ,$$

we find

$$f'(a) = \int_{-\infty}^{\infty} z\, n(az)\, n(z)\, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{(a^2z^2+z^2)}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{(a^2+1)z^2}{2}} dz \ .$$

Next, we let $y = (a^2 + 1)^{1/2} z$ and find

$$f'(a) = \frac{(a^2+1)^{-1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y\, n(y)\, dy \ .$$

But this integral is just the mean of a standard normal so that

$$f'(a) = 0$$

or

$$f(a) = k \ . \tag{4.1}$$

To evaluate k we set a = 0 in the original integral and have

$$f(0) = \int_{-\infty}^{\infty} N(0)\ \eta(z)\ dz = \frac{1}{2}$$

and, from (4.1), f(0) = k, so that k = 1/2 and f(a) = 1/2.

Part (i) is not used directly, but we do need its result to prove part (ii).

(ii) We use the same method as for part (i).

$$f(c) = E\{N(az + c)\} = \int_{-\infty}^{\infty} N(az + c)\ \eta(c)\ dz \ .$$

$$f'(c) = \int_{-\infty}^{\infty} \eta(az + c)\ \eta(c)\ dc$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(a^2z^2 + 2acz + c^2 + z^2)}\ dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(a^2+1)\ z^2 + 2acz + c^2]}\ dz \ .$$

To complete the square on the exponent of e, we let

$$y = (a^2+1)^{1/2}\ z + (a^2+1)^{-1/2}\ ac \ .$$

Then

$$f'(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (a^2+1)^{-1/2} e^{-\frac{c^2}{2(a^2+1)}} \eta(y)\, dy$$

$$= \frac{1}{\sqrt{2\pi}} (a^2+1)^{-1/2} e^{-\frac{c^2}{2(a^2+1)}} .$$

To find $f(c)$ we may integrate $f'(c)$ from any arbitrary lower limit to an upper limit c. We pick our lower limit as $-\infty$. Then

$$f(c) = (a^2+1)^{-1/2} \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2(a^2+1)}} dt + k .$$

Let $u = t(a^2+1)^{-1/2}$ and we have

$$f(c) = \int_{-\infty}^{c(a^2+1)^{-1/2}} \eta(u)\, du + k$$

or

$$(4.2) \qquad f(c) = N(c/\sqrt{a^2+1}\,) + k .$$

But

$$f(0) = \int_{-\infty}^{\infty} N(az)\, \eta(z)\, dz = \frac{1}{2}$$

and from (4.2)

$$f(0) = \frac{1}{2} + k \ .$$

Hence, k = 0 and

$$f(c) = N(c/\sqrt{a^2 + 1}\ ) \ .$$

<u>Corollary</u>: If z is a standard normal random variable,

$$E\{N(z + c)\} = N\left(\frac{c}{\sqrt{2}}\right) ,$$

where c is a constant.

Proof: Let a = 1 in part (ii).

There is another proof of this corollary, a proof which uses convolution theory. We include it only for academic interest. First, we let

$$N(z + c) = P(y \leq z + c) \quad \text{for} \quad y \cap N(0,1) \ .$$

Next we set u = y - z so that

$$P(y \leq z + c) = P(u \leq c)$$

$$= \int_{-\infty}^{c} f(u)\, du$$

(4.3)
$$= \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(u+z)^2}{2}}\, du$$

since $u \cap N(-z,1)$.

Now we let $v = (u+z)/\sqrt{2}$, $du = \sqrt{2}\, dv$, and (4.3) becomes

$$\int_{-\infty}^{\frac{c}{\sqrt{2}} + \frac{z}{\sqrt{2}}} \frac{\sqrt{2}}{\sqrt{2\pi}} \cdot e^{-\frac{v^2}{(2)(1/2)}} dv$$

$$= \int_{-\infty}^{\frac{c}{\sqrt{2}} - \left(\frac{-z}{\sqrt{2}}\right)} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/2}} e^{-\frac{v^2}{2(1/2)}} dv$$

$$= N\left[\frac{c}{\sqrt{2}} - \left(\frac{-z}{\sqrt{2}}\right) \middle| \mu = 0,\ \sigma^2 = \frac{1}{2}\right] .$$

Now

$$E\{N(z + c)\} = \int_{-\infty}^{\infty} N\left[\frac{c}{\sqrt{2}} - \left(\frac{-z}{\sqrt{2}}\right) \middle| \mu = 0,\ \sigma^2 = \frac{1}{2}\right] dN[z|\mu = 0,\ \sigma^2 = 1] .$$

But

$$dN(z|\mu = 0,\ \sigma^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-z)^2}{2}} .$$

Next we let $t = z/\sqrt{2}$ so that

$$f(t) = \frac{\sqrt{2}}{\sqrt{2\pi}} e^{-\frac{t^2}{(2)(1/2)}} = \frac{1}{\sqrt{2\pi}\ \sqrt{1/2}} e^{-\frac{(-t)^2}{(2)(1/2)}} .$$

Then

$$E\{N(z + c)\} = \int_{-\infty}^{\infty} N\left[\frac{c}{\sqrt{2}} - (-t)\right] dN(-t)$$

$$= N\left(\frac{c}{\sqrt{2}}\right)$$

by convolution theorems.

(iii) This part can be proved using the differentiation with respect to a parameter method, but in this case it is easier to just change the order of integration. We have

$$I_3 = E\{zN(z + c)\} = \int_{-\infty}^{\infty} zN(z + c)\ \eta(z)\ dz$$

$$= \int_{-\infty}^{\infty} z\ \eta(z) \int_{-\infty}^{z+c} \eta(t)\ dt\ dz\ .$$

Changing the order of integration gives us

$$I_3 = \int_{-\infty}^{\infty} \eta(t) \int_{t-c}^{\infty} z\ \eta(z)\ dz\ dt$$

$$= \int_{-\infty}^{\infty} \eta(t)\ dt\ [- \eta(z)|_{t-c}^{\infty}]$$

$$= \int_{-\infty}^{\infty} \eta(t)\ \eta(t - c)\ dt$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} \; e^{-(1/2)(2t^2 - 2ct + c^2)} \; dt \; .$$

But

$$2t^2 - 2ct + c^2 = \left(\sqrt{2}\; t - \frac{c}{\sqrt{2}}\right)^2 + \frac{c^2}{2}$$

so

$$I_3 = \frac{1}{\sqrt{2\pi}} \; e^{-\, c^2/4} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \; e^{-\,(1/2)(\sqrt{2}\; t \; - \; c/\sqrt{2})^2} \; dt \; .$$

Now let $w = \sqrt{2}\; t - c/\sqrt{2}$ and we have

$$I_3 = \frac{e^{-c^2/4}}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \eta(w) \; dw$$

$$= \frac{e^{-c^2/4}}{2\sqrt{\pi}} \; .$$

(iv) We use the same method as in part (iii).

$$I_4 = E\{z^2 N(z + c)\} = \int_{-\infty}^{\infty} z^2 \; \eta(z) \int_{-\infty}^{z+c} \eta(t) \; dt \; dz$$

$$= \int_{-\infty}^{\infty} \eta(t) \int_{t-c}^{\infty} z^2 \, \eta(z) \, dz \, dt$$

$$= \int_{-\infty}^{\infty} \eta(t) \, [(t - c) \, \eta(t - c) + 1 - N(t - c)] \, dt \, .$$

We break I_4 down into three integrals. $I_4 = (a) + (b) + (c)$.

$$(a) = \int_{-\infty}^{\infty} (t - c) \, \eta(t) \, \eta(t - c) \, dt \, .$$

Let $w = \sqrt{2}\, t - c/\sqrt{2}$.

$$(a) = \frac{e^{-c^2/4}}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\frac{w}{\sqrt{2}} - \frac{c}{2} \right) \eta(w) \, dw$$

$$= - \frac{c e^{-c^2/4}}{4\sqrt{\pi}} \, ,$$

$$(b) = \int_{-\infty}^{\infty} \eta(t) \, dt = 1 \, ,$$

$$(c) = - \int_{-\infty}^{\infty} N(t - c) \, \eta(t) \, dt$$

$$= - N\left(- \frac{c}{\sqrt{2}}\right)$$

by part (ii)

$$= N\left(\frac{c}{\sqrt{2}}\right) - 1 .$$

Thus,

$$I_4 = N\left(\frac{c}{\sqrt{2}}\right) - \frac{ce^{-c^2/4}}{4\sqrt{\pi}} ,$$

and the theorem is proved.

Unfortunately, these methods break down for finding expected values of terms involving $N^k(z + c)$ for $k > 1$ and such terms have to be referred to the computer for numerical integration.

We now look at the bias of $\hat{\theta}$.

$$E(\hat{\theta}|\theta) = E(z\sigma_{\bar{x}} + \theta) + (1 - k)\ E\{\sigma_{\bar{x}}\ N(z + c_b)$$

$$- \sigma_{\bar{x}}\ N(z + c_a) + \theta N(z + c_a) - \theta N(z + c_a)\}$$

$$= \theta + (1 - k)\left[\sigma_{\bar{x}}\left(\frac{1}{2\sqrt{\pi}}\ e^{-c_b^2/4} - \frac{1}{2\sqrt{\pi}}\ e^{-c_a^2/4}\right)\right.$$

$$\left. + \theta\left\{N\left(\frac{c_b}{\sqrt{2}}\right) - N\left(\frac{c_a}{\sqrt{2}}\right)\right\}\right] ,$$

so that

$$\text{(4.4)} \qquad B(\hat{\theta}|\theta) = (1-k)\left[\frac{\sigma_{\bar{x}}}{\sqrt{2}}\left\{-n\left(\frac{c_a}{\sqrt{2}}\right)+n\left(\frac{c_b}{\sqrt{2}}\right)\right\} + \theta\left\{N\left(\frac{c_b}{\sqrt{2}}\right)-N\left(\frac{c_a}{\sqrt{2}}\right)\right\}\right] .$$

Note the parallels between the functional form of $B(\hat{\theta}|\theta)$ and that of $B(\hat{\theta}p|\theta)$ in (3.1.2). A comparison of the two bias functions is given graphically in Figure 2 for $k = .5$ and $\sigma_{\bar{x}} = 1$. Similar to the $B(\hat{\theta}p|\theta)$, the $B(\hat{\theta}|\theta)$ is negative for $\theta > 0$, positive for $\theta < 0$, zero for $\theta = 0$, and symmetric with respect to the origin.

The $MSE(\hat{\theta}|\theta)$ function is given by

$$MSE(\hat{\theta}|\theta) = var(\hat{\theta}|\theta) + B^2(\hat{\theta}|\theta)$$

$$= E(\hat{\theta}^2|\theta) + \theta^2 - 2\theta E(\hat{\theta}|\theta) .$$

We have

$$E(\hat{\theta}^2|\theta) = E[(z^2\ \sigma_{\bar{x}}^{\ 2} + 2z\sigma_{\bar{x}}\ \theta + \theta^2)$$

$$\{1 + 2\ (1-k)\ (N(z + c_b) - N(z + c_a))$$

$$+ (N(z + c_b) - N(z + c_a))^2\}] ,$$

which is considered separately in five parts.

$$\text{(a)} \qquad E(z^2\ \sigma_{\bar{x}}^{\ 2} + 2z\sigma_{\bar{x}}\ \theta + \theta^2) = \sigma_{\bar{x}}^{\ 2} + \theta^2$$

