DEVELOPMENTS IN DECISION-THEORETIC VARIANCE ESTIMATION

Jon M. Maatta
University of Missouri

and

George Casella
Cornell University

BU-977-M

June, 1988
First Revision May, 1989
Final Revision October, 1989

Abstract

This article traces the history of the problem of estimating the variance, σ^2, based on a random sample from a normal distribution with mean μ unknown. Considered are both the point estimation and confidence interval cases. We see that improvement over both usual estimators follows a remarkably parallel development and stemmed from the innovative ideas presented in Stein (1964). We examine developments through the most recent dealing with improved confidence intervals and conditional evaluations of interval estimators.

1. *Introduction*

The chronological history behind the estimation of the multivariate normal mean is well known. Stein's 1955 paper, which demonstrated the existence of an estimator that improved upon the usual maximum likelihood estimator with respect to squared error loss, was followed by the famous 1961 paper of James and Stein. Each subsequent step, and there are many, lay out the history for all to see. The list of participants in this history reads like a who's - who in modern 20*th* century statistics. Much less well known, though not any less illustrious in its participants, is the chronological history surrounding the estimation of the normal variance. In many ways, the two histories parallel each other. For example, both start with the innovative ideas of Charles Stein. Both continue to include not only improvements with regard to point estimation, but also to encompass improvements in set estimation. Both continue to include statements about conditional confidence properties based on the ideas originally put forth by Fisher (1956a) and later expanded by Buehler (1959) and Robinson (1979a).

In this paper, we trace the history of this estimation problem starting with Stein's (1964) elegant proof of the inadmissibility of the "usual" estimator of variance. Later results flow from Stein's result in a natural sequence. First, Brown (1968) then Brewster and Zidek (1974) improved upon Stein's result for point estimation. (All three papers include results that are more general than will be discussed here.) Using Brown's result, Cohen (1972) constructed intervals for variance that were the same length as the "usual" interval (based on the minimum length interval) but with uniformly higher coverage probability. About ten years later, Shorrock (1982) used Brewster and Zidek's result to construct an interval that improved upon Cohen's. More recently, adapting Shorrock's techniques, Goutis (1989) has produced an interval estimator that is better than that of Shorrock, and improves on both coverage probability and length over the usual intervals. The conditional confidence properties of the usual intervals and the improved versions of Cohen and Shorrock were investigated in papers by Maatta and Casella (1987) and Casella and Maatta (1987).

A brief review of general notation is needed before we proceed. Let

$X = (X_1, X_2, \cdots, X_n)$ be iid random variables from a normal distribution with mean, μ, and variance, σ^2, both unknown, i.e.,

$$X_1, X_2, \cdots, X_n \sim \text{iid } N(\mu, \sigma^2). \tag{1.1}$$

The problem of interest is the estimation of σ^2. Let $\bar{X} = \Sigma X_i/n$ and $S^2 = \Sigma(X_i - \bar{X})^2$, the usual mean and sum of squared deviations about the mean, respectively, calculated from the random sample. Define

$$Z = \sqrt{n}\, \bar{X}/S . \tag{1.2}$$

With our normality assumption, we have the following sampling distributions:

$$S^2/\sigma^2 \sim \chi^2_\nu \quad , \nu = n\text{-}1 \tag{1.3}$$

$$n\bar{X}^2/\sigma^2 \sim \chi^2_1(\tfrac{1}{2}\, \eta^2) \quad , \qquad \eta = \mu/\sigma , \tag{1.4}$$

a central and noncentral chi squared distribution, respectively, the former with ν degrees of freedom and the latter with 1 degree of freedom and noncentrality parameter $\frac{1}{2}\eta^2$. The density of Z, $f_Z(z|\eta)$, depends on μ and σ only through the parameter η.

In the point estimation case, we will be considering the performance of estimators, δ, with respect to the loss function

$$L(\delta, \sigma^2) = \frac{1}{\sigma^4}(\delta - \sigma^2)^2 , \tag{1.5}$$

which is a scaled version of squared error loss. In terms of admissibility considerations, it is equivalent to looking at performance with respect to squared error loss. Also, under this loss function, the decision problem is invariant under affine transformations.

2. *Point Estimation*

At the time of Stein's 1964 paper, two basic results concerning the point estimation of the normal variance were known. First, if a sample is drawn from a normal distribution with *known* mean, μ, and unknown variance, σ^2, the estimator

$$\delta_0(X_1,\cdots,X_n) = \Sigma(X_i-\mu)^2/(\nu+3)\ , \tag{2.1}$$

where $\nu = n\text{-}1$, is admissible for estimating σ^2 with squared error loss. (See Hodges and Lehmann (1951) or Girshick and Savage (1951).) The second result was that, in the unknown mean case, the minimizing value of c for estimators of the form cS^2 (with respect to squared error loss) was $c = \frac{1}{\nu+2}$. That is, the estimator,

$$\delta_1(X_1,\cdots,X_n) = \Sigma(X_i-\bar{X})^2/(\nu+2) = S^2/(\nu+2) \tag{2.2}$$

minimizes risk with respect to squared error loss. Furthermore, $\delta_1(X)$ is the best among all scale and translation invariant estimators and is minimax with respect to the loss (1.5).

For Stein, interest lie in considering a larger class of estimators, those that were scale invariant. These estimators are of the form

$$\delta(\bar{X},S^2) = \phi(Z)S^2 \tag{2.3}$$

where Z is defined by (1.2), and ϕ is a real-valued function. Note that this estimator depends on both $\bar{X}$ and S^2 while those that are a constant function of S^2 ((2.2) for example) depend on $\bar{X}$ only through its appearance in S^2. This represents the first time that it was thought that an estimator for a mean could be used to "improve" the estimator of a variance. Stein showed that an estimator of the form (2.3) dominates the usual estimator (2.2) with respect to loss (1.5).

Theorem 2.1 (*Stein*): With assumptions (1.1), define

$$\phi_S(Z) = \min\left(\frac{1}{\nu+2}\, ,\, \frac{1+Z^2}{\nu+3}\right), \tag{2.4}$$

where Z is defined by (1.2). Then the estimator $\phi_S(Z)S^2$ dominates $\frac{1}{\nu+2}\, S^2$ with respect to the loss (1.5).

Stein's proof is magnificent and will be repeated in detail since Brown's and Brewster and Zidek's results flow naturally from it. Within the proof, two innovative ideas lead to the result and these ideas are germane to the subsequent results of Brown and Brewster and

Zidek.

Proof: Consider the risk of the estimator (2.3) with respect to the loss (1.5),

$$
\begin{aligned}
\mathrm{E}\,\frac{\left(\phi(Z)S^2 - \sigma^2\right)^2}{\sigma^4} &= \mathrm{E}\left(\phi(Z)\left(\frac{S^2}{\sigma^2}\right) - 1\right)^2 \\
(2.5) \qquad &= \mathrm{E}\left(\phi^2(Z)\left(\frac{S^2}{\sigma^2}\right)^2 - 2\phi(Z)\left(\frac{S^2}{\sigma^2}\right) + 1\right) .
\end{aligned}
$$

By iterating the expectation, (2.5) is equal to

$$
\begin{aligned}
&\mathrm{E}_Z\left\{\mathrm{E}\left[\phi^2(Z)\left(\frac{S^2}{\sigma^2}\right)^2 - 2\phi(Z)\left(\frac{S^2}{\sigma^2}\right) + 1\right]\middle| Z\right\} \\
(2.6) \qquad &= \mathrm{E}_Z\left\{\phi^2(Z)\mathrm{E}\left(\left(\frac{S^2}{\sigma^2}\right)^2 \middle| Z\right) - 2\phi(Z)\mathrm{E}\left(\frac{S^2}{\sigma^2} \middle| Z\right) + 1\right\} .
\end{aligned}
$$

Consider the term inside the curly brackets of (2.6). Stein noted that, for fixed μ, σ^2, and for each Z, this is a quadratic in $\phi(\cdot)$ with minimum at

$$
(2.7) \qquad \phi^*(Z) = \frac{\mathrm{E}\left(S^2/\sigma^2 \,|\, Z\right)}{\mathrm{E}\left((S^2/\sigma^2)^2 \,|\, Z\right)} ,
$$

a function only of $|Z|$ and $|\eta| = |\mu|/\sigma$ (See Figure 1).

