Improved Invariant Set Estimation

for General Scale Families

Constantinos Goutis[1]

University College London

George Casella[2]

Cornell University

BU-1090-M July, 1990

AMS 1980 subject classifications. Primary 62F25; Secondary 62C99.

Key words and phrases. Optimal intervals, coverage probability.

[1]This research is part of the author's Ph.D. dissertation at Cornell University.

[2]Research supported by National Science Foundation Grant No. DMS-890039.

ABSTRACT

Confidence intervals for any power of a scale parameter of a distribution are constructed. We first state the needed general distributional assumptions and show how to construct the minimum length location-scale invariant interval, having a predetermined coverage coefficient $1-\alpha$, which is the analog of the "usual" interval for a normal variance. If we then relax the invariance restriction, we obtain an improved interval. Using information relating the size of the location parameter to that of the scale parameter, we shift the minimum length interval closer to zero, simultaneously bringing the endpoints closer to each other. These intervals have guaranteed coverage probability uniformly greater than a predetermined value $1-\alpha$, and have uniformly shorter length.

1. Introduction. The problem of constructing optimal estimators for a scale parameter when the location parameter is unknown has a long history. The most common case is that of estimating the variance of the normal distribution with unknown mean, the history of which is given in Maatta and Casella (1990). This history can be traced back at least to Stein (1964), who showed that we can improve on the "usual" point estimator for the variance by using information about the size of the sample mean relative to the sample variance. His estimation procedure can be thought of as first testing the null hypothesis that the population mean is zero, and, if accepted, calculating the variance around zero rather than the sample mean. In this way, whenever the population mean seems to be small, another degree of freedom is gained and we are able to beat the usual estimator based on the sample variance alone. Brown (1968) extended Stein's results to more general loss functions and a larger class of distributions. His estimator applies to a general scale parameter when the location parameter is unknown. He uses the usual estimator s^2 for scale parameter whenever the estimate, y, of the location parameter seems large and a smaller multiple of s^2 whenever y seems small. The relative size of s^2 is measured by the statistic $t = y^2/s^2$.

Both Stein's and Brown's estimators are inadmissible, thus it is possible to improve upon these estimators. Brewster and Zidek (1974) were able to find better estimators by taking a finer partition of the set of possible values of t. Their estimator is "smooth" enough to be generalized Bayes and, under some conditions, admissible among scale invariant point estimators.

The problem of the interval estimation is in many ways similar to the problem of point estimation. Tate and Klett (1959) calculated the endpoints of the shortest confidence intervals for a normal variance, based on s^2 alone. Cohen (1972) was able to construct improved confidence intervals adapting Brown's (1968) techniques. Cohen's intervals keep the same length but, by shifting the endpoints towards zero whenever $t \leq K$, some fixed but arbitrary constant, he was able to dominate Tate and Klett's intervals in terms of coverage probability. Cohen's intervals are for the special case of normal variance.

Shorrock (1990) further improved on Cohen's result. In a manner analogous to Brewster and

Zidek, Shorrock partitioned the set of possible values of t. By successively adding more cutoff points he was able to construct a "smooth" version of Cohens interval. The resulting interval is highest posterior density region with respect to an improper prior and dominates Tate and Klett's interval based on s^2 alone. For both Shorrock- and Cohen-type intervals the domination is only in terms of coverage probability since, by construction, the length is kept fixed and equal to the usual length. Furthermore, the confidence coefficient remains equal to $1-\alpha$ since asymptotically, as the noncentrality parameter $\lambda = \mu^2/\sigma^2$ tends to infinity, the endpoints of the intervals coincide with the endpoints of the usual interval.

Goutis and Casella (1990) constructed intervals for a normal variance which improve upon the usual shortest interval based on s^2 alone in terms of length. They kept the minimum coverage probability equal to a predetermined value $1-\alpha$ and shifted the interval closer to zero whenever the sample indicates that the mean is close to zero. By shifting they were able to bring the endpoints closer to each other, hence producing shorter intervals. Using a method similar to Brewster and Zidek they constructed a family of "smooth" $(1-\alpha)\,100$ % intervals which are shorter than the usual interval and, consequently, Cohen and Shorrock type intervals. This paper contains a generalization of the construction to confidence intervals for a general scale parameter, without relying on normality. We first generalize conditions for constructing optimal location-scale invariant intervals, making as few distributional assumptions as possible. Then we show that, under suitable assumptions, we can use techniques similar to Goutis and Casella (1990) to construct families of intervals for the scale parameter when the location parameter is unknown.

2. Assumptions and location scale invariant intervals. We state the assumptions needed in the general scale parameter case, several of which are similar to Brown's (1968) distributional assumptions.