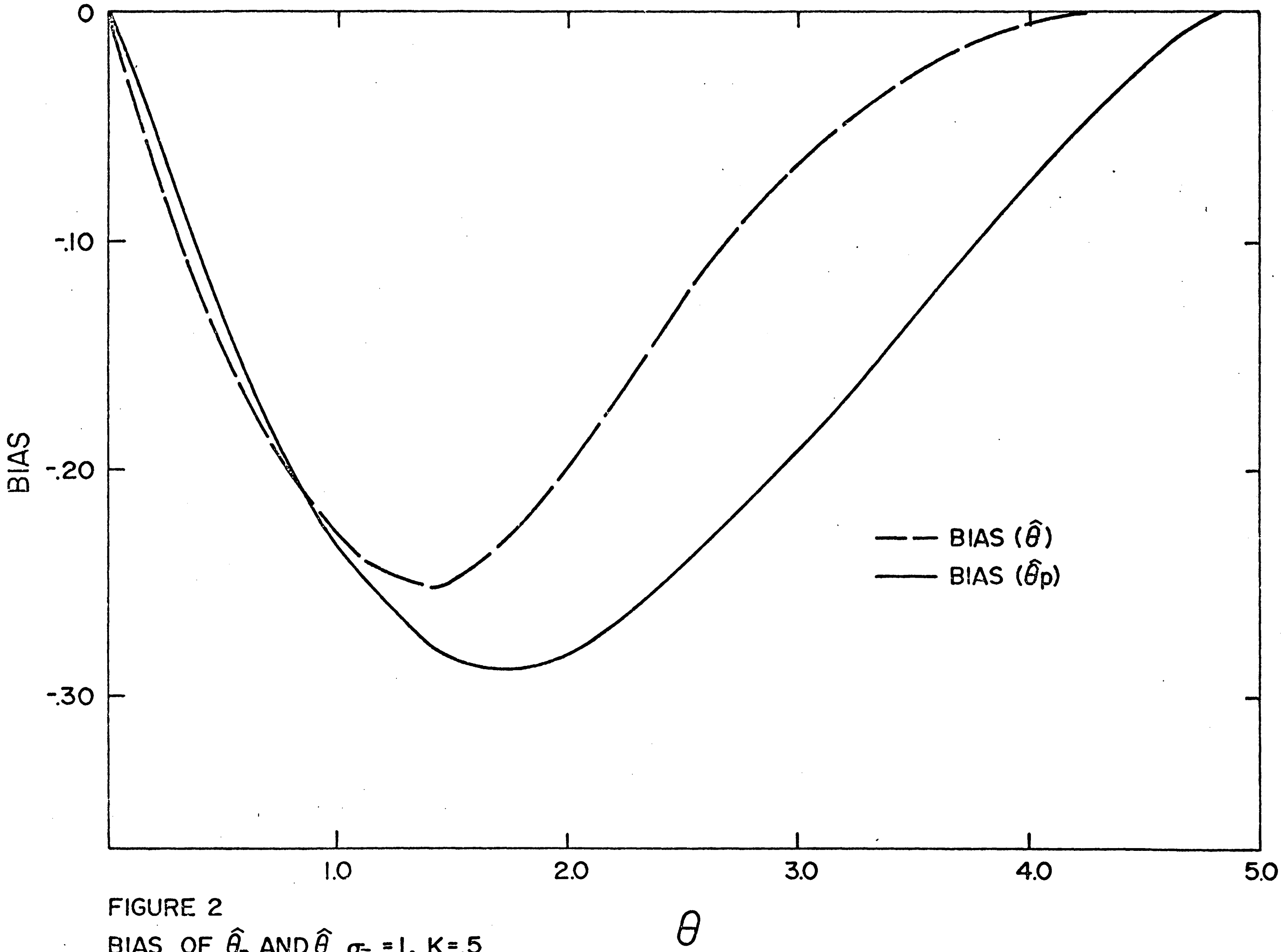

FIGURE 2
BIAS OF $\hat{\theta}_p$ AND $\hat{\theta}$ $\sigma_{\bar{x}} = 1$, K=.5

(b) $$E(2(1-k)\,\sigma_{\bar{x}}^{2}\,z^{2}\,\{N(z+c_b)-N(z+c_a)\})$$

$$= 2(1-k)\,\sigma_{\bar{x}}^{2}\left[N\left(\frac{c_b}{\sqrt{2}}\right) - \frac{c_b\,e^{-c_b^{2}/4}}{4\sqrt{\pi}} - N\left(\frac{c_a}{\sqrt{2}}\right) + \frac{c_a\,e^{-c_a^{2}/4}}{4\sqrt{\pi}}\right]$$

$$= 2(1-k)\,\sigma_{\bar{x}}^{2}\left[N\left(\frac{c_b}{\sqrt{2}}\right) - \frac{c_b}{2\sqrt{2}}\,\eta\left(\frac{c_b}{\sqrt{2}}\right) - N\left(\frac{c_a}{\sqrt{2}}\right) + \frac{c_a}{2\sqrt{2}}\,\eta\left(\frac{c_a}{\sqrt{2}}\right)\right]$$

(c) $$E(4\theta\sigma_{\bar{x}}(1-k)\,z\,\{N(z+c_b)-N(z+c_a)\})$$

$$= 2\sqrt{2}\,\sigma_{\bar{x}}(1-k)\,\theta\left[\eta\left(\frac{c_b}{\sqrt{2}}\right) - \eta\left(\frac{c_a}{\sqrt{2}}\right)\right]$$

(d) $$E(2\theta^{2}\,(1-k)\,\{N(z+c_b)-N(z+c_a)\})$$

$$= 2\theta^{2}\,(1-k)\left[N\left(\frac{c_b}{\sqrt{2}}\right) - N\left(\frac{c_a}{\sqrt{2}}\right)\right] .$$

These expressions account for all of the terms in $E(\hat{\theta}^{2}|\theta)$ except for

(e) $$E[(z^{2}\,\sigma_{\bar{x}}^{2} + 2z\sigma_{\bar{x}}\,\theta + \theta^{2})\,\{N(z+c_b)-N(z+c_a)\}^{2}\,(1-k)^{2}] .$$

Since

$$\theta^2 - 2\theta E(\hat{\theta}|\theta) = -\theta^2 - 2\theta\,(1-k)\left[\frac{\sigma_{\bar{x}}}{\sqrt{2}}\left\{-\,\eta\left(\frac{c_a}{\sqrt{2}}\right) + \eta\left(\frac{c_b}{\sqrt{2}}\right)\right\}\right.$$

$$\left. + \theta\left\{N\left(\frac{c_b}{\sqrt{2}}\right) - N\left(\frac{c_a}{\sqrt{2}}\right)\right\}\right] ,$$

then

$$\mathrm{MSE}(\hat{\theta}|\theta) = \sigma_{\bar{x}}^2 + \frac{\sigma_{\bar{x}}^2\,(1-k)}{\sqrt{2}}\left[c_a\;\eta\left(\frac{c_a}{\sqrt{2}}\right) - c_b\;\eta\left(\frac{c_b}{\sqrt{2}}\right)\right]$$

$$+ 2\sigma_{\bar{x}}^2\,(1-k)\left[N\left(\frac{c_b}{\sqrt{2}}\right) - N\left(\frac{c_a}{\sqrt{2}}\right)\right]$$

$$+ \sqrt{2}\,\sigma_{\bar{x}}\,(1-k)\,\theta\left[\eta\left(\frac{c_b}{\sqrt{2}}\right) - \eta\left(\frac{c_a}{\sqrt{2}}\right)\right] + (e)\;.$$

The (e) term may be evaluated through numerical integration. 10-point Gauss-Hermite quadrature was applied here, using the IBM 360/40 computer at the Virginia Polytechnic Institute.

Figure 3 shows the MSE for $\hat{\theta}$ with $k = .5$, $\sigma_{\bar{x}} = 1$ for $\theta > 0$. As with $\hat{\theta}p$, the points of intersection, a and b, have also been forced to be ± 1 and ± 2.464 and these curves are plotted for $\hat{\theta}$. The curves are symmetric about the MSE-axis.

A comparison of $MSE(\hat{\theta}p|\theta)$ and $MSE(\hat{\theta}|\theta)$ is given in Figure 4 for $k = .5$, $\sigma_{\bar{x}} = 1$, $a = -1.732$ and $b = 1.732$. We see that $\hat{\theta}$ clearly has a smaller MSE than $\hat{\theta}p$ for $|\theta| \leq 2.75$ although the MSE's are very close at $\theta = 0$ where $MSE(\hat{\theta}p) = .543$ while $MSE(\hat{\theta}) = .539$. For $|\theta| > 2.75$, $\hat{\theta}p$ is preferable to $\hat{\theta}$. Notice also that $\hat{\theta}$ has lower MSE than $\bar{x}$ for $|\theta| < 1.5$.

The estimators and their regions of preference are presented here not to dictate which procedure one should use but rather so that an experimenter may decide for himself which estimator, if both are considered applicable, would be preferable for his particular problem.

Nevertheless, we have been assuming that there exists a point estimate of θ, an estimate in which the experimenter has some degree of confidence. Under these circumstances it is unlikely that θ would actually lie in the region in which $\hat{\theta}p$ is preferable to $\hat{\theta}$.

This can be seen from Figure 4. With $\sigma_{\bar{x}} = 1$ and $k = .5$, $\hat{\theta}$ is superior to $\hat{\theta}p$ for θ values within 2.75 standard deviations ($\sigma_{\bar{x}}$) of $\theta = 0$. Requiring the experimenter to make a point "guesstimate" no more than 2.75 $\sigma_{\bar{x}}$ away from the true value of θ does not seem to be an unreasonable demand. If he does not feel he can make an initial estimate meeting this requirement he should just use $\bar{x}$ exclusively anyway. Thus, there seems to be little reason to use $\hat{\theta}p$ in light of the improvement offered by $\hat{\theta}$.