FIGURE 1 ABOUT HERE

Stein then showed that $\max_{\mu,\sigma^2}\phi^*(Z)$ is attained at $\mu = 0$, $\sigma^2 = 1$, which eliminated the need to work with non-central chi squareds. Now, straightforward calculations yield

$$
\begin{aligned}
\phi_{0,1}(Z) &= \frac{\mathrm{E}\left(S^2/\sigma^2 \mid Z\right)}{\mathrm{E}\left((S^2/\sigma^2)^2 \mid Z\right)} \qquad \text{with } \mu = 0,\ \sigma^2 = 1 \\
(2.8) \qquad &= \frac{\mathrm{E}\left(\chi^2_\nu \,|\, \chi^2_1/\chi^2_\nu = Z^2\right)}{\mathrm{E}\left((\chi^2_\nu)^2 \mid \chi^2_1/\chi^2_\nu = Z^2\right)} = \frac{1+Z^2}{\nu+3} ,
\end{aligned}
$$

using the fact that χ^2_1 and χ^2_ν are independent (a consequence of the independence of $\overline{X}$ and

S^2 in the normal case).

If $Z^2 < \frac{1}{\nu + 2}$, then $(1 + Z^2)/(\nu + 3) < \frac{1}{\nu + 2}$ which also implies that

$$\phi_{\mu,\sigma}(Z) \leq \phi_{0,1}(Z) \leq \frac{1}{\nu + 2}\ , \qquad \text{for all } \mu,\ \sigma^2 \tag{2.9}$$

that is, $\phi_{0,1}(Z)$ is closer to the minimizing value than $\frac{1}{\nu + 2}$. Define

$$\phi_S(Z) = \min\left[\frac{1 + Z^2}{\nu + 3}\ ,\ \frac{1}{\nu + 2}\right]\ . \tag{2.10}$$

Referring to Fig. 1 it is obvious that for each μ, σ^2, Z,

$$E\left(\left(\phi_S(Z)\left(\frac{S^2}{\sigma^2}\right) - 1\right)^2 \Big| Z\right) \leq E\left[\left(\frac{1}{\nu+2}\left(\frac{S^2}{\sigma^2}\right) - 1\right)^2 \Big| Z\right]$$

so $\phi_S(Z)S^2$ is better than $\frac{1}{\nu+2}S^2$ and the result is proved. □

There are two ideas in this proof that are innovative. In hindsight, these ideas will probably seem almost trivial, but they led to many new developments. The first innovative insight in Stein's proof was the conditioning argument that led to considering the quadratic function of $\phi(Z)$ in (2.6). Treating the risk conditionally (on what might be considered an ancillary statistic) leads to consideration of the quadratic. The second insight is realizing that a relationship like (2.9) is a possibility. That is, describing values of Z for which $\phi_{0,1}(Z)$ is closer to the minimizing value than $\frac{1}{\nu+2}$ is a major breakthrough. This argument can also be traced through subsequent work.

Stein knew that his estimator was not admissible, perhaps speculating that such estimators would be limits of Bayes solutions and hence analytic, as in the one parameter exponential family. (This circumstance, however, is not always the case, even in nice situations with convex losses, as seen in Cohen and Sackrowitz (1970).) However, it appears that Stein's enthusiasm for the problem had waned before the time of publication for he states that he finds "...... it hard to take the problem of estimating σ^2 with quadratic loss seriously.", or he may have continued like his successors.

The intuition behind Stein's result is also quite illuminating. Consider that if

$Z^2 < \frac{1}{\nu+2}$ we have

$$\phi_S(Z)S^2 = \frac{1+Z^2}{\nu+3}\, S^2 = \frac{1+(\nu+1)\bar{X}^2/S^2}{\nu+3}\, S^2 \tag{2.11}$$

$$= \frac{1}{\nu+3}\, \Sigma X_i^2 \ ,$$

a special case of the estimator (2.1). When μ is known to be zero, it is admissible. So if Z^2 is small, this can be interpreted as evidence that μ is equal to zero, and we can "pool" S^2 with $(\nu+1)\bar{X}^2$ to get an extra degree of freedom. In effect, we get an estimator that agrees with the admissible estimator (if $\mu = 0$) when Z^2 is small.

This, in fact, is the rationale Brown (1968) uses to justify his estimator. Brown gives the following argument, which has an empirical Bayes flavor.

> If the estimator, say $\bar{x}$, of the location parameter is small compared with the estimate, say s^2, of the variance then this indicates that the location parameter, μ, is near 0. However, when μ is known to be near 0 and $\bar{x}$ is also near 0, the best estimator of the variance is a smaller multiple of s^2 than it is when μ is unknown. Thus, by this reasoning one should use the usual estimator for σ^2 when $|\bar{x}|/s$ is large and a somewhat smaller multiple when $|\bar{x}|/s$ is small.

Clearly, this is a reasonable justification for considering a new estimator and is a rationale behind Stein's estimator. However, it is not needed to justify Stein's proof, which stands alone on a purely decision-theoretic basis. Brown's results are much wider ranging, going far beyond normality and squared error loss, presenting interesting results about different loss functions and best invariant estimators.

A special case of Brown's paper, concerning the normal variance, is to consider estimators of the form $\phi_B(Z)S^2$, where

$$\phi_B(Z) = \begin{cases} c & \text{if } Z^2 \le r^2 \\ d & \text{if } Z^2 > r^2 \end{cases} \tag{2.12}$$

for constants c, d, and r. For the loss (1.5), the risk of $\phi_B(Z)S^2$ is

$$E\left[\frac{\left(\phi_B(Z)S^2 - \sigma^2\right)^2}{\sigma^4}\right] = E\left[\left(c\left(\frac{S^2}{\sigma^2}\right) - 1\right)^2 \,\middle|\, Z^2 \le r^2\right] P(Z^2 \le r^2) \tag{2.13}$$

$$+ \, E\left[\left(d\left(\frac{S^2}{\sigma^2}\right) - 1\right)^2 \,\middle|\, Z^2 > r^2\right] P(Z^2 > r^2) \quad .$$

Differentiation of (2.13), with respect to c, shows that for each μ, σ^2, the best value of c is given by

$$\text{(2.14)} \qquad c = c_{\mu,\sigma^2}(r^2) = \frac{E\left(S^2/\sigma^2 | Z^2 \leq r^2\right)}{E\left((S^2/\sigma^2)^2 \mid Z^2 \leq r^2\right)} \quad .$$

Brown showed that, for every μ, σ^2, and r^2,

$$\text{(2.15)} \qquad c_{\mu,\sigma^2}(r^2) \leq c_{0,1}(r^2) < \frac{1}{\nu+2} \, ,$$

(Compare (2.15) to (2.9).) The estimator $\phi_B^*(Z^2)S^2$, with

$$\text{(2.16)} \qquad \phi_B^*(Z^2) = \begin{cases} c_{0,1}(r^2) & \text{if } Z^2 \leq r^2 \\ 1/(\nu+2) & \text{if } Z^2 > r^2 \end{cases}$$

uniformly dominates $\frac{1}{\nu+2}S^2$ with respect to squared error loss (since $c_{0,1}(r^2)$ is closer to the minimizing value of c (2.14) than is $\frac{1}{\nu+2}$) .

Minimization of (2.13) results in a best c (2.14) of the same form as Stein's (2.8) and the resulting inequality (2.15) concerning $c_{\mu,\sigma^2}(r^2)$ is similar to (2.9). Thus, the two innovative ideas in Stein's proof are repeated.

Like Stein's estimator, Brown's estimator was also inadmissible (as he knew). In 1974, Brewster and Zidek extended the argument used by Stein and Brown to improve upon Brown's estimator. At the same time that Brewster and Zidek were conducting their research, Strawderman (1974) also exhibited improved estimators of the normal variance using a different technique. He found minimax and generalized Bayes variance estimators using the representation of noncentral chi squared expectations conditionally as central chi squareds (by conditioning on an auxiliary Poisson random variable). Since Strawderman's ultimate results are similar to those of Brewster and Zidek, his techniques will not be described in detail here.