Let Y, S, Z be random variables taking the values y, s, z in $\Re^1 \times \Re^1 \times \Re^q$, $q \geq 0$ and $s > 0$. The variable Z is an ancillary statistic and may not exist. If this is the case, we take $q = 0$. We assume that the distribution of Z, f(z), does not depend on any unknown parameters. Given $Z = z$, the random variables Y and S have a conditional density with respect to Lebesgue measure of the form

$$f_{\mu,\sigma}(s, y \mid z) = \frac{1}{\sigma^2} f_{0,1}(\frac{s}{\sigma}, \frac{y-\mu}{\sigma} \mid z) \tag{2.1}$$

that is, the conditional density belongs to the location-scale family. The location parameter is denoted by μ and the scale parameter by σ. The random variable S has a density

$$f_\sigma(s \mid z) = \frac{1}{\sigma} f(\frac{s}{\sigma} \mid z), \tag{2.2}$$

for some functions f_σ and f, where f is independent of μ and σ.

We are interested in estimating the parameter σ^ρ where $\rho > 0$ a fixed known constant. Any location and scale invariant estimator is of the form $\psi(z)s^\rho$, for some function ψ. However, if we require invariance only under rescaling and change of sign of the data, the class of estimators is increased to those of the form $\phi(\mid y \mid /s, z)s^\rho$, for some function ϕ.

When constructing confidence intervals, we consider only connected confidence intervals. Although our results extend to confidence sets that are not connected, such procedures are intuitively unappealing. A sufficient, but not necessary, condition for the minimum length intervals to be connected is that the densities be unimodal, however, we will not make such an assumption. If there is no unimodality, it should be understood that by "minimum length intervals" we mean "minimum length connected intervals".

We need some additional assumptions about the density $f_\sigma(s \mid z)$ in order to determine the endpoints of the shortest location and scale invariant $1-\alpha$ confidence intervals for s^ρ. The confidence interval is invariant if it is of the form $(\psi_1(z)s^\rho, \psi_2(z)s^\rho)$, and to have the confidence coefficient to be $1-\alpha$, we must have

$$P\{\sigma^\rho \in (\psi_1(z)s^\rho, \psi_2(z)s^\rho)\} \geq 1-\alpha. \tag{2.3}$$

Conditioning on z the above inequality becomes

$$\int P\{\psi_1(z)s^\rho \leq \sigma^\rho \leq \psi_2(z)s^\rho) \mid z\} f(z)\, dz \geq 1-\alpha. \tag{2.4}$$

Observe that the conditional probability does not depend on any unknown parameters since f(z) of (2.2) is free of unknown parameters. Straightforward calculation shows that a sufficient condition for (2.2) is

$$\int_{\frac{1}{\psi_2(z)}}^{\frac{1}{\psi_1(z)}} f(s^\rho \mid z)\, ds^\rho = 1-\alpha \tag{2.5}$$

for almost all z. Although equation (2.5) is not necessary for (2.3), if the endpoints $\psi_1(z)$ and $\psi_2(z)$ satisfy (2.3) but not (2.5), then the interval would have undesirable conditional properties since the conditional coverage probability would be bounded on one side of $1-\alpha$ for a range of z values. In the future we will omit the explicit dependency of the endpoints on z, for notational convenience. It should be understood, however, that everything is conditional on ancillary statistics, whenever they exist.

Now we derive a convenient expression for the endpoints of the shortest location-scale invariant interval satisfying the probability constraint.

THEOREM 2.1. If $f_\sigma(s \mid z)$ is continuous and has connected support for almost all z then the minimum length location-scale invariant interval, subject to the condition (2.5), is given by

$$C_U(s, z) = (\, \psi_1 s^\rho,\ \psi_2 s^\rho\,) \tag{2.6}$$

where ψ_1 and ψ_2 satisfy equation (2.5) and

$$\left(\frac{1}{\psi_1}\right)^2 f\left(\frac{1}{\psi_1} \mid z\right) = \left(\frac{1}{\psi_2}\right)^2 f\left(\frac{1}{\psi_2} \mid z\right). \tag{2.7}$$

PROOF. Equation (2.7) can be derived by using the technique of Lagrange multipliers, minimizing $\psi_2 - \psi_1$ subject to the constraint (2.5). In order to use the Lagrange multipliers we need the differentiability of the upper limit of integration as a function of the lower limit, where the integrand is the density of s^ρ. If the assumption of continuity and connected support holds for $f_\sigma(s \mid z)$ then it also holds for $f(s^\rho \mid z)$ and we apply Lemma A.2 to derive equation (2.7).