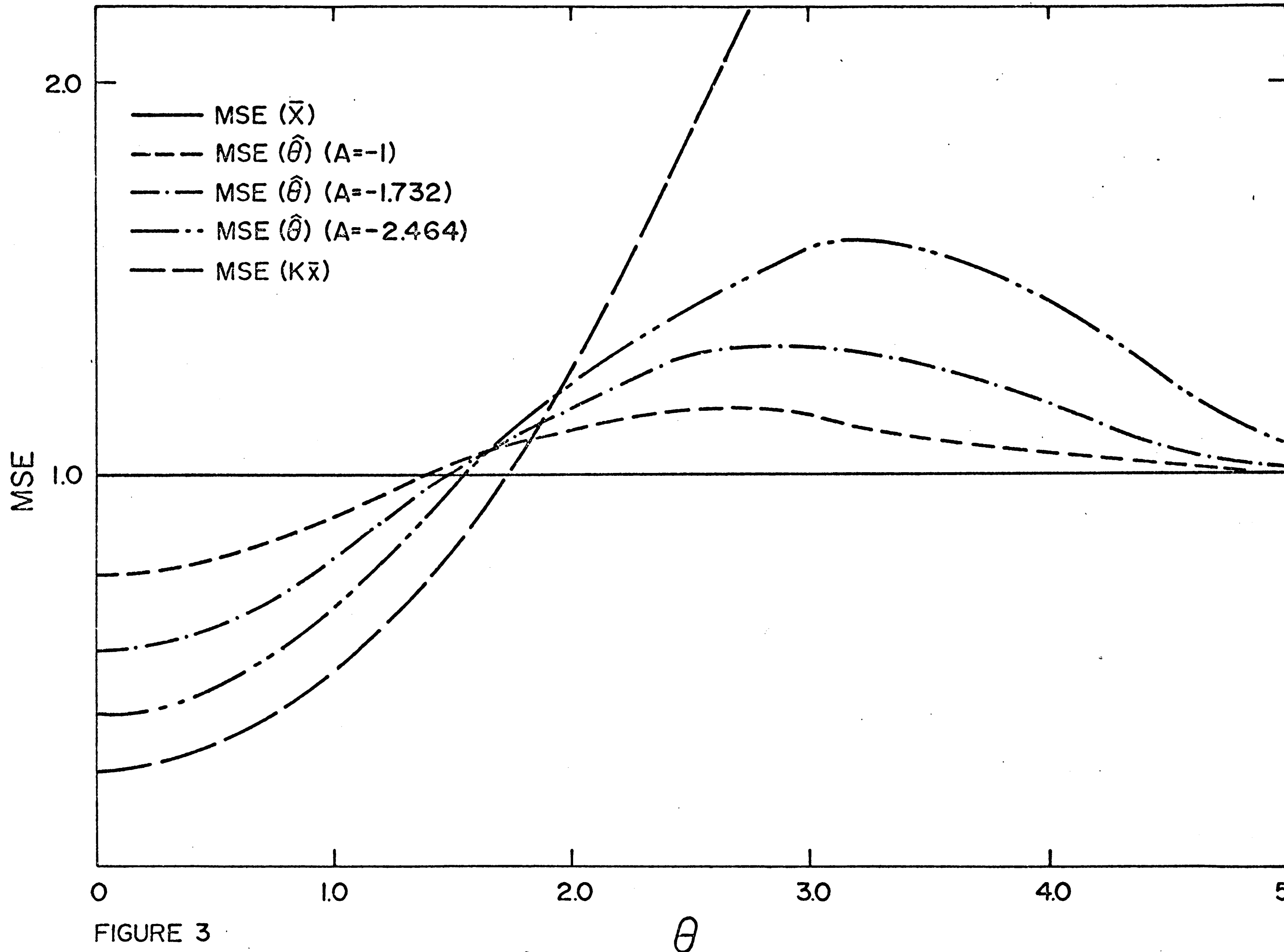

FIGURE 3
MSE OF WEIGHTED ESTIMATOR $\hat{\theta}$ $\sigma_{\bar{x}} = 1, K = .5$

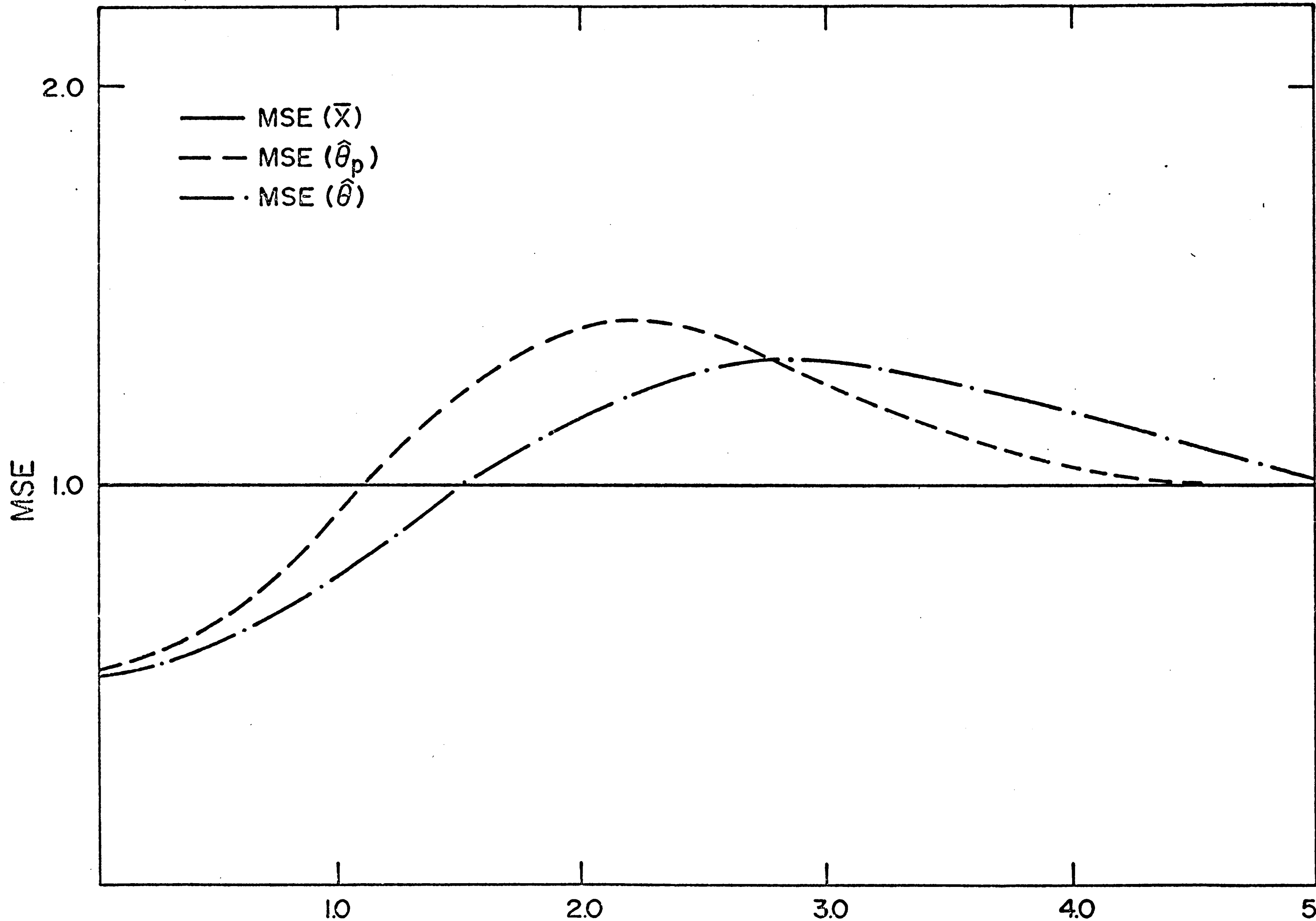

FIGURE 4
MSE OF $\hat{\theta}_p$ AND $\hat{\theta}$ $K=.5,\ \sigma_{\bar{x}}=1,\ A=-1.732$

Comparing $\hat{\theta}$ with $\bar{x}$ indicates that perhaps more accuracy on the original estimate will be required, say within 2.5 $\sigma_{\bar{x}}$ of $\theta = 0$, in order to justify straying away from exclusive use of $\bar{x}$. Essentially we are buying smaller MSE around the origin at the expense of larger MSE farther away. However, $\hat{\theta}$ has the desirable property of not having unbounded risk. Even if the original estimate was far from the true θ there would not be catastrophic results.

To examine the admissibility of $\hat{\theta}$ with respect to $\bar{x}$, we first note that

$$\hat{\theta} = \bar{x}\, g(\bar{x}) + k\bar{x}[1 - g(\bar{x})]$$

$$= h(\bar{x})\, \bar{x}$$

where

$$h(\bar{x}) = g(\bar{x})\ (1 - k) + k\ .$$

Thus, $k \leq h(\bar{x}) \leq 1$.

Next, we let S be such that

$$\bar{x} \in S \quad \text{when} \quad h(\bar{x}) < 1$$

and

$$\bar{x} \in \bar{S} \quad \text{when} \quad h(\bar{x}) = 1\ .$$

We assume that $P(\bar{x} \in S) \neq 0$. Then we examine $MSE(\bar{x}|\theta)$ and $MSE(\hat{\theta}|\theta)$ when $\theta = 0$. Thus

$$MSE(\bar{x}|\theta = 0) = \int_S \bar{x}^2\, f(\bar{x})\, d\bar{x} + \int_{\bar{S}} \bar{x}^2\, f(\bar{x})\, d\bar{x}$$

and

$$MSE(\hat{\theta}|\theta = 0) = \int_S h^2 (\bar{x})\ \bar{x}^2\ f(\bar{x})\ d\bar{x}$$

$$+ \int_{\bar{S}} \bar{x}^2\ f(\bar{x})\ d\bar{x}\ .$$

Since $h^2(\bar{x}) < 1$ on S,

$$MSE(\hat{\theta}|\theta = 0) < MSE(\bar{x}|\theta = 0)\ ,$$

and $\hat{\theta}$ is admissible with respect to $\bar{x}$.

The question of whether or not $\hat{\theta}$ is universally admissible is left unresolved. The methods applied to $\hat{\theta}p$ in Chapter III do not apply to $\hat{\theta}$.

CHAPTER V

A PT ESTIMATOR FOR THE VARIANCE

This chapter introduces the use of a PT estimator for the variance of a normal population whose mean is also unknown. The test makes use of information about the mean to provide an improved estimator for the variance.

Suppose one obtains a sample of size n from a normally distributed variate x whose mean μ and variance σ^2 are unknown. The usual estimator for σ^2 in this case is $s^2 = [\Sigma(x_i - \bar{x})^2/(n-1)]$, popular because it is unbiased. However, Goodman (1953) showed for squared error loss, that if $\theta E(Y|\theta)/E(Y^2|\theta) = A$, where A is known and independent of θ, then among all estimators of θ of the form cY the one which minimized the MSE is AY. For our example, $\theta = \sigma^2$, $Y = s^2$, $E(Y|\theta) = \sigma^2$ and $E(Y^2|\theta) = (n+1)\sigma^4/(n-1)$. Hence, $A = (n-1)/(n+1)$ and $[\Sigma(x_i - \bar{x})^2/(n+1)]$ has uniformly minimum MSE over the class of estimators cs^2.

If we assume that $\mu = \mu_0$ is known, the classical estimator for σ^2 is $w^2 = [\Sigma(x_i - \mu_0)^2/n]$. Again considering Goodman's theorem for $\theta = \sigma^2$, $Y = w^2$, and letting $\mu = \mu_0 = 0$, we have

$$E(Y|\theta) = E\left(\frac{\Sigma x_i^2}{n}\right) = \sum \frac{E(x_i^2)}{n}$$

$$= \sum \frac{[\mathrm{var}\ x_i + E^2(x_i)]}{n} = \sigma^2 ;$$

and

$$E(Y^2|\theta) = E\left(\frac{\Sigma x_i^2}{n}\right) = \frac{1}{n^2} E\left[\sum x_i^4 + 2 \sum_{i \neq j}\sum x_i^2 x_j^2\right] .$$

But

$$E(\Sigma x_i^4) = \Sigma[\mu_4 + 4\mu \mu_3 + 6\mu^2 \sigma^2 + \mu^4] = 3n \sigma^4 ,$$

since for the normal $\mu_4 = 3\sigma^4$ and $\mu_3 = 0$. Also

$$E\left(\sum_{i \neq j}\sum x_i^2 x_j^2\right) = \frac{n(n-1)}{2} \sigma^4 ,$$

so that

$$E(Y^2|\theta) = \frac{3n \sigma^4 + n^2 \sigma^4 - n \sigma^4}{n^2} = \sigma^4 \left(\frac{n+2}{n}\right) .$$

Thus A should equal $n/(n+2)$ and then $\Sigma(x_i - \mu_0)^2/(n + 2)$ has uniformly minimum MSE among estimators of the form cw^2.

The first PT estimator for σ^2 to be considered will be one which chooses between s^2 and $nw^2/(n+2)$. This estimator will be compared with s^2 and $nw^2/(n+2)$, each of which is best in its respective class. Since

$$MSE(s^2) = \frac{2\sigma^4}{n-1}$$

and

$$MSE\left(\frac{nw^2}{n+2}\right) = \frac{2\sigma^4 (n+2) + n^2(\mu-\mu_0)^2}{(n+2)^2} ,$$

we find that

$$\text{MSE}\left(\frac{nw^2}{n+2}\right) < \text{MSE}(s^2) \ , \tag{5.1}$$

when $[(\mu-\mu_0)/\sigma]^4 < 6(n+2)/[n^2(n-1)]$. In other words, given n and σ, when $(\mu-\mu_0)$ is small, $nw^2/(n+2)$ is superior to s^2. This region in which $nw^2/(n+2)$ is superior to s^2 becomes very small for large n. Thus, we consider only small sample sizes.

It was mentioned in Chapter II that Kale (1966) estimated the scale parameter σ in the presence of a nuisance location parameter θ. Here we wish to estimate σ^2 in the presence of the nuisance parameter μ. We will again assume that we have an *a priori* point estimate or guesstimate, μ_0, of the mean μ. Then the preliminary test will be made for this point estimate. The PT estimator for this situation is defined as follows:

$$\hat{\sigma}^2 = \begin{cases} s^2, & \text{if} \quad |t| = \left|\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| > t_{\frac{\alpha}{2},\, n-1} \ \varepsilon \ R \\[2ex] \dfrac{nw^2}{n+2}, & \text{if} \quad |t| < t_{\frac{\alpha}{2},\, n-1} \ \varepsilon \ \bar{R} \ . \end{cases} \tag{5.2}$$

The rationale for this estimator is clear. If a preliminary t-test indicates that the mean is actually close to μ_0, we use the estimator for σ^2 which is better than s^2 when $\mu = \mu_0$. If the t-test rejects $\mu = \mu_0$, we use the same estimator (s^2) that is usually used anyway.