Extending the work of Brown, Brewster and Zidek selected, for fixed r as above, r', with $0 < r' < r$ and considered the estimator $\phi(Z^2)S^2$, where

$$\phi(Z^2) = \begin{cases} c_{0,1}(r'^2) & \text{if } Z^2 \leq r'^2 \\ c_{0,1}(r^2) & \text{if } r'^2 < Z^2 \leq r^2 \\ 1/(\nu+2) & \text{if } r^2 < Z^2 \ . \end{cases} \tag{2.17}$$

By noticing that $c_{0,1}(r'^2) \leq c_{0,1}(r^2)$ and by repeating the arguments of Stein and Brown, they conclude that (2.17) yields a better estimator than (2.16). Obviously, this process can continue: take r'', with $0 < r'' < r' < r$ and construct a corresponding new $\phi(Z^2)$ function. In this way, they built new estimators that are each better than the previous. Again, all of these resulting estimators are inadmissible.

However, Brewster and Zidek had another innovative idea. For $i = 1,2,\cdots m_i$, they showed that they could select a finite partition, R_i, where

$$0 = r_{i,0} < r_{i,1} < \cdots < r_{i,m_{i-1}} < r_{i,m_i} = \infty \tag{2.18}$$

and define

$$\phi^{(i)}(R_i) = c_{0,1}(r_{ij}^2) \qquad \text{for} \quad r_{ij-1}^2 < R_i \leq r_{ij}^2 \qquad j = 1,\cdots,m_i \tag{2.19}$$

with $c_{0,1}(\infty) = \frac{1}{\nu+2}$. Furthermore, letting

$$\lim_{i\to\infty} r_{i,m_{i-1}} = \infty \ , \quad \text{and} \quad \lim_{i\to\infty} \max_{1\leq j\leq m_i-1} |r_{ij} - r_{i,j-1}| = 0 \ ,$$

results in

$$\lim_{i\to\infty} \phi^{(i)}(R_i) = \phi^*(Z^2)$$

where

$$\phi^*(z^2) = c_{0,1}(z^2) = \frac{E\left(S^2/\sigma^2 | Z^2 \leq z^2\right)}{E\left((S^2/\sigma^2)^2 | Z^2 \leq z^2\right)} \ , \qquad \mu = 0, \ \sigma^2 = 1. \tag{2.20}$$

Clearly, by construction, for $i' > i$, $\phi^{(i')}(R_{i'})S^2$ is better than $\phi^{(i)}(R_i)S^2$, i.e.,

$$E\left(\phi^{(i')}(R_{i'})S^2 - \sigma^2\right)^2 \leq E\left(\phi^{(i)}(R_i)S^2 - \sigma^2\right)^2$$

and for all $i = 1,2,\cdots$,

$$E\left(\phi^*(Z^2)S^2 - \sigma^2\right)^2 \leq E\left(\phi^{(i)}(R_i)S^2 - \sigma^2\right)^2 ,$$

showing that $\phi^*(Z^2)S^2$ is superior to any of the estimators based on a finite partition. The intuitive appeal of the estimator is obvious, repeated application of a process that improved an estimator would continue to improve the estimator in the limit. Brewster and Zidek go on to show that $\phi^*(Z^2)S^2$ is generalized Bayes, and is admissible in the class of scale-equivariant procedures. Results by Proskin (1985) have shown that $\phi^*(Z^2)S^2$ is admissible among all estimators of σ^2 (using the loss (1.5).)

It is interesting to note Brewster and Zidek's argument for showing that $\phi^*(Z^2)S^2$ is generalized Bayes, since Shorrock (1990) uses the same argument to show that his improved interval is also generalized Bayes. Brewster and Zidek argue that

$$\begin{aligned}(2.21)\qquad \frac{1}{\sigma^4}E\left[\left(\phi(Z^2)S^2 - \sigma^2\right)^2 |Z\right] &= E\left[\left(\phi(Z^2)(S^2/\sigma^2) - 1\right)^2 |Z\right] \\ &= \int\left(\phi(z^2)t - 1\right)^2 f(z^2,t|\eta)dt ,\end{aligned}$$

where $f(z^2,t|\eta)$ is a function of μ and σ^2 only through $\eta = \mu/\sigma$. Using a (possibly improper) prior on η, $\pi(\eta)$, a (possibly generalized) Bayes estimator against this prior will minimize, for each z^2, the posterior loss

$$(2.22)\qquad \int_0^\infty (\phi(z^2)t - 1)^2\, g_\pi(z^2,t)dt,$$

where

$$g_\pi(z^2,t) = \int_{-\infty}^{\infty} f(z^2,t|\eta)\pi(\eta)d\eta .$$

Expression (2.22) is minimized by taking $\phi(z^2)$ equal to

$$\phi^\pi(z^2) = \frac{\int_{-\infty}^{\infty} t g_\pi(z^2,t)dt}{\int_{-\infty}^{\infty} t^2 g_\pi(z^2,t)dt} ,$$

and $\phi^\pi(z^2) = \phi^*(z^2)$, giving the Brewster-Zidek estimator if

(2.23) $$\pi(\eta) = \int_0^\infty u^{-1/2}(1+u)^{-1} e^{-un\eta^2/2} du \quad .$$

If we return to the original setting of the problem (recall $\eta = \mu/\sigma$), then the Brewster-Zidek estimator is the posterior mean starting from the prior $\sigma^{-1}\pi(\mu/\sigma)d\mu\, d\sigma/\sigma$, where $\pi(\cdot)$ is given by (2.23).

While the estimator of Brewster and Zidek is an admissible estimator of σ^2 with respect to loss (1.5), it wasn't until Rukhin (1987) that the relative improvement was investigated. Rukhin also considers locally optimal minimax shrinkage estimators and observes that Brewster and Zidek's estimator has a risk function that is very close to the locally optimal estimators. However, the maximum relative improvement of Rukhin's estimators, over the usual estimator (2.2), is only 4%, suggesting that there would be very little practical benefit associated with these improved estimators. This, perhaps, confirms Stein's original intuition concerning estimation of σ^2 with respect to squared error loss. However, as we shall see in Section 5, there are interesting cases where substantial improvement is possible.

3. *Interval Estimation*

The development of improved estimators in the interval case followed directly (about 6-8 years later) from the improvements in the point estimation case. Cohen (1972) improved upon the "usual" estimator using arguments similar to those in Brown (1968). Shorrock (1982, 1990) produces an improved estimator (improved over Cohen's) and went on to produce further improvements using techniques similar to Brewster and Zidek.

As an introduction, we mention the "usual" estimators of σ^2 based on S^2 only. (See Tate and Klett (1959) for a more complete review.) These intervals are of the form

(3.1) $$C(S^2) = \left\{\sigma^2 : aS^2 \le \sigma^2 \le bS^2\right\}$$

where $P\left(1/b \le \chi^2_\nu \le 1/a\right\} = 1 - \alpha$. To uniquely determine a and b, an additional constraint is needed. For example, the most well known interval of the form (3.1) is the equal-tailed interval that has added constraint

$$I_{ET}:\ P\left(\chi^2_\nu \geq 1/a\right) = P\left(\chi^2_\nu \leq 1/b\right) = \alpha/2 \quad . \tag{3.2}$$

At least two other intervals are worthy of mention. The first, the minimum length interval, is found by minimizing the length of (3.1) and is determined by the added constraint

$$I_{ML}:\ f_{\nu+4}(1/a) = f_{\nu+4}(1/b) \tag{3.3}$$

where $f_m(\cdot)$ is the chi squared density function with m degrees of freedom. The second interval, the shortest unbiased, is associated with the inverse of the family of uniformly most powerful unbiased tests of the hypothesis $H_0: \sigma^2 = \sigma_0^2$ vs. $H_1: \sigma^2 \neq \sigma_0^2$ and is determined by the added constraint

$$I_{SU}:\ f_{\nu+2}(1/a) = f_{\nu+2}(1/b) \quad . \tag{3.4}$$

The choice of interval should depend on more than just ease of calculation, which is the only favorable factor associated with I_{ET}. It is generally accepted that length is the overriding criterion when interval estimation is concerned. Thus, we could consider the "best" (1 - α) 100% confidence interval for σ^2 based only on S^2 to be I_{ML}.