To find the bias and the MSE for $\hat{\sigma}^2$ we must take expectations over the joint distribution of $\bar{x}$ and s^2, noting also that

$$\frac{n}{n+2} w^2 = \left(\frac{n-1}{n+2}\right)\left[s^2 + \frac{n(\bar{x}-\mu_0)^2}{n-1}\right].$$

Then

$$(5.3)\quad E(\hat{\sigma}^2|\sigma^2) = \int_{s^2=0}^{\infty} \int_{\bar{x}\epsilon\bar{R}} \left(\frac{n-1}{n+2}\right)\left[s^2 + \frac{n(\bar{x}-\mu_0)^2}{n-1}\right] g(s^2)\, f(\bar{x})\, d\bar{x}\, ds^2$$

$$+ \int_0^{\infty} \int_{\bar{x}\epsilon R} s^2\, g(s^2)\, f(\bar{x})\, d\bar{x}\, ds^2$$

$$= \int_0^{\infty} g(s^2) \int_{\bar{x} = \mu_0 - t_{\alpha/2}\, s/\sqrt{n}}^{\bar{x} = \mu_0 + t_{\alpha/2}\, s/\sqrt{n}} f(\bar{x})\, d\bar{x}\, ds^2$$

$$+ \int_0^{\infty} s^2\, g(s^2) \int_{\bar{x} = -\infty}^{\bar{x} = \mu_0 - t_{\alpha/2}\, s/\sqrt{n}} f(\bar{x})\, d\bar{x}\, ds^2$$

$$+ \int_0^{\infty} s^2\, g(s^2) \int_{\bar{x} = \mu_0 + t_{\alpha/2}\, s/\sqrt{n}}^{\bar{x} = \infty} f(\bar{x})\, d\bar{x}\, ds^2 .$$

Since

$$f(\bar{x}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\bar{x}}} e^{-\frac{(\bar{x}-\mu)^2}{2\sigma_{\bar{x}}^2}},$$

we let $z = (\bar{x}-\mu)/\sigma_{\bar{x}}$,

$$c_2(s) = (\mu_0 - \mu + t_{\alpha/2}\, s/\sqrt{n})/\sigma_{\bar{x}} ,$$

and (5.4)

$$c_1(s) = (\mu_0 - \mu_0 - t_{\alpha/2}\, s/\sqrt{n})/\sigma_{\bar{x}}$$

and then

$$\mathrm{Bias}(\hat{\sigma}^2|\sigma^2) = E(\hat{\sigma}^2|\sigma^2) - \sigma^2$$

$$= \int_0^\infty \Bigg(\left\{\frac{-3s^2}{n+2} + \frac{n}{n+2}\,[\sigma_{\bar{x}}^2 + (\mu-\mu_0)^2]\right\} \{N(c_2(s)) - N(c_1(s))\}$$

$$- \frac{n}{n+2}\,\{2\sigma_{\bar{x}}\,(\mu-\mu_0) + \sigma_{\bar{x}}^2\, c_2(s)\}\, \eta(c_2(s))$$

$$- \frac{n}{n+2}\,\{2\sigma_{\bar{x}}\,(\mu-\mu_0) + \sigma_{\bar{x}}^2\, c_2(s)\}\, \eta(c_2(s))\Bigg)\, g(s^2)\, ds^2 ,$$

where

$$g(s^2) = \frac{(n-1)^{\frac{n-1}{2}} (s^2)^{\frac{n-3}{2}} e^{-\frac{s^2(n-1)}{2\sigma^2}}}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}} \sigma^{n-1}} .$$

A closed form expression for this integral was not found for the general case in which $c_2(s)$ and $c_1(s)$ are of the form $k + cs$ for $k = (\mu_0 - \mu)/\sigma_{\bar{x}}$ and $c = \pm t_{\alpha/2}/\sigma$ where both k and c are constants. However, we can obtain a closed form when $k = 0$; i.e., when $\mu = \mu_0$. This is the case in which our a priori point estimate of the mean is the same as the true mean.

To find a closed form for $\mathrm{Bias}(\hat{\sigma}^2 | \mu = \mu_0)$ and $\mathrm{MSE}(\hat{\sigma}^2 | \mu = \mu_0)$ we will need the following theorem.

Theorem 5.1: If x is a standard normal variate and $s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$, then

(i) $$\int_0^\infty s^p\, n(cs)\, g(s^2)\, ds^2 = \frac{(n-1)^{\frac{n-1}{2}} 2^{p/2} \sigma^p \Gamma\left(\frac{n-1+p}{2}\right)}{\sqrt{2\pi}\, \Gamma\left(\frac{n-1}{2}\right) (n-1+\sigma^2 c^2)^{\frac{n-1+p}{2}}}$$

and

(ii) $$\int_0^\infty s^p\, N(cs)\, g(s^2)\, ds^2 =$$

$$= q(n) \sum_{j=1}^{\left(\frac{n+p-1}{2}\right)} \frac{c\binom{n+p-2}{2j-1}(2j-1)!}{(4n-4)^j\,(n-1+\sigma^2 c^2)^{\frac{n+p}{2}-j}\binom{\frac{n+p-2}{2}}{j}^2 (j!)^2}$$

$+ k$, if $(n+p)$ is odd;

$$= q(n)\left[\sum_{j=1}^{\frac{n+p-2}{2}} \frac{c\binom{n+p-2}{2j-1}(2j-1)!}{(4n-4)^j\,(n-1+\sigma^2 c^2)^{\frac{n+p}{2}-j}\binom{\frac{n+p-2}{2}}{j}^2 (j!)^2}\right.$$

$$\left.+ \frac{(n+p-1)!\,(n+p-2)!}{(4n-4)^{\frac{n+p-2}{2}}\left(\frac{n+p-2}{2}\right)!\,\sigma\sqrt{n-1}} \tan^{-1}\frac{c\sigma}{\sqrt{n-1}}\right]$$

$+ k$, if $(n+p)$ is even;

where c is a constant,

$$q(n) = \frac{(n-1)^{\frac{n-1}{2}} \, 2^{p/2} \, \sigma^{p+1} \, \Gamma\left(\frac{n+p}{2}\right)}{\sqrt{\pi} \, \Gamma\left(\frac{n-1}{2}\right)} ,$$

$$\text{and } k = \frac{1}{2} E(s^p) .$$

Proof:

$$\text{(i)} \quad \int_0^\infty s^p \, n(cs) \, g(s^2) \, ds^2 = h(n) \int_0^\infty e^{-\frac{c^2 s^2}{2}} \, (s^2)^{\frac{n-3+p}{2}} \, s^{-\frac{s^2(n-1)}{2\sigma^2}} \, ds^2$$

where

$$h(n) = \frac{\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}}}{\sqrt{2\pi} \, \Gamma\left(\frac{n-1}{2}\right) \, \sigma^{n-1}} .$$

We let

$$y = \left[\frac{n-1}{2\sigma^2} + \frac{c^2}{2}\right] s^2 ,$$

$$dy = \left[\frac{n-1}{2\sigma^2} + \frac{c^2}{2}\right] ds^2 ,$$

and the integral becomes

$$h(n) \left(\frac{n-1}{2\sigma^2} + \frac{c^2}{2}\right)^{-\left(\frac{n-1+p}{2}\right)} \int_0^\infty e^{-y} \; y^{\frac{n-3+p}{2}} \, dy .$$

This integral is simply

$$\Gamma\left(\frac{n-1+p}{2}\right) .$$

Thus,

$$(5.5) \quad \int_0^\infty s^p \, \eta(cs) \; g(s^2) \; ds^2 = h(n) \left(\frac{2\sigma^2}{n-1+\sigma^2 c^2}\right)^{\frac{n-1+p}{2}} \Gamma\left(\frac{n-1+p}{2}\right)$$

$$= \frac{(n-1)^{\frac{n-1}{2}} \; 2^{p/2} \; \sigma^p \; \Gamma\left(\frac{n-1+p}{2}\right)}{\sqrt{2\pi} \; \Gamma\left(\frac{n-1}{2}\right) \; (n-1 + \sigma^2 \, c^2)^{\frac{n-1+p}{2}}} .$$

(ii) We use the method of differentiation with respect to a parameter. Then

$$f(c) = \int_0^\infty s^p N(cs)\ g(s^2)\ ds^2 ,$$

and

$$f'(c) = \int_0^\infty s^{p+1}\ \eta(cs)\ g(s^2)\ ds^2 .$$

From (5.5) we have

$$f'(c) = h(n)/(n-1 + \sigma^2\ c^2)^{\frac{n+p}{2}} ,$$

where

$$h(n) = \frac{(n-1)^{\frac{n-1}{2}}\ 2^{p/2}\ \sigma^{p+1}\ \Gamma\left(\frac{n+p}{2}\right)}{\sqrt{\pi}\ \Gamma\left(\frac{n-1}{2}\right)} .$$

Then

$$f(c) = h(n) \int_0^\infty \frac{dc}{(n-1 + \sigma^2\ c^2)^{\frac{n+p}{2}}} .$$

Note that the arbitrary lower limit in this integral is 0. Since

$$\int \frac{dx}{(a+bx^2)^{m+1}} = \frac{1}{2ma} \frac{x}{(a+bx^2)^m} + \frac{2m-1}{2ma} \int \frac{dx}{(a+bx^2)^m} ,$$

and

$$\int \frac{dx}{a+bx^2} = \frac{1}{\sqrt{ab}} \tan^{-1} \frac{x\sqrt{ab}}{a} ,$$

we have the desired result (ii). k is obtained in the usual manner, substituting $c = 0$ into both expressions for $f(c)$.

These formulas can be used in conjunction with (5.4) to obtain a closed form for $B(\hat{\sigma}^2 | \mu = \mu_0)$. This function is graphed in Figure 5 for $n = 5$, $\alpha = .2$ and $.5$. Note that the function diverges as σ increases. Also we see that for $\mu = \mu_0$, $|B|$ for a smaller α is uniformly greater than that for a larger α. This is easily explained by the fact that we accept the preliminary test of hypothesis more often with smaller α, thus choosing the biased estimator $nw^2/(n+2)$ more often.

The MSE function for $\hat{\sigma}^2$ is

$$MSE(\hat{\sigma}^2) = E(\hat{\sigma}^2 - \sigma^2)^2$$

$$= \int_0^\infty \int_{\bar{x}\varepsilon\bar{R}} \left\{\left(\frac{n-1}{n+2}\right)\left[s^2 + \frac{n(\bar{x}-\mu_0)^2}{n-1}\right] - \sigma^2\right\}^2 g(s^2)\ f(\bar{x})\ ds^2\ d\bar{x}$$

$$+ \int_0^\infty \int_{\bar{x}\varepsilon R} (s^2 - \sigma^2)^2\ g(s^2)\ f(\bar{x})\ ds^2\ d\bar{x}$$

$$= \int_0^\infty g(s^2) \int_{\bar{x}\varepsilon\bar{R}} \left[\left\{\left(\frac{n-1}{n+2}\right)^2 s^4 - 2\left(\frac{n-1}{n+2}\right) \sigma^2\ s^2 + \sigma^4\right\}\right.$$

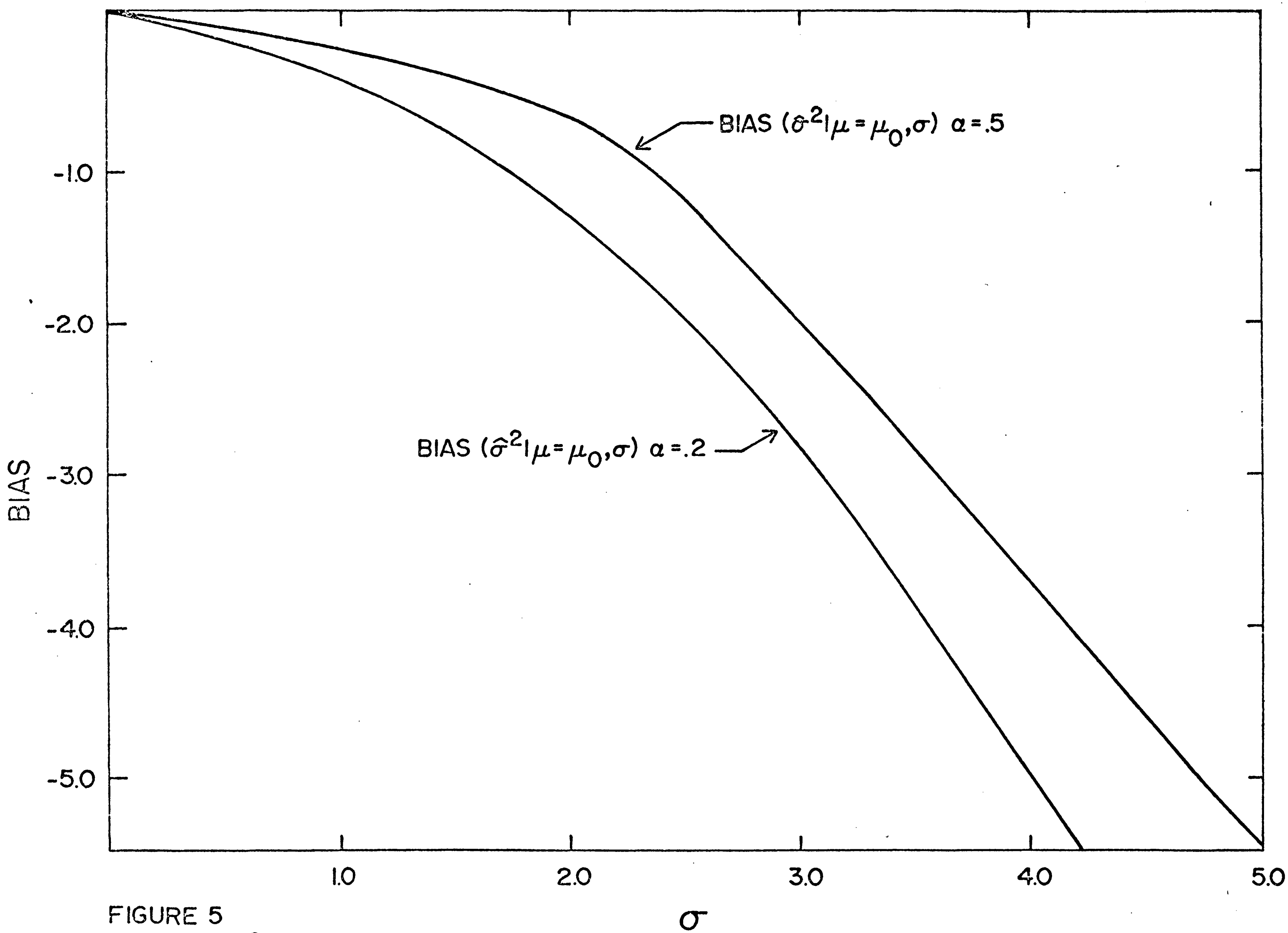

FIGURE 5
BIAS OF $\hat{\sigma}^2$ WHEN $\mu = \mu_0$ N=5, $\alpha = .5, .2$

$$+ \left\{\frac{2n(n-1)}{(n+2)^2} s^2 - \frac{2n\sigma^2}{n+2}\right\} (\bar{x} - \mu_0)^2$$

$$\left. + \left(\frac{n}{n+2}\right)^2 (\bar{x} - \mu_0)^4 \right] f(\bar{x})\, d\bar{x}\, ds^2$$

$$+ \int_0^\infty (s^2 - \sigma^2)^2 g(s^2) \int_{\bar{x}\varepsilon R} f(\bar{x})\, d\bar{x}\, ds^2 .$$

We use the relationships

$$\int_{c_1(s)}^{c_2(s)} z^4 \eta(z)\, dz = - c_2^3(s)\, \eta(c_2(s)) + c_1^3(s)\, \eta(c_1(s))$$

$$+ 3[N(c_2(s)) - N(c_1(s)) - c_2(s)\, \eta(c_2(s))$$

$$+ c_1(s)\, \eta(c_1(s))] ,$$

and

$$\int_{c_1(s)}^{c_2(s)} z^3 \eta(z)\, dz = - c_2^2(s)\, \eta(c_2(s)) + c_1^2(s)\, \eta(c_1(s))$$

$$+ 2[- \eta(c_2(s)) + \eta(c_1(s))] ,$$

where $z = (\bar{x} - \mu)/\sigma_{\bar{x}}$ and $c_1(s)$ and $c_2(s)$ are defined at (5.4). Hence,

$$\mathrm{MSE}(\hat{\sigma}^2|\sigma^2) = \int_0^\infty \frac{2}{(n+2)^2} \; [a(s) \; \{N(c_2(s)) - N(c_1(s))\}$$

$$+ \; b(s) \; \eta(c_1(s)) + d(s) \; \eta(c_2(s))$$

$$+ \; (n+2)^2 \; (s^4 - 2s^2 \; \sigma^2 + \sigma^4)] \; g(s^2) \; ds^2 \; ,$$

where

$$a(s) = (-\,6n - 3) \; s^4 + [(6n + 12) \; \sigma^2 + 2n(n-1) \; (\sigma_{\bar{x}}^2 + (\mu - \mu_0)^2)] \; s^2$$

$$- \; 2(n+2) \; \sigma^2 \; (\sigma_{\bar{x}}^2 + (\mu - \mu_0)^2) + 3n^2 \; \sigma_{\bar{x}}^4$$

$$+ \; n^2[6\sigma_{\bar{x}}^2 \; (\mu - \mu_0)^2 + (\mu - \mu_0)^4]$$

$$b(s) = [2n(n-1) \; s^2 - 2n(n+2) \; \sigma^2] \; [2\sigma_{\bar{x}} \; (\mu - \mu_0) + \sigma_{\bar{x}}^2 \; c_1(s)]$$

$$+ \; n^2[\sigma_{\bar{x}}^4 \; c_1^3(s) + 3\sigma_{\bar{x}}^4 \; c_1(s) + 4\sigma_{\bar{x}}^3 \; (\mu - \mu_0) \; c_1^2(s)$$

$$+ \; 8\sigma_{\bar{x}}^3 \; (\mu - \mu_0) + 4\sigma_{\bar{x}} \; (\mu - \mu_0)^3 + 6\sigma_{\bar{x}}^2 \; (\mu - \mu_0)^2 \; c_1(s)] \; ,$$

and

$$d(s) = -\,[2n(n-1) \; s^2 - 2n(n+2) \; \sigma^2] \; [2\sigma_{\bar{x}} \; (\mu - \mu_0) + \sigma_{\bar{x}}^2 \; c_2(s)]$$

$$- \; n^2[\sigma_{\bar{x}}^4 \; c_2^3(s) + 3\sigma_{\bar{x}}^4 \; c_2(s) + 4\sigma_{\bar{x}}^3 \; (\mu - \mu_0) \; c_2^2(s)$$

$$+ \; 8\sigma_{\bar{x}}^3 \; (\mu - \mu_0) + 4\sigma_{\bar{x}} \; (\mu - \mu_0)^3 + 6\sigma_{\bar{x}}^2 \; (\mu - \mu_0)^2 \; c_2(s)] \; .$$

We can use the results of Theorem 5.1 to obtain a closed form for $MSE(\hat{\sigma}^2|\mu = \mu_0)$. This expression is cumbersome but it is easily seen that all terms in it are of the order σ^4. Thus, since

$$MSE(s^2) = \frac{2\sigma^4}{n-1},$$

we have that the relative efficiency of $\hat{\sigma}^2$ to s^2, when $\mu = \mu_0$, is a constant independent of σ. This efficiency, when $\mu = \mu_0$, becomes larger as α becomes smaller, approaching the efficiency of $nw^2/(n+2)$ to s^2 when $\mu = \mu_0$ as α approaches 0. Of course, as α approaches unity, the efficiency at $\mu = \mu_0$ approaches 1.

For $\mu \neq \mu_0$, the $B(\hat{\sigma}^2|\sigma^2)$ and $MSE(\hat{\sigma}^2|\sigma^2)$ functions must be evaluated through numerical integration. 32-point Gauss-Laguerre quadrature was used in most cases. However, for both very large and very small values of σ, Simpson's Rule was used since it provided greater accuracy.

We note that the bias and MSE functions for $\hat{\sigma}^2$ approach those for s^2. Thus, as mentioned earlier, $\hat{\sigma}^2$ is primarily useful as a small sample estimator. We have chosen to investigate $\hat{\sigma}^2$ for n = 3, 5, 7 and 10 and for α = .1, .2, .5 and .8. For all cases efficiency is defined as the ratio of the MSE of some standard estimator to the MSE of the estimator under consideration. The numbers in the tables have been checked for accuracy at $\mu = \mu_0$ and all decimal places submitted are significant.

The case n = 5 is examined in the greatest detail. Tables I, II, III, and IV show the efficiencies of $\hat{\sigma}^2$ relative to s^2 when n = 5 for the above four α levels at selected values of $(\mu - \mu_0)$ and σ. For α = .8, .5, and .2 the efficiencies appear to uniformly increase as α

TABLE I

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .8$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.19	1.19	1.19	1.19	1.19	1.19	1.19
.25	1.09	1.13	1.16	1.17	1.17	1.19	1.19
.50	1.01	1.05	1.10	1.12	1.14	1.16	1.17
.70	1.00	1.02	1.05	1.08	1.10	1.13	1.16
1.00	1.00	1.00	1.02	1.03	1.06	1.10	1.14
1.20	1.00	1.00	1.01	1.02	1.04	1.07	1.12
1.50	1.00	1.00	1.00	1.00	1.02	1.04	1.10
2.00	1.00	1.00	1.00	1.00	1.00	1.02	1.06
3.00	1.00	1.00	1.00	1.00	1.00	1.00	1.02

TABLE II

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .5$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.50	1.50	1.50	1.50	1.50	1.50	1.50
.25	1.31	1.40	1.45	1.47	1.48	1.49	1.50
.50	1.08	1.20	1.32	1.37	1.41	1.45	1.48
.70	1.01	1.09	1.21	1.27	1.34	1.40	1.46
1.00	1.00	1.01	1.09	1.14	1.22	1.32	1.41
1.20	1.00	1.00	1.04	1.09	1.16	1.26	1.37
1.50	1.00	1.00	1.01	1.03	1.09	1.18	1.32
2.00	1.00	1.00	1.00	1.00	1.02	1.09	1.22
3.00	1.00	1.00	1.00	1.00	1.00	1.01	1.09

TABLE III

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .2$

$\mu - \mu_0$	σ = .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.65	1.65	1.65	1.65	1.65	1.65	1.65
.25	1.59	1.64	1.65	1.65	1.65	1.65	1.65
.50	1.32	1.51	1.60	1.62	1.64	1.65	1.65
.70	1.12	1.32	1.51	1.57	1.61	1.64	1.65
1.00	1.01	1.11	1.32	1.43	1.53	1.60	1.64
1.20	1.00	1.04	1.21	1.32	1.45	1.56	1.63
1.50	1.00	1.01	1.09	1.18	1.32	1.48	1.60
2.00	1.00	1.00	1.01	1.05	1.14	1.32	1.53
3.00	1.00	1.00	1.00	1.00	1.01	1.09	1.32

TABLE IV

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 5, \alpha = .1$

	σ						
$\mu - \mu_0$	.5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.66	1.66	1.66	1.66	1.66	1.66	1.66
.25	1.58	1.63	1.65	1.66	1.66	1.66	1.66
.50	1.27	1.48	1.59	1.62	1.64	1.65	1.65
.70	1.05	1.27	1.49	1.55	1.60	1.63	1.65
1.00	.98	1.05	1.27	1.39	1.51	1.59	1.64
1.20	.99	.99	1.14	1.27	1.42	1.54	1.62
1.50	1.00	.98	1.03	1.12	1.27	1.45	1.59
2.00	1.00	1.00	.98	.99	1.08	1.27	1.51
3.00	1.00	1.00	1.00	.99	.98	1.03	1.27

becomes smaller and also to be uniformly greater than or equal to 1, apparently causing s^2 to be inadmissible. For $\alpha = .1$ the efficiencies are slightly higher around $\mu = \mu_0$ than those for $\alpha = .2$, but they stay larger over only a small portion of the parameter space. It appears that somewhere between $\alpha = .2$ and $\alpha = .1$ there is an optimum value of α which uniformly outperforms all larger α values while still making $\hat{\sigma}^2$ uniformly superior to s^2. In the absence of a closed form expression for $MSE(\hat{\sigma}^2|\sigma^2)$ this α value would be difficult to find. Also, the small difference in efficiencies makes it of no practical significance to know the optimal α level.

In Figure 6 the σ vs. $(\mu - \mu_0)$ space is broken down into areas in which each α level performs the best for $n = 5$, while Figure 7 shows the areas in which each α level performs the worst of the four. We note that $\alpha = .2$ is best over a much wider region than is $\alpha = .1$ and that $\alpha = .8$ appears to have a poor performance over a large set of parameter points. Figure 8 shows the relative efficiency contours for $\alpha = .2$. The shaded area represents an efficiency of 1.0.

$\alpha = .2$ also appears to be the overall best significance level for other values of n. Tables V, VI, and VII give the relative efficiencies for $\hat{\sigma}^2$ with $\alpha = .2$ and $n = 3, 7$, and 10. Again $\hat{\sigma}^2$ appears to be uniformly better than s^2 when $\alpha = .2$ for these n.

One consideration in the use of $\hat{\sigma}^2$ is the size of the bias function. We saw in Figure 5 the bias when $\mu = \mu_0$. The bias becomes smaller, of course, as $|\mu - \mu_0|$ becomes larger, since the unbiased estimator s^2 is then used more often. Just how much bias one is willing to tolerate is left up to the experimenter.