Of course, there are arguments in favor of measures of volume other than length. One popular alternative for scale parameters is the ratio of the endpoints. Using this criterion for normal variance intervals shows that the interval with smallest endpoint ratio satisfies (3.4), that is, the shortest unbiased interval also minimizes the ratio of endpoints. Furthermore, the constructions outlined here to improve on I_{ML} can also be applied to I_{SU} to construct intervals with smaller endpoint ratio.

Returning to the construction of shorter intervals, we first note a notationally simpler formulation of I_{ML} is

$$I_{ML}:\left\{\sigma^2 : a_0 S^2 \leq \sigma^2 \leq (a_0 + c_0) S^2\right\} \tag{3.5}$$

with a_0 and c_0 satisfying

$$P\left(\frac{1}{a_0+c_0} \leq \chi^2_\nu \leq \frac{1}{a_0}\right) = 1 - \alpha$$

and c_0 is minimum among all c. By differentiation, the minimum length constraint is satisfied if $f_{\nu+4}(1/a) = f_{\nu+4}(1/(a+c))$. It is this minimum length interval that Cohen (and later Shorrock) uses for the starting point of his improved estimator. Both intervals are improvements over I_{ML} since they will have the same length but uniformly greater probability of coverage.

Cohen's (1972) improvement uses Brown's technique applied to confidence intervals. Starting with I_{ML}, fix $r > 0$, and define

$$I_C(Z^2,S^2) = \begin{cases} \left(\phi_C(r)S^2, \quad (\phi_C(r) + c_0)S^2\right) & \text{if } Z^2 \leq r^2 \\ \left(a_0S^2 \quad , \quad (a_0 + c_0)S^2\right) & \text{if } Z^2 > r^2 \end{cases} . \tag{3.6}$$

Note that if $Z^2 > r^2$, (3.6) is just the usual minimum length interval, and if $Z^2 \leq r^2$, (3.6) still has the same length but no longer agrees with I_{ML}. Cohen proved that $\phi_C(r)$ can be chosen so that $\phi_C(r) < a_0$ and

$$P\left(\sigma^2 \in I_C(Z^2,S^2)\middle|Z^2 \leq r^2\right) > P\left(\sigma^2 \in I_{ML}\middle|Z^2 \leq r^2\right) \qquad \forall\ \mu,\sigma^2 , \tag{3.7}$$

and hence this interval dominates I_{ML} (has greater probability of coverage).

The rationale behind Cohen's interval, (3.6), is similar to Brown's motivation for improvement of the point estimator. When μ is known to be near 0 and $\bar{X}$ is near 0, the best estimator of σ^2 is a smaller multiple of S. So, when $|\bar{X}|/S$ is large, use I_{ML}, but when $|\bar{X}|/S$ is small use a somewhat smaller multiple of S^2. This is what (3.6) does. When $|\bar{X}|/S$ is small, it shifts the interval toward zero while still maintaining the overall minimum length.

Shorrock (1982, 1990) extended the work of Cohen. He showed that for fixed r, the best choice of $\phi_C(r)$ (in the sense of Brown), of minimizing the conditional expectations, is the unique root, ϕ, of

$$f_{\nu+4}\left(\tfrac{1}{\phi}\right)P\left(\chi_1^2 \leq r^2/\phi\right) = f_{\nu+4}\left(\tfrac{1}{\phi+c_0}\right)P\left(\chi_1^2 \leq \frac{r^2}{\phi + c_0}\right) . \tag{3.8}$$

Shorrock denoted this root as $\phi_0(r^2)$ and noted the limiting features of $\phi_0(r^2)$:

$$\text{as } r^2 \to \infty , \quad \phi \to a_0 , \quad \text{the minimum length choice,}$$

and

$$\text{as } r^2 \to \infty\ , \quad \phi \to a_0^{n+1}\ ,$$

the best choice for n+1 observations and length c_0. Shorrock uses these properties of $\phi_0(\cdot)$ for development of his improvements.

The setup is the same as that of Brewster and Zidek. Consider a partition

$$0 = r_{i0} < r_{i1} < \cdots < r_{im_i} = \infty\ ,$$

and define

$$I^{(i)}(R_i, Z^2, S^2) = \left(\phi_0\left(r_{ij}^2\right)S^2,\ \left(\phi_0(r_{ij}^2) + c_0\right)S^2\right) \qquad \text{for } r_{ij-1}^2 < Z^2 \le r_{ij}^2\ . \tag{3.9}$$

By construction, each $I^{(i')}$ dominates $I^{(i)}$ for $i' > i$, and $I^{(0)} = I_C$ dominates I_{ML}. Shorrock proves that as $i \to \infty$, if

$$r_{i,m_{i-1}} \to \infty \qquad \text{and} \qquad \max_j |r_{ij} - r_{i,j-1}| \to 0\ ,$$

then

$$I^{(i)}(R_i, Z^2, S^2) \to I_S(Z^2, S^2),$$

where

$$I_S(Z^2, S^2) = \left(\phi_0(Z^2)S^2,\ \left(\phi_0(Z^2) + c_0\right)S^2\right), \tag{3.10}$$

and, for each $t > 0$, $\phi_0(t)$ is uniquely defined by the relationship

$$f_{\nu+4}\left(\frac{1}{\phi_0(t)}\right)P\left(\chi_1^2 \le \frac{t}{\phi_0(t)}\right) = f_{\nu+4}\left(\frac{1}{\phi_0(t)+c_0}\right)P\left(\chi_1^2 \le \frac{t}{\phi_0(t)+c_0}\right). \tag{3.11}$$

Furthermore,

$$P(\sigma^2 \in I_S) > P(\sigma^2 \in I_{ML}) \qquad \text{for all } \mu, \sigma^2, \tag{3.12}$$

thus I_S has uniformly greater coverage probability than I_{ML} while maintaining the same length as I_{ML}.

Shorrock proved that I_S is generalized Bayes in the following sense: Fix c_0, and consider all intervals of the form

$$I_\phi = \left(\phi(Z^2)S^2\ ,\ \left(\phi(Z^2) + c_0\right)S^2\right)\ . \tag{3.13}$$

Let the loss be

$$L\left(\sigma^2, I_\phi\right) = 1 - I\left(\sigma^2 \in I_\phi\right)\ , \tag{3.14}$$

then the corresponding risk function is

$$R\left(\sigma^2, I_\phi\right) = 1 - P\left(\sigma^2 \in I_\phi\right)\ ,$$

the probability of noncoverage. Further, consider the posterior Bayes risk, conditional on z^2,

$$R(\sigma^2, I_\phi \mid z^2) = 1 - \int I(\sigma^2 \in I_\phi)\pi(\sigma^2|z^2)d\sigma^2, \tag{3.15}$$

with

$$\pi(\sigma^2|z^2) = \frac{\int\int f(s^2,z^2|\mu,\sigma^2)\pi(\mu,\sigma^2)d\mu ds^2}{\int\int\int f(s^2,z^2|\mu,\sigma^2)\pi(\mu,\sigma^2)d\mu ds^2 d\sigma^2}\ , \tag{3.16}$$

where $f(s^2,z^2\,|\,\mu,\sigma^2)$ is the joint density function of S^2 and Z^2. The posterior $\pi(\sigma^2|z^2)$ depends on μ and σ^2 only through $\eta = \mu/\sigma$, so (3.16) is really of the form $\pi(\eta|z^2)$. If we take $\pi(\mu,\sigma) = \pi(\mu)\pi(\eta)$, then the posterior risk of I_ϕ is

$$R\left(\eta, I_\phi|z^2\right) = 1 - \int_{1/\phi+c_0}^{1/\phi} \pi(\eta|z^2)d\eta\ . \tag{3.17}$$

For $\pi(\eta)$ equal to the prior (2.24), I_S minimizes $R(\eta, I_\phi|z^2)$, i.e., ϕ_0 minimizes (3.17) for $\pi(\eta)$ of (2.23).

Although the intervals of Cohen and Shorrock have uniformly higher coverage probability than the minimum length interval, I_{ML}, they have the same length. They did not attempt the dual problem, that is, to obtain an interval with uniformly shorter length than I_{ML}, but the same coverage probability. Cohen (1972) did establish the following existence theorem, which shows it is possible to improve upon both length and coverage probability simultaneously.