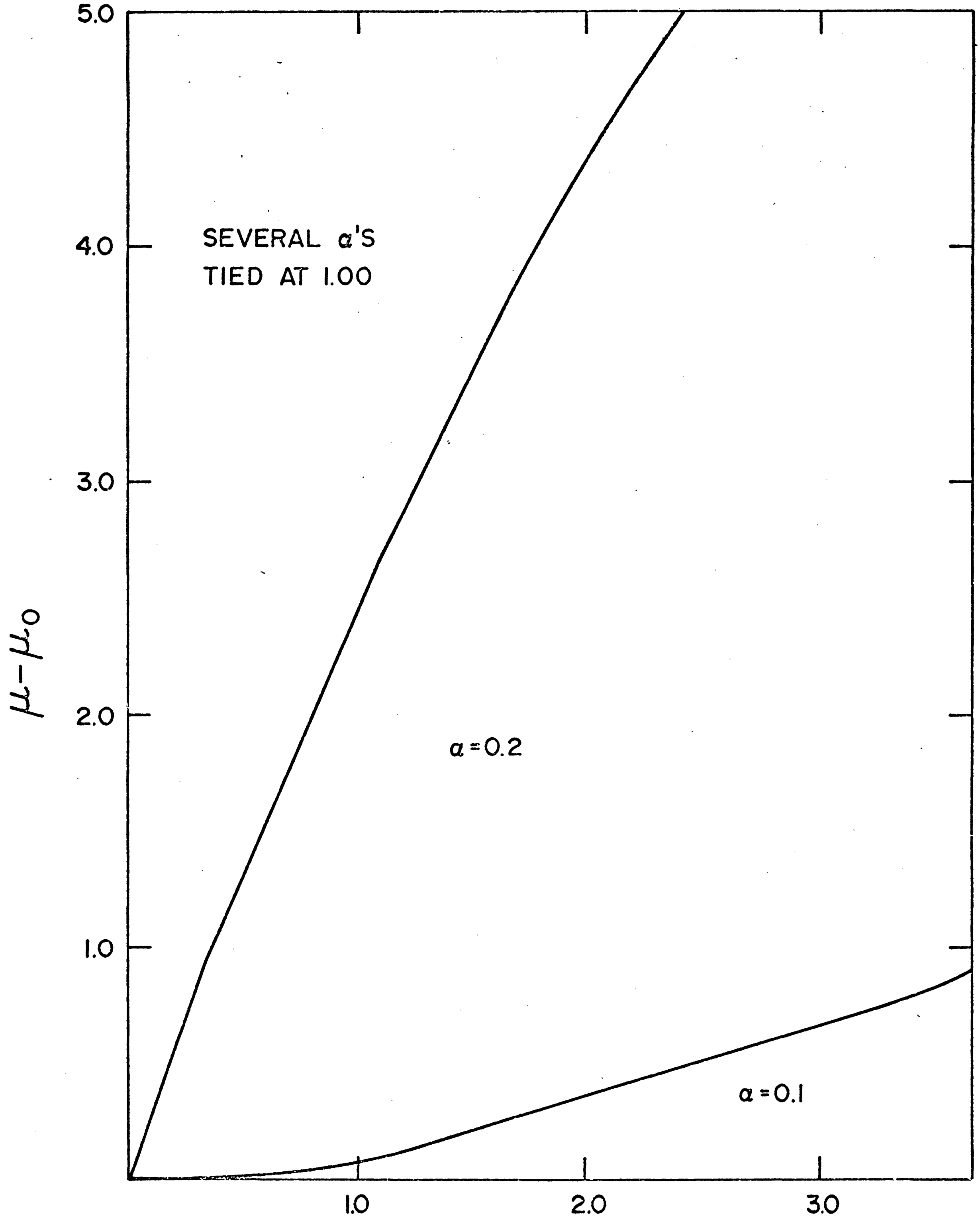

FIGURE 6
BEST α LEVEL FOR $\hat{\sigma}^2$ N=5

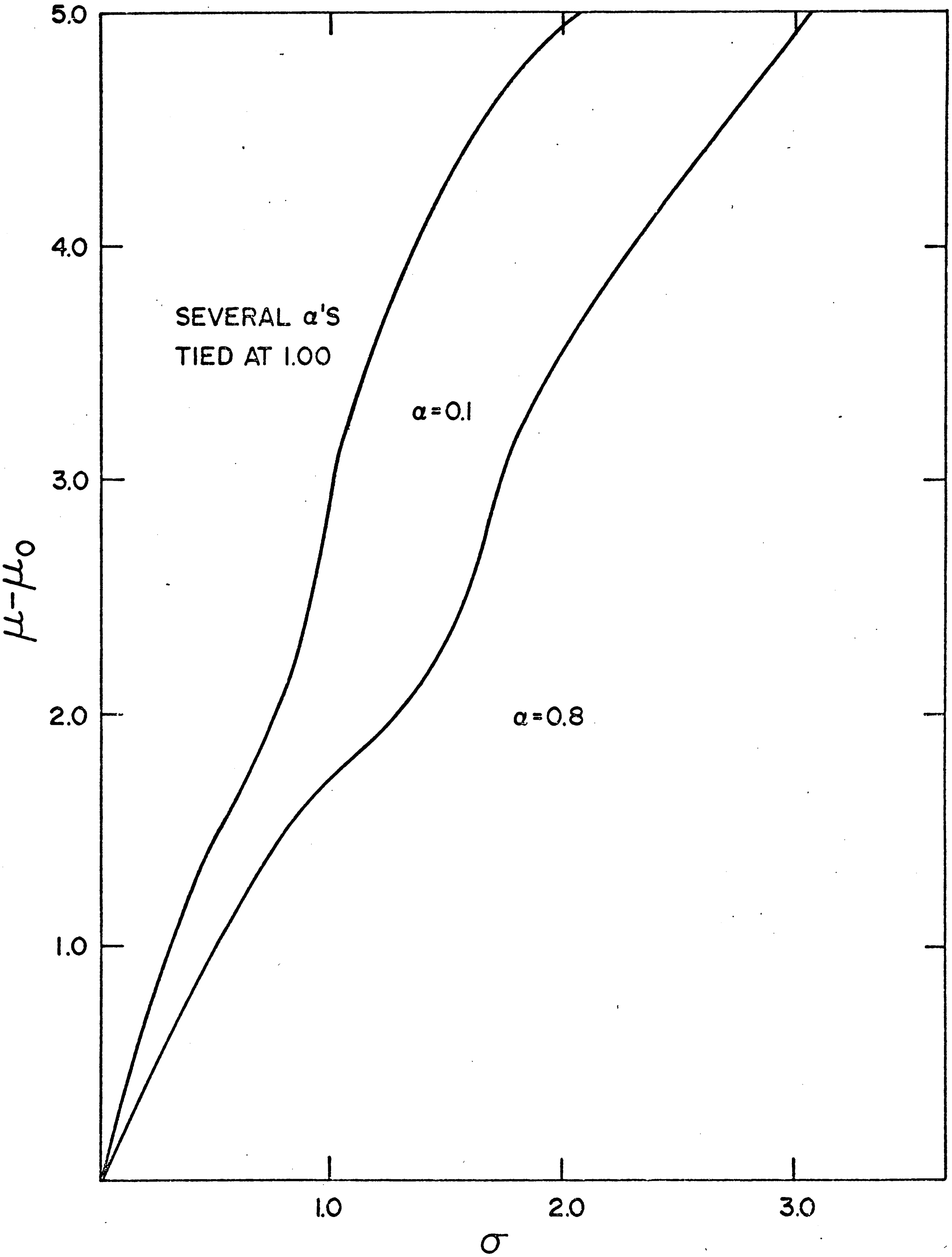

FIGURE 7
POOREST α LEVEL FOR $\hat{\sigma}^2$ N=5

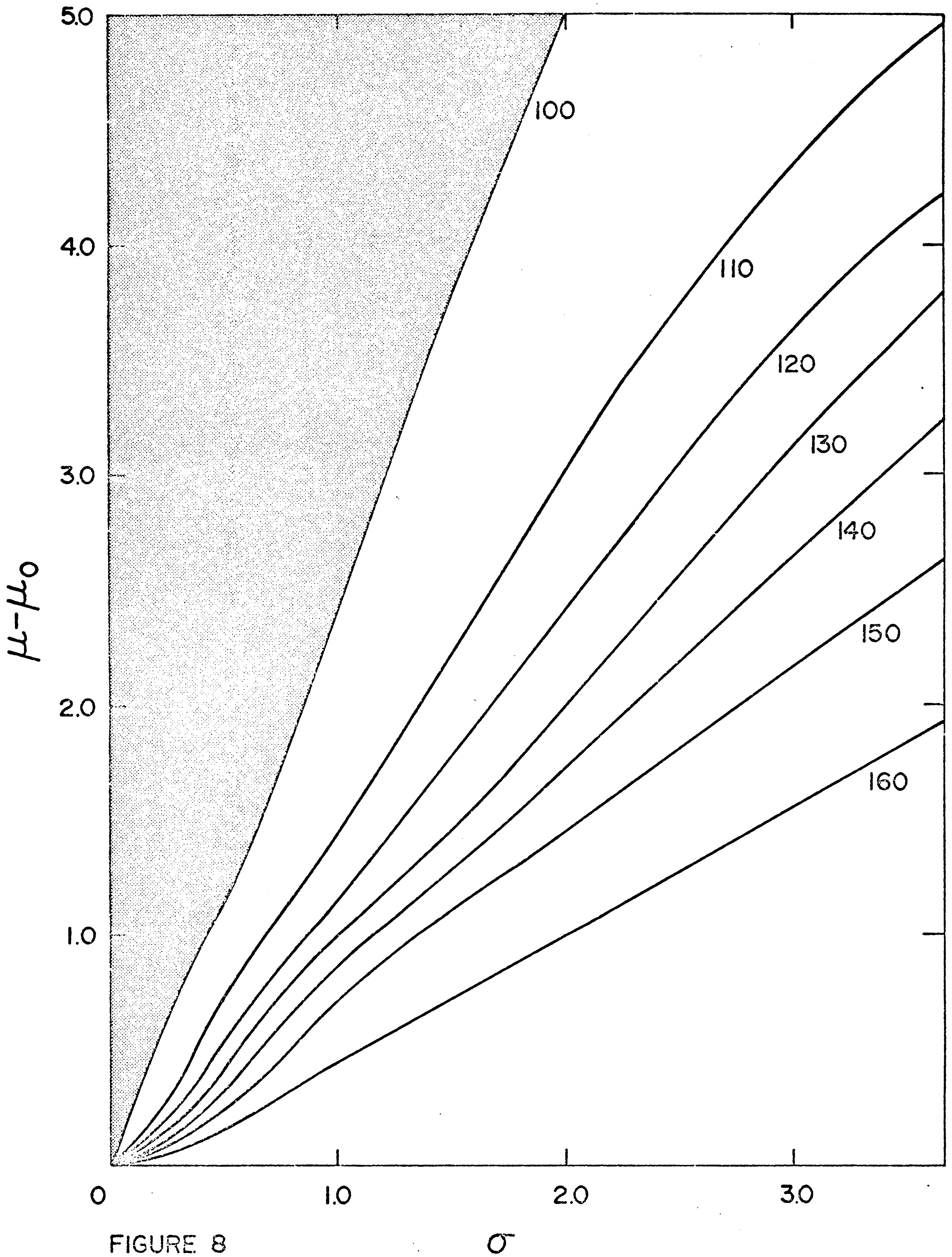

FIGURE 8

PERCENT EFFICIENCY CURVES FOR $\hat{\sigma}^2$ RELATIVE TO S^2

$N = 5$, $\alpha = .2$

SHADED AREA REPRESENTS 100 % EFFICIENCY

TABLE V

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 3$, $\alpha = .2$

$\mu - \mu_0$	σ: .5	.7	1.0	1.2	1.5	2.0	3.0
.00	2.26	2.26	2.26	2.26	2.26	2.26	2.26
.25	2.23	2.25	2.26	2.26	2.26	2.26	2.26
.50	1.95	2.15	2.23	2.24	2.25	2.26	2.26
.70	1.58	1.95	2.16	2.21	2.24	2.25	2.26
1.00	1.19	1.55	1.95	2.08	2.17	2.23	2.25
1.20	1.07	1.34	1.76	1.95	2.10	2.20	2.25
1.50	1.01	1.14	1.50	1.72	1.95	2.13	2.23
2.00	1.00	1.02	1.19	1.37	1.64	1.95	2.17
3.00	1.00	1.00	1.01	1.06	1.19	1.50	1.95

TABLE VI

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 7$, $\alpha = .2$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.44	1.44	1.44	1.44	1.44	1.44	1.44
.25	1.37	1.41	1.43	1.43	1.44	1.44	1.44
.50	1.13	1.27	1.37	1.40	1.42	1.43	1.44
.70	1.02	1.13	1.28	1.34	1.38	1.42	1.43
1.00	1.00	1.02	1.13	1.21	1.29	1.37	1.42
1.20	1.00	1.00	1.06	1.13	1.22	1.33	1.40
1.50	1.00	1.00	1.01	1.05	1.13	1.25	1.37
2.00	1.00	1.00	1.00	1.01	1.03	1.13	1.29
3.00	1.00	1.00	1.00	1.00	1.00	1.01	1.13

TABLE VII

Efficiency of $\hat{\sigma}^2$ Relative to s^2

$N = 10$, $\alpha = .2$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.30	1.30	1.30	1.30	1.30	1.30	1.30
.25	1.21	1.26	1.28	1.29	1.29	1.29	1.30
.50	1.04	1.12	1.21	1.24	1.27	1.28	1.29
.70	1.00	1.04	1.13	1.18	1.23	1.26	1.29
1.00	1.00	1.00	1.04	1.08	1.14	1.21	1.27
1.20	1.00	1.00	1.01	1.04	1.09	1.17	1.25
1.50	1.00	1.00	1.00	1.00	1.04	1.11	1.21
2.00	1.00	1.00	1.00	1.00	1.00	1.04	1.14
3.00	1.00	1.00	1.00	1.00	1.00	1.01	1.05