Theorem 3.1 (*Cohen*): Under assumptions (1.1) there exists a confidence procedure with coverage probability greater than 1-α for all (μ,σ^2) and whose length, on a set of positive probability, is less than the length of I_{ML}.

Goutis (1989), using a construction similar to, but more general than, that of Shorrock, produces procedures that simultaneously improve upon the coverage probability and length of I_{ML}. These procedures are also shown to be generalized Bayes with respect to priors that are similar to (2.23). Goutis uses a construction and proof similar to that of Shorrock, but considers a more general class of intervals given by

$$I_\tau(Z^2,S^2) = \left(\phi_1(Z^2)S^2,\ \phi_2(Z^2)S^2\right), \tag{3.18}$$

where $\phi_1(t)$ and $\phi_2(t)$ satisfy the two conditions

$$f_{\nu+4}\left(\frac{1}{\phi_1(t)}\right)P\left(\chi_1^2 \le \frac{\tau(t)}{\phi_1(t)}\right) = f_{\nu+4}\left(\frac{1}{\phi_2(t)}\right)P\left(\chi_1^2 \le \frac{\tau(t)}{\phi_2(t)}\right),$$

(3.19)

$$\left[\frac{d\phi_1(t)}{dt}\right]f_{\nu+4}\left(\frac{1}{\phi_1(t)}\right)P\left(\chi_1^2 \le \frac{t}{\phi_1(t)}\right) = \left[\frac{d\phi_2(t)}{dt}\right]f_{\nu+4}\left(\frac{1}{\phi_2(t)}\right)P\left(\chi_1^2 \le \frac{t}{\phi_2(t)}\right),$$

where $\tau(t)$ is a positive function satisfying $\tau(t) \ge t$. Note that if $\tau(t) = t$ than I_τ reduces to Shorrock's interval, as the second condition in (3.19) reduces to $\phi_2(t) = \phi_1(t) + c_0$. If $\tau(t) > t$, however, the intervals are different and I_τ provides a length decrease over I_S, while maintaining 1-α coverage.

4. *Conditional Properties*

In this section, we present results of investigations into the conditional properties of the normal variance intervals mentioned in Section 3. Conditional properties of all frequentist confidence procedures have been of interest in recent years partly due to criticisms leveled at these procedures by Bayesians and others (See Berger and Wolpert (1984), Cox (1958), Fisher (1956a).) Suppose that X has a distribution depending on a parameter θ, and there exists a subset, $\mathcal{A}$, of the sample space such that

$$P(\theta \in C(X)|\mathcal{A}) < 1 - \alpha \qquad \forall\, \theta \tag{4.1}$$

for a $1-\alpha$ confidence procedure, $C(X)$. The statement that we are $1 - \alpha$ confident in $C(X)$ is certainly less than satisfying. If such a set, $\mathcal{A}$, exists, we would want to identify the set and, if $X \in \mathcal{A}$, we would want to either quote a different confidence or perhaps modify $C(X)$ to alleviate the problem of (4.1).

Fisher (1956a) first suggested the existence of such sets and called them *recognizable sets*. His suggestion to eliminate problems like (4.1) was to look at confidence conditional on an ancillary statistic. Fisher's solution worked to a degree, however, ancillary statistics sometimes don't exist or may be hard to find. A more formal structure for the evaluation of conditional properties of frequentist procedures was presented by Buehler (1959) following Fisher's lead. He presented an argument based on a two-person game, where for a specific form of bet the expected gain (or loss) can be interpreted as a conditional probability. It wasn't until 20 years later, however, that the theory was suitably formalized to allow for relatively easy evaluations of conditional properties. Robinson (1979a,1979b) explicitly formalized the theory with definitions on the types of bets and the biases involved. His formulation is much more general but can be used to evaluate a procedures conditional properties.

The basic idea is as follows. A confidence procedure is a pair, $<C(x), \gamma(x)>$, where $C(x)$ is a set estimator and $\gamma(x)$ is a quoted confidence (function). If you assert confidence $\gamma(x)$, you should be willing to take bets for or against coverage, with odds based on $\gamma(x)$. For example, if a person bets for coverage, the person risks $\gamma(x)$ to win $1 - \gamma(x)$ and vice versa if the bet is against coverage. A betting strategy or rule, $k(x)$, is a bounded function of x, which we can assume to satisfy $-1 \leq k(x) \leq 1$ without loss of generality. The function $k(x)$ can be interpreted as follows: $k(x)$ negative implies a bet against coverage; $k(x)$ positive implies a bet for coverage. Furthermore, $k(x)$ is said to be <u>relevant</u> if

$$E_\theta\left\{\left[I\left(\theta \in C(X)\right) - \gamma(X)\right]k(X)\right\} \geq \epsilon E_\theta|k(X)| \tag{4.2}$$

for all θ and some $\epsilon > 0$, where $I(\cdot)$ is the indicator function of a set, and <u>semirelevant</u> if

$$E_\theta\left\{\left[I\left(\theta \in C(X)\right) - \gamma(X)\right]k(X)\right\} \geq 0 \tag{4.3}$$

for all θ, with strict inequality for some θ. Note that the lefthand side of (4.2) or (4.3) is just the expected gain of the betting strategy k(x). So, in the general betting setup, relevancy is a statement about the existence of a winning strategy.

A special case of the betting setup leads us to a statistical (conditional inference) interpretation. For conditional evaluations of frequentist confidence procedures take $\gamma(x) = 1 - \alpha$, which is the case if C(X) is a frequentist procedure $\left(\text{since } P_\theta\left(\theta \in C(X)\right) \geq 1 - \alpha\right)$. Consider the betting strategy $k(x) = I(x \in A)$ where A is a subset of the sample space. If k is relevant then either

$$P_\theta(\theta \in C(X)|X \in A) \geq 1 - \alpha + \epsilon \qquad \forall\ \theta. \tag{4.4}$$

or

$$P_\theta(\theta \in C(X)|X \in A) \leq 1 - \alpha - \epsilon \qquad \forall\ \theta. \tag{4.5}$$

Similar statements hold if k is semirelevant except the corresponding statements for (4.4) and (4.5) would have no ϵ. The inequality (4.4) is positively biased (semi) relevant and (4.5) negatively biased (semi) relevant. If such a strategy exists we have a subset of the sample space where the conditional probability doesn't agree with the unconditional. Certainly, negative bias is most serious: the stated $(1 - \alpha)$ level of confidence is not attained if $x \in A$ of (4.5). Although still a problem, positive bias is less worrisome since $1 - \alpha$ would represent a conservative confidence level.

Procedures that are free of relevant betting are usually considered to have good conditional properties, although the exact desirable conditions are not yet agreed upon, indeed, not yet known. See Bondar (1977), Robinson (1979a,1979b), or Casella (1988) for a more extensive discussion of these ideas.

We next consider how to show that a confidence procedure, $<C(x), \gamma(x)>$, has good conditional properties, i.e., is free from relevant betting. Suppose we have $X \sim f(x|\theta)$, $<C(x), \gamma(x)>$, and we want to know when there does not exist k(x) such that (4.2) holds. Let

$\pi_m(\theta)$ be a prior distribution. If (4.2) is true then we have

$$\int_\Theta E_\theta\left\{\left[I\big(\theta \in C(x)\big) - \gamma(X)\right]k(X)\right\} \pi_m(\theta)d\theta \geq \epsilon \int_\Theta E_\theta|k(X)|\pi_m(\theta)d\theta \; . \tag{4.6}$$

If the order of integration can be changed, the lefthand side of (4.6) is equal to

$$\int_{\mathfrak{X}}\left[\int_\Theta I(\theta \in C(x)\pi_m(\theta|\chi)d\theta - \gamma(x)\right]k(x)m(x)dx \; , \tag{4.7}$$

where $\pi_m(\theta|x)$ and $m(x)$ are the posterior and marginal distributions resulting from the prior distribution, $\pi_m(\theta)$, respectively. If $\gamma(x)$ is equal to the posterior probability of $C(X)$, or a limit of posterior probabilities, then under suitable conditions (4.7) $\to$ 0 and, hence, the lefthand side of (4.6) $\to$ 0. However, the righthand side of (4.6) does not equal zero unless k is trivial. This is a contradiction. Thus, we can conclude that procedures that are limits of Bayes rules and satisfy

$$\gamma(x) = \lim_m \int_\Theta I\big(\theta \in C(x)\big)\pi_m(\theta|x)d\theta \quad , \tag{4.8}$$

are free of conditional problems (relevant betting).

Conditional properties of some classical procedures have been established, with most of the results applying to the case of $X_1, X_2, \cdots, X_n \sim$ iid $N(\mu,\sigma^2)$ with both μ and σ^2 unknown. Buehler and Fedderson (1963) showed that for $n = 2$ the usual Student's t confidence procedure $<\bar{X} \pm tS/\sqrt{n}, 1\text{-}\alpha>$, where t is the appropriate 1-α cutoff point, allows positively biased relevant subsets. Brown (1967) extended this result to arbitrary sample sizes, showing that sets of the form $\{(\bar{X},S) : |\bar{X}|/S \leq k\}$ are positively biased relevant for certain choices of the constant k. Thus, on sets that can be interpreted as *acceptance regions* of the hypothesis H_0: $\mu = 0$, the conditional confidence can be bounded above 1-α. Furthermore, Robinson (1976) showed that no negatively biased relevant sets exist for the t procedure, demonstrating acceptable conditional behavior.

Negatively biased semirelevant sets do exist for the t interval, a fact which follows from the above results. The interesting interpretation here is that there are negatively biased semirelevant sets of the form $\{(\bar{X},S) : |\bar{X}|/S \geq k\}$, sets that can be interpreted as *rejection* *regions* of the hypothesis H_0: $\mu = 0$. These results have been extended by Olshen (1973) to the analysis of variance to show that, conditional on rejecting the null hypothesis that all treatment means are equal, the Scheffé intervals have conditional confidence that can be bounded below 1-α.

Conditional investigations in case of two means with unequal, unknown, variances (the Behrens-Fisher problem) were done by Fisher (1956b) and Robinson (1976). Fisher showed that negatively biased relevant subsets exists for the intervals of Welch (1947), and Robinson showed that the Behrens-Fisher solution allowed no negatively biased relevant subsets. Robinson (1979b) also investigated conditional properties of procedures in more general situations, showing, in particular, that relevant subsets do not exist for location or scale procedures based on Pitman-type estimators.

We are now prepared to look at variance intervals. We start with intervals of the form (3.1) and consider conditional properties of the interval estimator: $<C(S^2),\ 1 - \alpha>$. Note that this includes all the usual interval estimators, both one-sided and two-sided. To show that $<C(S^2),\ 1 - \alpha>$ admits no relevant betting, it is sufficient to find a sequence $\pi_m(\mu,\sigma^2)$ that satisfies

$$\begin{aligned}(4.9)\qquad \lim_{m\to\infty}\int\int E_{\mu,\sigma^2}\left\{\left[I\left(\sigma^2 \in C(S^2)\right) - (1-\alpha)\right]k(\bar{X},S^2)\right\}\pi_m(\mu,\sigma^2)d\mu d\sigma^2 \\ \leq \lim_{m\to\infty}\int\int E_{\mu,\sigma^2}\left[\left|k(\bar{X},S^2)\right|\right]\pi_m(\mu,\sigma^2)d\mu d\sigma^2\end{aligned}$$

for all bounded $k(\bar{X},S^2)$.

There exist some technical problems in the establishment of (4.9), however. For the variance intervals of (3.1), the sequence of improper priors

$$(4.10)\qquad \pi(\mu,\sigma^2|r,a) \propto \left(\frac{1}{r\sigma^2}\right)^{1/2} e^{-1/2\mu^2/r\sigma^2}\left(\frac{1}{\sigma^2}\right)^{a}.$$

can be used to show that there is no relevant betting against the usual variance intervals I_{ET}, I_{ML}, and I_{SU} (Maatta and Casella, 1987).

It is interesting to note that both I_{ML} and I_{SU} are Bayes' highest posterior density regions using the following improper priors.

$$\text{(4.11)} \qquad I_{ML}: \pi(\mu,\sigma^2) = \left(\frac{1}{\sigma^2}\right)^3 d\mu d\sigma^2 \quad \text{and} \quad I_{SU}: \pi(\mu,\sigma^2) = \left(\frac{1}{\sigma^2}\right)^2 d\mu d\sigma^2 .$$

Furthermore, the posterior probabilities of coverage satisfy

$$\gamma_{ML}(S^2) = P\left(\sigma^2 \in I_{ML} | S^2\right) > 1 - \alpha \quad \text{and} \quad \gamma_{SU}(S^2) = P\left(\sigma^2 \in I_{SU} | S^2\right) = 1 - \alpha \quad \forall\ S^2 .$$

These results show that there is no relevant betting against

$$\langle I_{ML}, \gamma_{ML}(S^2)\rangle \qquad \text{or} \qquad \langle I_{SU}, \gamma_{SU}(S^2)\rangle .$$

In addition, since $\gamma_{ML}(S^2) > 1 - \alpha$, no negatively biased semirelevant betting exists against $\langle I_{ML}, 1 - \alpha\rangle$. Similarly, since $\gamma_{SU}(S^2) = 1 - \alpha$, neither negatively biased nor positively biased semirelevant betting exists for $\langle I_{SU}, 1 - \alpha\rangle$, an extremely strong conditional property for a frequentist procedure.

Though no relevant betting exists for the equal tailed interval, I_{ET}, the stronger conditional properties exhibited by I_{ML} and I_{SU} are not characteristic of I_{ET}. In fact, there exists semirelevant betting against $\langle I_{ET}, 1 - \alpha\rangle$. If we bet against coverage if $\bar{X}^2/S^2 < q_0$ for some (no too large) constant q_0, and otherwise don't bet, this strategy is negatively biased. Our intuition suggests that perhaps I_{ET} would not have the best properties when $\bar{X}^2/S^2 < q_0$, since this is the case when S^2 can be improved on as a point estimate. We would therefore expect an interval that is shifted toward zero when $\bar{X}^2/S^2 < q_0$ to be an improvement. This is exactly the case for I_{ML} and I_{SU}, and they both have better conditional properties than does I_{ET}.

As mentioned before, intervals of the form (3.1) also include the one-sided intervals. In particular, the theorem applies to the upper tailed interval $C_1(S^2) = \left\{\sigma^2 : \sigma^2 \leq bS^2\right\}$, which results from the inversion of a uniformly most powerful unbiased test, and the lower tailed interval $C_2(S^2) = \left\{\sigma^2 : \sigma^2 \geq aS^2\right\}$, which results from the inversion of a uniformly most powerful test. Further conditional properties of these intervals are studied by Maatta and Casella (1987).

The conditional properties of the improved variance intervals of Cohen (3.6) and Shorrock (3.11) were investigated by Casella and Maatta (1987) and agreed with expectations: both procedures allow no relevant betting. In addition, the smooth procedure (Shorrock's) allows no negatively biased betting while the discontinuous procedure (Cohen's) does. These results are in general agreement with the statement that procedures with good conditional properties are Bayes or limit of Bayes rules and thus must be smooth.