Since Goodman showed that $(n-1)\ s^2/(n+1)$ has uniformly lower MSE than s^2 we replace s^2 by $(n-1)\ s^2/(n+1)$ in the PT estimator $\hat{\sigma}^2$ to form a modified estimator $\tilde{\sigma}^2$; i.e.,

$$\tilde{\sigma}^2 = \left(\frac{n-1}{n+1}\right) s^2, \quad \text{if} \quad |t| > t_{\alpha/2}\ \varepsilon\ R$$

$$= \frac{nw^2}{n+2}, \quad \text{if} \quad |t| < t_{\alpha/2}\ \varepsilon\ \bar{R}\ .$$

The bias and MSE expressions for $\tilde{\sigma}^2$ can be readily derived. We have

$$E(\tilde{\sigma}^2|\sigma^2) = \int_0^\infty \int_{\bar{x}\varepsilon R} \left(\frac{n-1}{n+1}\right) s^2\ g(s^2)\ f(\bar{x})\ ds^2\ d\bar{x}$$

$$+ \int_0^\infty \int_{\bar{x}\varepsilon \bar{R}} \left(\frac{n}{n+2}\right) w^2\ g(s^2)\ f(\bar{x})\ ds^2\ d\bar{x}$$

$$= \int_0^\infty \int_{\bar{x}\varepsilon R} \left[1 - \frac{2}{n+1}\right] s^2\ g(s^2)\ f(\bar{x})\ ds^2\ d\bar{x}$$

$$+ \int_0^\infty \int_{\bar{x}\varepsilon \bar{R}} \left(\frac{n}{n+2}\right) w^2\ g(s^2)\ f(\bar{x})\ ds^2\ d\bar{x}\ .$$

Thus,

$$B(\tilde{\sigma}^2|\sigma^2) = B(\hat{\sigma}^2|\sigma^2) - \frac{2}{n+1}\int_0^\infty \int_{\bar{x}\in R} s^2\, g(s^2)\, f(\bar{x})\, ds^2\, d\bar{x}\ .$$

This final double integral is the only new term we need to compute in order to evaluate this bias function. The term is always negative. Hence, since the bias of $\hat{\sigma}^2$ is always negative, the bias of $\tilde{\sigma}^2$ is always greater in the negative direction than the bias of $\hat{\sigma}^2$. This would be expected since we are choosing between two negatively biased estimators now.

For the $MSE(\tilde{\sigma}^2|\sigma^2)$ we have

$$MSE(\tilde{\sigma}^2|\sigma^2) = \int_0^\infty \int_{\bar{x}\in R} \left[s^2 \left(\frac{n-1}{n+1}\right) - \sigma^2\right]^2 g(s^2)\, f(\bar{x})\, ds^2\, d\bar{x}$$

$$+ \int_0^\infty \int_{\bar{x}\in \bar{R}} \left[\frac{n}{n+2} w^2 - \sigma^2\right] g(s^2)\, f(\bar{x})\, ds^2\, d\bar{x}$$

$$= MSE(\hat{\sigma}^2|\sigma^2) + \int_0^\infty \int_{\bar{x}\in R} \left[-\frac{4ns^4}{n+1} + \frac{4s^2\ \sigma^2}{n+1}\right] g(s^2)\, f(\bar{x})\, ds^2\, d\bar{x}\ .$$

In this case the last term is neither always positive nor always negative. This means that even though $(n-1)\ s^2/(n+1)$ has uniformly lower MSE than s^2, using $(n-1)\ s^2/(n+1)$ instead of s^2 does not lead to uniformly lower MSE for the PT estimator.

Tables VIII, IX, X, and XI give the efficiencies for $\tilde{\sigma}^2$ with n = 5 and α = .8, .5, .2, and .1. These efficiencies are based on comparing $\tilde{\sigma}^2$ with s^2. Here $\tilde{\sigma}^2$ for all the α levels makes s^2 inadmissible apparently. Also $\tilde{\sigma}^2$ makes $\hat{\sigma}^2$ inadmissible for α = .8 and α = .5. For α's less than some α close to but less than .5, $\hat{\sigma}^2$ becomes admissible with respect to $\tilde{\sigma}^2$.

The efficiency of $(n-1)\ s^2/(n+1)$ with respect to s^2 is $(n+1)/(n-1)$ which for n = 5 is 1.50. We note that for α = .2 and α = .1 the efficiencies of $\tilde{\sigma}^2$ drop below 1.50 over much of the parameter space while α = .8 and α = .5 produce a $\tilde{\sigma}^2$ which makes $(n-1)\ s^2/(n+1)$ inadmissible. Based on Goodman's work it might be improperly argued that one should use $(n-1)\ s^2/(n+1)$ all the time but here is a PT estimator which has uniformly lower MSE than Goodman's estimator.

Just as α = .2 was for $\hat{\sigma}^2$, it appears that α = .5 is the best overall significance level for $\tilde{\sigma}^2$. Table XII shows the efficiency of $\hat{\sigma}^2$ with α = .2 relative to $\tilde{\sigma}^2$ with α = .5. In the absence of bias considerations it would appear that $\tilde{\sigma}^2$ is the better estimator.

However, some observers might want to look at the bias functions for the two estimators. Tables XIII and XIV show the bias for $\hat{\sigma}^2$ (α = .2) and $\tilde{\sigma}^2$ (α = .5) for n = 5. As can be seen from these tables, $\hat{\sigma}^2$ uniformly has a smaller absolute bias than $\tilde{\sigma}^2$. However, the difference becomes smaller for large sample sizes.

If bias is very important to the experimenter he could consider a PT estimator which chooses between the estimators s^2 and w^2. s^2 is unbiased and w^2 is unbiased if $\mu = \mu_0$. The PT estimator for σ^2 for this situation is defined as follows:

TABLE VIII

Efficiency of $\tilde{\sigma}^2$ Relative to s^2

$N = 5,\ \alpha = .8$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.52	1.52	1.52	1.52	1.52	1.52	1.52
.25	1.51	1.51	1.51	1.52	1.52	1.52	1.52
.50	1.50	1.51	1.51	1.51	1.51	1.51	1.52
.70	1.50	1.50	1.51	1.51	1.51	1.51	1.51
1.00	1.50	1.50	1.50	1.51	1.51	1.51	1.51
1.20	1.50	1.50	1.50	1.50	1.50	1.51	1.51
1.50	1.50	1.50	1.50	1.50	1.50	1.51	1.51
2.00	1.50	1.50	1.50	1.50	1.50	1.50	1.51
3.00	1.50	1.50	1.50	1.50	1.50	1.50	1.50

TABLE IX

Efficiency of $\tilde{\sigma}^2$ Relative to s^2

$N = 5,\ \alpha = .5$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.53	1.53	1.53	1.53	1.53	1.53	1.53
.25	1.52	1.53	1.53	1.53	1.53	1.53	1.53
.50	1.51	1.52	1.52	1.53	1.53	1.53	1.53
.70	1.50	1.51	1.52	1.52	1.52	1.53	1.53
1.00	1.50	1.50	1.51	1.52	1.52	1.52	1.53
1.20	1.50	1.50	1.51	1.51	1.52	1.52	1.53
1.50	1.50	1.50	1.50	1.50	1.51	1.52	1.52
2.00	1.50	1.50	1.50	1.50	1.50	1.52	1.52
3.00	1.50	1.50	1.50	1.50	1.50	1.52	1.51

TABLE X

Efficiency of $\tilde{\sigma}^2$ Relative to s^2

$N = 5,\ \alpha = .2$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.54	1.54	1.54	1.54	1.54	1.54	1.54
.25	1.49	1.52	1.53	1.53	1.54	1.54	1.54
.50	1.40	1.45	1.49	1.51	1.52	1.53	1.54
.70	1.42	1.41	1.45	1.48	1.50	1.52	1.53
1.00	1.48	1.42	1.41	1.43	1.46	1.49	1.52
1.20	1.50	1.46	1.41	1.41	1.43	1.47	1.51
1.50	1.50	1.49	1.43	1.41	1.41	1.44	1.49
2.00	1.50	1.50	1.48	1.45	1.41	1.41	1.46
3.00	1.50	1.50	1.50	1.50	1.48	1.43	1.41

TABLE XI

Efficiency of $\tilde{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .1$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.59	1.59	1.59	1.59	1.59	1.59	1.59
.25	1.47	1.53	1.56	1.57	1.58	1.58	1.58
.50	1.17	1.33	1.46	1.50	1.53	1.56	1.58
.70	1.08	1.17	1.34	1.41	1.48	1.53	1.56
1.00	1.27	1.08	1.17	1.26	1.36	1.46	1.53
1.20	1.42	1.15	1.10	1.17	1.28	1.40	1.51
1.50	1.49	1.33	1.09	1.09	1.17	1.31	1.46
2.00	1.50	1.49	1.27	1.13	1.08	1.17	1.36
3.00	1.50	1.50	1.49	1.44	1.27	1.09	1.17

TABLE XII

Efficiency of $\hat{\sigma}^2$ (α = .2) Relative to $\tilde{\sigma}^2$ (α = .5)

N = 5

	σ						
$\mu - \mu_0$	.5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.08	1.08	1.08	1.08	1.08	1.08	1.08
.25	1.05	1.08	1.08	1.08	1.08	1.08	1.08
.50	.87	.99	1.05	1.06	1.07	1.08	1.08
.70	.75	.87	.99	1.03	1.05	1.07	1.08
1.00	.67	.74	.87	.94	1.00	1.05	1.07
1.20	.67	.69	.80	.87	.95	1.03	1.07
1.50	.67	.67	.73	.78	.87	.97	1.05
2.00	.67	.67	.67	.70	.76	.87	.99
3.00	.67	.67	.67	.67	.67	.73	.87

TABLE XIII

Bias of $\hat{\sigma}^2$

$N = 5, \alpha = .2$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	-.08	-.16	-.32	-.46	-.72	-1.28	-2.87
.25	-.06	-.13	-.30	-.44	-.69	-1.25	-2.85
.50	-.02	-.08	-.24	-.37	-.63	-1.18	-2.78
.70	-.01	-.05	-.18	-.31	-.55	-1.10	-2.69
1.00	.00	-.01	-.09	-.20	-.42	- .95	-2.51
1.20	.00	.00	-.05	-.13	-.33	- .83	-2.37
1.50	.00	.00	-.02	-.07	-.21	- .65	-2.13
2.00	.00	.00	.00	-.01	-.08	- .37	-1.68
3.00	.00	.00	.00	.00	.00	- .07	- .84

TABLE XIV

Bias of $\tilde{\sigma}^2$

$N = 5,\ \alpha = .5$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	-.09	-.18	-.38	-.54	-.85	-1.51	-3.39
.25	-.09	-.18	-.37	-.54	-.84	-1.50	-3.38
.50	-.08	-.17	-.36	-.52	-.83	-1.49	-3.37
.70	-.08	-.17	-.35	-.51	-.81	-1.47	-3.35
1.00	-.08	-.16	-.34	-.50	-.79	-1.44	-3.31
1.20	-.08	-.16	-.34	-.49	-.78	-1.42	-3.28
1.50	-.08	-.16	-.33	-.48	-.76	-1.39	-3.23
2.00	-.08	-.16	-.33	-.48	-.75	-1.36	-3.16
3.00	-.08	-.16	-.33	-.48	-.75	-1.34	-3.05

$$\bar{\sigma}^2 = s^2, \quad \text{if} \quad |t| > t_{\alpha/2} \; \varepsilon \; R$$

$$= w^2, \quad \text{if} \quad |t| < t_{\alpha/2} \; \varepsilon \; \bar{R},$$

where t is the same as for $\hat{\sigma}^2$ and $\tilde{\sigma}^2$.

Bias $(\bar{\sigma}^2|\sigma^2)$ and MSE$(\bar{\sigma}^2|\sigma^2)$ are calculated in the same manner as for $\hat{\sigma}^2$ and $\tilde{\sigma}^2$. Tables XV, XVI, XVII, and XVIII show the MSE of $\bar{\sigma}^2$ for $n = 5$ with $\alpha = .8, .5, .2$ and $.1$ while Table XIX gives the bias of $\bar{\sigma}^2$ for $n = 5$ and $\alpha = .5$ since $\alpha = .5$ appears to be the best α level for $\bar{\sigma}^2$. The results are just as one would expect. The bias of $\bar{\sigma}^2$ is lower but the MSE is greater than for $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ with comparable values of n and α.