Shorrock's interval is a Bayes HPD region in the following sense. Start with the prior

$$\pi(\mu,\sigma^2) = \left(\frac{1}{\sigma^2}\right)^{5/2}\int_0^\infty e^{-vn\mu^2/2\sigma^2}\frac{v^{-1/2}}{(n+v)}\,dv\ , \tag{4.12}$$

which yields a posterior

$$\pi(\sigma^2\,|\,\bar{X},S^2) = \frac{f_{\nu+4}\left(S^2/\sigma^2\right)P\left(\chi_1^2 < n\bar{X}^2/\sigma^2\right)}{\frac{S^2}{\nu(\nu+4)}\,P\left(F_{1,\nu} < \nu(n\bar{X}^2/S^2)\right)}\ . \tag{4.13}$$

The posterior probability of an interval of the form $\left(\phi(Z^2)S^2,\ (\phi^2(Z^2)+c_o)S^2\right)$ is

$$\gamma(\bar{X},S^2) = \int_{\phi(Z^2)S^2}^{(\phi(Z^2)+c_0)S^2} \pi(\sigma^2\,|\,\bar{X},S^2)\,d\sigma^2\ . \tag{4.14}$$

This is maximized, for a fixed c_0, by choosing $\phi = \phi_0$ (as in 3.8). Call the resulting posterior probability $\gamma_0(\bar{X},S^2)$. Using the argument at (4.6)-(4.8), it is easy to see that no relevant betting exists against $<I_S,\ \gamma_0(\bar{X},S^2)>$. In addition, it can be shown that $\gamma_0(\bar{X},S^2) \geq 1-\alpha\ \forall\ \bar{X},S^2$, which implies that no negatively biased betting exists against $<I_S,\ 1-\alpha>$. This is the desired result, since $<I_S,\ 1-\alpha>$ is a frequentist procedure (as $1-\alpha$ is a pre-experimental confidence report), but $<I_S,\ \gamma_0(\bar{X},S^2)>$ is not (strictly) a frequentist procedure, as $\gamma_0(\bar{X},S^2)$ is not a pre-experimental confidence report.

Goutis (1989) also demonstrated that his intervals have acceptable conditional properties. For the intervals $I_{\mathcal{T}} = \left(\phi_1(Z^2)S^2,\ \phi_2(Z^2)S^2\right)$ defined in (3.18) and (3.19), define the confidence procedure $<I_{\mathcal{T}},\ \gamma_{\mathcal{T}}(\bar{X},S^2)>$, where $\gamma_{\mathcal{T}}(\bar{X},S^2)$ is given (analogous to (4.14)) by

$$\gamma_{\mathcal{T}}(\bar{X},S^2) = \int_{\phi_1(Z^2)}^{\phi_2(Z^2)} \pi(\sigma^2\,|\,\bar{X},S^2)\,d\sigma^2\ . \tag{4.15}$$

Then $<I_{\mathcal{T}},\ \gamma_{\mathcal{T}}(\bar{X},S^2)>$ allows no relevant subsets. Furthermore, $\gamma_{\mathcal{T}}(\bar{X},S^2) \geq 1-\alpha$, so the confidence procedure $<I_T, 1\text{-}\alpha>$ allows no negatively biased relevant subsets.

5. *Practical Improvements*

As previously mentioned, Rukhin (1987) has shown that the maximum relative improvement obtainable in the point estimation case (in terms of relative risk) is only 4%. Thus, the improvement associated with the Brewster-Zidek point estimator, and the confidence intervals of Shorrock and Goutis, can only be expected to be minimal. However, these (somewhat) negative statements only apply to the univariate case, where we observe one sample mean along with our variance estimate. The multivariate case, better known as the generalized linear model, offers the possibility of more substantial improvement.

All of the results presented here immediately apply to the generalized linear model case, where we observe $X = (X_1, X_2)$, a $(\nu+p)\times 1$ vector with $X_1 = (X_1,\cdots,X_\nu)$ and $X_2 = (X_{\nu+1},\cdots,X_{\nu+p})$. We assume that X is a multivariate normal random vector with mean $(0,\mu)$, where 0 is of order ν and $\mu = (\mu_1,\cdots,\mu_p)$ is unknown. The covariance matrix is σ^2 times the ν+p identity matrix, and we are interested in estimating the unknown parameter σ^2. This set-up is the familiar analysis of variance.

Continuing with our general definitions, we have

$$S^2 = X_1{}'X_1, \text{ with } S^2/\sigma^2 \sim \chi^2_\nu \text{ and } Y^2 = X_2{}'X_2, \text{ with } Y^2/\sigma^2 \sim \chi^2_p(\lambda),$$

with $\lambda = \mu'\mu/\sigma^2$. (In analysis of variance terminology, Y^2 represents the model sum of squares and S^2 represents the error sum of squares.) In this set-up, Stein's estimator of (2.4) becomes $\phi_S(Z)S^2$, where

$$\phi_S(Z) = \min\left(\frac{1}{\nu+2}, \frac{1+Z^2}{\nu+p+2}\right), \quad Z^2 = Y^2/S^2. \tag{5.1}$$

As before, this estimator is better than the usual estimator under squared error loss. The point estimators of Brown and Brewster and Zidek can similarly be generalized, and their optimality properties also carry over.

In the confidence interval problem the analogous generalizations hold, and optimality properties are also preserved. For example, Shorrock's interval generalizes to

$$I_S(Z^2,S^2) = \Big(\phi_0(Z^2)S^2,\ \big(\phi_0(Z^2) + c_0\big)S^2\Big)\,, \tag{5.2}$$

where c_0 is the length of I_{ML} and $\phi_0(Z^2)$ is the unique root of

$$f_{\nu+4}\Big(\frac{1}{\phi}\Big)P\Big(\chi_p^2 \le \frac{Z^2}{\phi}\Big) = f_{\nu+4}\Big(\frac{1}{\phi+c_0}\Big)P\Big(\chi_p^2 \le \frac{Z^2}{\phi+c_0}\Big)\,. \tag{5.3}$$

The interval I_S retains its dominance over I_{ML}, and, contrary to Rukhin's findings in the univariate point estimation case, the improvement here can be substantial. For example, Shorrock reports numerical studies that show for large p (p=29, n=29) the maximum coverage probability of a nominal 90% interval can reach 92.3%. Other cases show a similar degree of improvement, with the maximum coverage generally increasing in p. (The coverage probability of Shorrock's interval is a nonconstant function of the noncentrality parameter λ, with the maximum being attained for a medium value, and coverage decreasing to the nominal level as $\lambda \to \infty$.)

In practice, the improvements of Shorrock's interval, although reasonable, are not tangible. That is, an experimenter still has an interval of the same length and same nominal coverage (although higher actual coverage). With the intervals of Goutis, however, a tangible gain is realized in that the length is decreased while keeping the coverage probability above the nominal level. These intervals also carry over to the case of the generalized linear model, with all of their optimality properties intact. For the generalized linear model the class of intervals is given by

$$I_\tau(Z^2,S^2) = \Big(\phi_1(Z^2)S^2,\ \phi_2(Z^2)S^2\Big)\,, \tag{5.4}$$

where $\phi_1(t)$ and $\phi_2(t)$ satisfy

$$f_{\nu+4}\Big(\frac{1}{\phi_1(t)}\Big)P\Big(\chi_p^2 \le \frac{\tau(t)}{\phi_1(t)}\Big) = f_{\nu+4}\Big(\frac{1}{\phi_2(t)}\Big)P\Big(\chi_p^2 \le \frac{\tau(t)}{\phi_2(t)}\Big)\,,$$

(5.5)

$$\Big[\frac{d\phi_1(t)}{dt}\Big]f_{\nu+4}\Big(\frac{1}{\phi_1(t)}\Big)P\Big(\chi_p^2 \le \frac{t}{\phi_1(t)}\Big) = \Big[\frac{d\phi_2(t)}{dt}\Big]f_{\nu+4}\Big(\frac{1}{\phi_2(t)}\Big)P\Big(\chi_p^2 \le \frac{t}{\phi_2(t)}\Big)\,,$$

for a positive function $\tau(t)$ satisfying $\tau(t) \geq t$. The goal of reducing length (over either I_{ML} or I_S), while maintaining coverage probability close to (but no smaller than) the nominal level, results in reasonable practical improvements, as the following table illustrates.

Table 1 about here

From Table 1 we can see that the amount of improvement possible is large enough to warrant the use of these improved procedures, particularly in the analysis of variance. Although these newer procedures require more computing power, there is widespread availability of such power, so this requirement is no longer a drawback. Furthermore, these procedures all have the minimax property that they are uniformly superior to the usual procedures. Thus, even if the situation is one where only a minimal improvement is possible (small p), it is reasonable to try for that small improvement.

6. *Discussion*

The problem of estimating the normal variance, using a decision-theoretic approach, has an illustrious history. The seed of the idea used to improve the usual point estimator of σ^2, stems from Stein and flows naturally to Brown and then to Brewster and Zidek. Improvement of the interval estimator also stems from Stein (and Brown) and flows naturally, in similar fashion, to Cohen, Shorrock, and Goutis.

Outside of the normal case, there have been many advances in the point estimation of scale parameters. Berger (1980) investigated simultaneous estimation of gamma scale parameters for a variety of loss functions, and discovered a Stein-type phenomenon there. Numerous authors have shown how to improve on the usual estimators of exponential scale. In particular, Rukhin and Strawderman (1982) consider the more general case of improved estimation of exponential quantiles, and produced estimators that substantially improve on risk. Less effort has been made, outside of the normal case, in the interval estimation case, a case that is ripe for consideration.

All of the interval estimation results considered here implicitly use a loss function with two components, coverage probability and length. Work of Cohen and Strawderman (1973) relate the coverage probability-length loss function to a loss function with components of coverage probability and probability of false coverage. They prove that if a procedure is admissible with respect to the coverage probability-length loss function then it is *almost admissible* (a slightly weaker condition) with respect to the coverage probability-probability of false coverage loss function. (The proof uses the identity of Pratt (1961).) The conditions of the Cohen-Strawderman result are satisfied for interval estimation of the normal variance, so length optimality will transfer over (almost) to false coverage optimality.

While Rukhin (1987) has shown that the maximum relative improvement is minimal in the univariate case, the possible improvement in the generalized linear model (analysis of variance) case can be substantial. Regardless of the amount of improvement possible in these cases, the innovative ideas presented in the original proof of Stein, and the subsequent modifications, certainly merit recognition as groundbreaking work. The fact that so much work in the point and interval estimation cases has come from these ideas, ideas that seem almost trivial in retrospect, is testimony to their innovation and importance.

Intersecting this progression of improvement has been the result of the conditional evaluations of interval estimators by Maatta and Casella (1987). While showing that the usual intervals generally have adequate conditional properties, it is shown that I_{SU}, the shortest unbiased interval has stellar conditional characteristics. In addition, the work of Casella and Maatta concerning Cohen and Shorrock's improved intervals serves as a starting point for subsequent improvements. The most recent addition of the work of Goutis (1989) shows that it is possible to construct generalized Bayes invariant intervals that improve upon the intervals of Shorrock and also maintain acceptable conditional properties.

Acknowledgement. We thank the referees for many helpful comments.

References

Berger, J.O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* 8: 545-571.

Berger, J.O. and Wolpert, R.W. (1984). *The Likelihood Principle.* Institute of Mathematical Statistics, Lecture Notes-Monograph Series.

Bondar, J.V. (1977). A conditional confidence principle. *Ann. Statist.* 5 :881-891.

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

Brown, L. (1967). The conditional level of Student's t-test. *Ann. Math. Statist.* 38: 1068-1071.

Brown, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.* 39: 29-48.

Buehler, R.J. (1959). Some validity criteria for statistical inference. *Ann. Math. Statist.* 30:845-863.

Buehler, R.J. and Fedderson, A.P. (1963). Note on a conditional property of Student's t. *Ann. Math. Statist.* 34:1098-1100.

Casella, G. (1988). Conditionally acceptable frequentist solutions. *Statistical Decision Theory and Related Topics IV* (S.S. Gupta and J.O. Berger, eds.), 73-117.

Casella, G. and Maatta, J.M. (1987). Conditional properties of generalized Bayes variance intervals. BU-939-M, Biometrics Unit, Cornell University.

Cohen, A. (1972). Improved confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* 67: 382-387.

Cohen, A. and Sackrowitz, H. (1970). Estimation of the last mean of a monotone sequence. *Ann. Math. Statist.* 41: 2021-2034.

Cohen, A. and Strawderman, W.E. (1973). Admissibility implications for different criteria in confidence estimation. *Ann. Statist.* 1: 363-366.

Cox, D.R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.*

29: 357-372.

Fisher, R.A. (1956a). *Statistical Methods and Scientific Inference*, *3rd* Ed., Hafner, New York.

Fisher, R.A. (1956b). On a test of significance in Pearson's *Biometrika* tables (no.11). *J. Royal Statist. Soc. Ser. B* **18**: 56-60.

Girshick, M.A. and Savage, L.J. (1951). Bayes and minimax estimates for quadratic loss functions, *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, 53-73.

Goutis, C. (1989). Improved Invariant Set Estimation of a Normal Variance with Generalizations. Ph.D. Thesis, Cornell University.

Hodges, J.L. and Lehmann, E.L. (1951). Some applications of the Cramér-Rao inequality, *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, 13-22.

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, 361-379.

Maatta, J.M. and Casella, G. (1987). Conditional properties of interval estimators of the normal variance *Ann. Statist.* **15**: 1372-1388.

Olshen, R.A. (1973). The conditional level of the F test. *J. Amer. Statist. Assoc.* **68**: 692-698.

Pratt, J.W. (1959). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56**: 549-567.

Proskin, H.M. (1985). An admissibility theorem with applications to the estimation of the variance of the normal distribution. Unpublished Ph.D. dissertation, Rutgers University, Department of Statistics.

Robinson, G.K. (1976). Properties of Student's t and the Behrens-Fisher solution to the two means problem. *Ann. Statist.* **4**: 963-971.

Robinson, G.K. (1979a). Conditional properties of statistical procedures. *Ann. Statist.* **7**: 742-755.

Robinson, G.K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist.* **7**: 756-771.

Rukhin, A.L. (1987). How much better are better estimators of a normal variance. *J. Amer. Statist. Assoc.* **82**:925-928.

Rukhin, A.L. and Strawderman, W.E. (1982). Estimating a quantile of an exponential distribution. *J. Amer. Statist. Assoc.* **77**:159-162.

Shorrock, G. (1982). A minimax generalized Bayes confidence interval for a normal variance. Unpublished Ph.D. Thesis, Department of Statistics, Rutgers University.

Shorrock, G. (1990). Improved confidence intervals for a normal variance. *Ann. Statist* 18. (To appear).

Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, 197-206.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16**: 155-160.

Strawderman, W.E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.* **2**: 190-198.

Tate, R.F. and Klett, G.W. (1959). Optimal confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **54**: 674-682.

Welch, B.L. (1947). The generalization of Student's problem. *Biometrika* **34**: 28-34.

Table 1. Relative length reduction (in percent) of the interval I_τ over I_{ML}, for $\tau(t) = 2t$. Values in parentheses are the maximum coverage probability.

$1 - \alpha = .950$

n	p = 1	5	6	10
10	.22 (.951)	3.3 (.951)		
25			2.9 (.952)	5.3 (.952)

$1 - \alpha = .900$

n	p = 5	10	20
50	1.1 (.902)	2.4 (.902)	3.3 (.902)

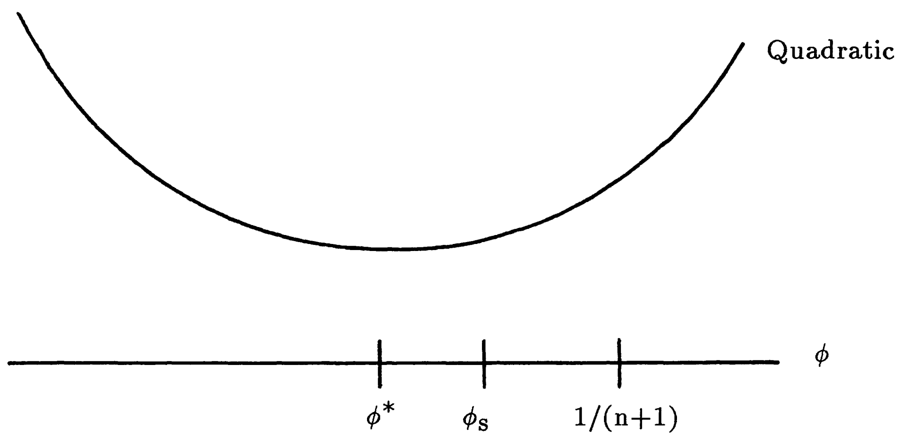

Figure 1. For fixed z, the quadratic $\phi^2(z)E\left((S^2/\sigma^2)^2|z\right) - 2\phi(z)E(S^2/\sigma^2|z) + 1$. The value ϕ^* is the minimum (see equation (2.7)), and ϕ_S yields Stein's estimator (see equation (2.10)).