However, there is some bias present even using $\bar{\sigma}^2$. If the statistician desires an unbiased estimator of the variance, he should use s^2. On the other hand, if he is willing to tolerate bias, then either $\hat{\sigma}^2$ or $\tilde{\sigma}^2$ should be used, depending on which area of the parameter space he feels contains μ and σ.

This chapter has presented and evaluated three alternative estimators of the variance which perform better than the usual estimators. Further investigation of the estimators could be done. For example, some mathematical closed form expression for the MSE's might enable one to find exactly which significance level is best in each case. Also, the problem of confidence intervals on the parameters might be considered. This would involve the use of joint confidence intervals and would necessitate finding the distributional form of the estimators. In addition, there are other combinations of estimators which could be considered in the PT estimation procedure. Perhaps some optimality conditions could be found in terms of bias, risk and sample size.

TABLE XV

Efficiency of $\bar{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .8$

$\mu - \mu_0$	σ: .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.12	1.12	1.12	1.12	1.12	1.12	1.12
.25	1.06	1.09	1.11	1.11	1.12	1.12	1.12
.50	1.01	1.04	1.07	1.08	1.09	1.11	1.12
.75	1.00	1.01	1.04	1.05	1.07	1.09	1.11
1.00	1.00	1.00	1.01	1.02	1.04	1.07	1.09
1.20	1.00	1.00	1.00	1.01	1.03	1.05	1.08
1.50	1.00	1.00	1.00	1.00	1.01	1.03	1.07
2.00	1.00	1.00	1.00	1.00	1.00	1.01	1.04
3.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01

TABLE XVI

Efficiency of $\bar{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .5$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.28	1.28	1.28	1.28	1.28	1.28	1.28
.25	1.17	1.22	1.25	1.26	1.27	1.27	1.28
.50	1.04	1.11	1.17	1.20	1.23	1.25	1.27
.70	1.01	1.04	1.11	1.15	1.19	1.22	1.25
1.00	1.00	1.01	1.04	1.08	1.12	1.17	1.23
1.20	1.00	1.00	1.02	1.04	1.08	1.14	1.21
1.50	1.00	1.00	1.00	1.02	1.04	1.10	1.17
2.00	1.00	1.00	1.00	1.00	1.01	1.04	1.12
3.00	1.00	1.00	1.00	1.00	1.00	1.00	1.05

TABLE XVII

Efficiency of $\bar{\sigma}^2$ Relative to s^2

$N = 5$, $\alpha = .2$

	σ						
$\mu - \mu_0$	.5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.27	1.27	1.27	1.27	1.27	1.27	1.27
.25	1.11	1.18	1.22	1.24	1.25	1.26	1.26
.50	.95	1.02	1.11	1.15	1.19	1.22	1.25
.70	.94	.95	1.03	1.08	1.13	1.18	1.23
1.00	.99	.94	.95	.99	1.04	1.11	1.19
1.20	1.00	.97	.93	.95	1.00	1.07	1.16
1.50	1.00	.99	.95	.93	.95	1.01	1.11
2.00	1.00	1.00	.99	.96	.94	.95	1.04
3.00	1.00	1.00	1.00	1.00	.99	.95	.95

TABLE XVIII

Efficiency of $\bar{\sigma}^2$ Relative to s^2

$N = 5,\ \alpha = .1$

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	1.23	1.23	1.23	1.23	1.23	1.23	1.23
.25	.96	1.08	1.15	1.18	1.20	1.21	1.22
.50	.67	.80	.96	1.03	1.09	1.15	1.20
.70	.62	.67	.81	.90	.99	1.08	1.16
1.00	.79	.62	.66	.73	.84	.96	1.09
1.20	.93	.69	.62	.66	.75	.88	1.04
1.50	.99	.85	.63	.62	.66	.78	.96
2.00	1.00	.99	.79	.67	.62	.66	.84
3.00	1.00	1.00	.99	.95	.79	.63	.66

TABLE XIX

Bias of $\bar{\sigma}^2$

N = 5, α = .5

$\mu - \mu_0$	σ .5	.7	1.0	1.2	1.5	2.0	3.0
.00	-.03	-.05	-.10	-.15	-.23	-.40	-.91
.25	-.02	-.04	-.09	-.13	-.21	-.39	-.89
.50	.00	-.02	-.06	-.10	-.18	-.36	-.86
.75	.00	-.01	-.04	-.07	-.15	-.32	-.81
1.00	.00	.00	-.01	-.04	-.09	-.25	-.73
1.20	.00	.00	-.01	-.02	-.06	-.20	-.66
1.50	.00	.00	.00	-.01	-.03	-.13	-.56
2.00	.00	.00	.00	.00	-.01	-.06	-.38
3.00	.00	.00	.00	.00	.00	.00	-.13

This investigation has shown that, in terms of risk, the usual unbiased estimators should not necessarily be considered. Also we have shown that one can improve upon the unique best variance estimator of the form cs^2. The basic intuitive reason that we were able to improve upon estimators of the form cs^2 is that the PT procedure enables one to utilize information from prior estimates or informed guesstimates.

CHAPTER VI

SUMMARY

In this dissertation three types of estimators have been investigated, all of which may be used as alternatives to the usual unbiased estimators. In general it was shown that these estimators are most valuable when the experimenter has some prior knowledge of the parameter space.

The first investigation was made on $\hat{\theta}p$, a preliminary test estimator for an unknown mean θ with the variance known. $\hat{\theta}p$ is defined by

$$\hat{\theta}p = \bar{x}, \quad \text{if} \quad \bar{x} \in R$$

$$= k\bar{x}, \quad \text{if} \quad \bar{x} \in \bar{R},$$

where $0 < k < 1$, R is defined as the region of θ in which $MSE(\bar{x}|\theta) < MSE(k\bar{x}|\theta)$, and $\bar{R}$ is the complement of R.

A method of calculating the optimal value of k for $\hat{\theta}p$ was given. Also, it was shown that $\hat{\theta}p$ is inadmissible universally but is admissible with respect to $\bar{x}$. In addition, regions in which $\hat{\theta}p$ was preferred to $\bar{x}$ were specified. It was seen that there is a region of the parameter space, however, where $\bar{x}$ had much smaller MSE than $\hat{\theta}p$ and for this reason another estimator of the mean, $\hat{\theta}$, was examined.

$\hat{\theta}$ was defined by

$$\hat{\theta} = \bar{x}\, g(\bar{x}) + k\bar{x}\, [1 - g(\bar{x})] ,$$

where $g(\bar{x})$ is a weighting function incorporating fiducial probabilities. In order to carry out the calculations of the bias and MSE of $\hat{\theta}$ as far as possible, it was necessary to derive expressions for

$$E[N(az)] ,$$

$$E[N(az + c)] ,$$

$$E[z\ N(z + c)] ,$$

and

$$E[z^2\ N(z + c)] ,$$

where z is a standard normal variate, $N(\cdot)$ represents the normal distribution function and a and c are constants.

It was shown that $\hat{\theta}$ has a lower MSE than $\hat{\theta}p$ for values of θ reasonably close to the hypothesized value of θ. It was also determined that $\hat{\theta}$ is admissible with respect to $\bar{x}$.

The concluding estimator was a preliminary test estimator for an unknown variance with the mean also unknown. A preliminary t-test was used for the mean and, if rejected, some function of s^2 was used; if accepted, some function of the crude sum of squares was chosen as the final estimator for the variance. Choices were made between s^2 and w^2, s^2 and $nw^2/(n+2)$ and between $(n-1)\ s^2/(n+1)$ and $nw^2/(n+2)$. For several different significance levels the usual estimator s^2 was made inadmissible.

In conjunction with these variance estimators closed form expressions were found for

$$E[s^p\ \eta(cs)]$$

and

$$E[s^p\ N(cs)] ,$$

where c and p are constants, $\eta(\cdot)$ represents the normal density functions and $s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$.

Recommendations were made for using all the new estimators when a reasonable guesstimate of the parameter's location can be made.

Further investigation of these estimators could be made. First, the question of the admissibility of $\hat{\theta}$ could be looked into. For the variance estimators, the problem of the distribution of the estimators and the resulting confidence interval would be a worthwhile investigation. Also, the question of which significance level in the preliminary test is optimal might be considered.

BIBLIOGRAPHY

Alam, K. and Thompson, J. R. (1968). A class of modified estimates. Unpublished mimeographed report, Indiana University, Bloomington, Indiana.

Arnold, J. C. (1967). Estimation through preliminary tests. Unpublished Ph.D. thesis, Florida State University, Tallahassee, Florida.

Asano, C. (1960a). Estimation after preliminary tests of significance and their applications to biometrical researches. Bull. Math. Statist. 9, 1-24.

________ (1960b). Tests due to pooling data through preliminary tests on biological direct assay. Bull. Math. Statist. 9, 25-39.

________ (1961). Some considerations on the combination of estimates from different biological assays. Bull. Math. Statist. 10, 17-32.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. Ann. Math. Statist. 15, 190-204.

Bennett, B. M. (1952). Estimation of means on the basis of preliminary tests of significance. Ann. Inst. Statist. Math. 4, 31-43.

________ (1956). On the use of preliminary tests in certain statistical procedures. Ann. Inst. Statist. Math. 8, 49-57.

Cohen, A. (1965). Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution. Ann. Math. Statist. 36, 78-87.

Goodman, L. A. (1953). A simple method for improving some estimators. Ann. Math. Statist. 22, 114-117.

Kale, B. K. (1966). An example in which the preliminary test of significance leads to a uniformly better estimator. Research report, Annual meeting of IMS, New Brunswick, N.J., Sept. 1966.

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

Kitagawa, T. (1963). Estimation after preliminary tests of significance. University of California Publications in Statistics, U.C. Press, Berkeley and Los Angeles, Vol. 3, No. 4, 147-186.

Mosteller, F. (1948). On pooling data. Jour. Amer. Statist. Assoc. 43, 231-242.

Rao, C. R. (1965). Linear Statistical Inference and Its Applications. John Wiley and Sons, New York.

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

Thompson, J. R. (1968). Some shrinkage techniques for estimating the mean. Jour. Amer. Statist. Assoc. 46, 113-124.

Widder, D. V. (1961). Advanced Calculus. Prentice-Hall, Inc., Englewood Cliffs, N.J.

The vita has been removed from the scanned document

AN INVESTIGATION OF SOME ALTERNATIVE

ESTIMATION PROCEDURES

by

Robert Loyal Davis

Abstract

Estimators are investigated which provide alternatives to the usual unbiased estimators for the mean and variance. It is assumed that the experimental situation is one in which the statistician has at his disposal some information concerning the parameter space. In particular, the experimenter may have reason to believe that the parameter is in a certain region of the parameter space.

The first estimator examined is reasonably close to the natural origin. The admissibility of $\hat{\theta}$ with respect to $\bar{x}$ is also exhibited.

The final estimator examined is a preliminary test estimator for the variance when the mean is also unknown. The estimator chooses between functions of the form cs^2, where s^2 is the usual unbiased estimator, and functions of the form kw^2, where w^2 is a crude sum of squares, as the final estimator of the variance on the basis of a preliminary t-test on the mean. s^2 is apparently made inadmissible by several of these intermediate values of θ, while the $MSE(\hat{\theta}p|\theta)$ and $MSE(\bar{x}|\theta)$ are shown to be approximately the same for relatively large values of θ. $\hat{\theta}p$ is shown to be inadmissible universally but admissible with respect to $\bar{x}$.

In order to improve upon $\hat{\theta}p$ in the areas in which it performs poorly a new estimator, $\hat{\theta}$, is proposed. $\hat{\theta}$ is a weighted estimator

between $\bar{x}$ and $k\bar{x}$ and its weighting functions are fiducial probabilities. The $MSE(\hat{\theta}|\theta)$ function is shown to be smaller than the $MSE(\hat{\theta}p|\theta)$ for values of θ so-called preliminary test estimator for an unknown mean θ with the variance known and is defined by

$$\begin{aligned} \hat{\theta}p &= \bar{x}, \quad \text{if} \quad \bar{x} \in R \\ &= k\bar{x}, \quad \text{if} \quad \bar{x} \in \bar{R}, \end{aligned}$$

where $0 < k < 1$, R is the region of θ in which $MSE(\bar{x}|\theta) < MSE(k\bar{x}|\theta)$ and $\bar{R}$ is the complement of R. Procedures are given for determining optimal values of k. Bias and MSE for $\hat{\theta}p$ are found. The $MSE(\hat{\theta}p|\theta)$ function is found to be smaller than that for $\bar{x}$ around the "natural origin", larger than $MSE(\bar{x}|\theta)$ for estimators as is

$$\left(\frac{n-1}{n+1}\right) s^2$$

which has uniformly minimum MSE among estimators of the form cs^2. Formulas are given for calculation of the MSE in closed form when the mean is actually the same as the hypothesized value.