Comparing the difference $\psi_2 - \psi_1$ to the possibly multiple solutions of (2.5) and (2.7) guarantees that we have a minimum. □

The assumptions of continuity and connectedness of support of $f_\sigma(s \mid z)$ may be stronger than necessary, because the shortest $1-\alpha$ interval may exist even if they are not met. However, we do not consider it as a major drawback since most densities of practical interest satisfy these conditions.

3. Improving upon location scale invariant intervals. Now we look for intervals that are superior to interval $C_U(s, z)$ by no longer requiring the interval to be location and scale invariant. The manner of improving is similar to that used in Goutis and Casella (1990), that is shifting the endpoints of the interval whenever $|y|/s$ seems small. In order to have consistent notation we define the statistic $t = y^2/s^2$. Observe that there is a one-to-one relation between t and $|y|/s$.

In the proof of Theorem 3.1 we will assume that given z, y^2 and s are conditionally independent. Even though this does not seem to be a crucial condition for the construction to work, if it holds, the distribution of y^2/σ^2 depends on the parameters only through μ^2/σ^2. Therefore by taking, without loss of generality, $\sigma = 1$ the distribution depends only on μ. We will denote the cumulative distribution function of y^2 by $F_\mu(y \mid z)$.

Two additional assumptions, which are important for the construction, are that the distribution $F_\mu(y \mid z)/F_0(y \mid z)$ is increasing in y and that $F_\mu(y \mid z)$ is continuous as a function of y. Note that if the density of y is symmetric around μ, then $F_\mu(y \mid z)$ depends on μ only through its absolute value. In that case it suffices to require $F_\mu(y \mid z)$ to have the monotone likelihood ratio property in $|\mu|$.

For fixed constant K define the interval $I_1(s, t, z, K)$ as follows:

$$I_1(s,t,z,K) = \begin{cases} (\psi_1 s^\rho, \psi_2 s^\rho) & \text{if} \quad t > K \\ (\phi_1(K) s^\rho, \phi_2(K) s^\rho) & \text{if} \quad t \le K \end{cases} \tag{3.1}$$

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

$$\int_{\phi_1(K)}^{\phi_2(K)} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K}{x^{2/\rho}} \mid z\right) dx = \int_{\psi_1}^{\psi_2} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K}{x^{2/\rho}} \mid z\right) dx \tag{3.2}$$

and $\phi_1(K) \le \psi_1$. Then we have the following theorem.

THEOREM 3.1. The coverage probability of $\mathbf{I_1}(s,t,z,K)$ is no less than the coverage probability of $C_U(s, z) = (\psi_1 s^\rho, \psi_2 s^\rho)$.

PROOF. If $P_{\mu,1}\{ t \le K \} = 0$ there is nothing to prove. Otherwise, observing that the intervals $\mathbf{I_1}(s,t,z,K)$ and $C_U(s, z)$ differ only when $t \le K$, it suffices to work with the joint probability

$$P\{\sigma^\rho \in \mathbf{I_1}(s,t,z,K),\ t \le K\} = P\{\frac{1}{\phi_1(K)} \le \frac{s^\rho}{\sigma^\rho} \le \frac{1}{\phi_2(K)}, \frac{y^2}{s^2} \le K\}. \tag{3.3}$$

Taking $\sigma = 1$ without loss of generality and conditioning on $s^\rho = w$, yields

$$P\{\sigma^\rho \in \mathbf{I_1}(s,t,z,K),\ t \le K\} = \int_{\frac{1}{\phi_2(K)}}^{\frac{1}{\phi_1(K)}} f(w \mid z)\, P_\mu(y^2 \le K w^{2/\rho} \mid s^\rho = w, z)\, dw$$

$$= \int_{\frac{1}{\phi_2(K)}}^{\frac{1}{\phi_1(K)}} f(w \mid z)\, F_\mu(K w^{2/\rho} \mid z)\, dw. \tag{3.4}$$

Using the transformation $x = 1/w$, the theorem is proved if we establish

$$\int_{\phi_1(K)}^{\phi_2(K)} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_\mu\left(\frac{K}{x^{2/\rho}} \mid z\right) dx \ge \int_{\psi_1}^{\psi_2} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_\mu\left(\frac{K}{x^{2/\rho}} \mid z\right) dx \tag{3.5}$$

for every μ. The proof of equation (3.5) uses the same lines of reasoning as in Theorem 2.1 of Goutis and Casella (1990). For fixed γ and μ define $g_{\gamma,\mu}(w)$ as the solution to

$$\gamma = \int_{w}^{g_{\gamma,\mu}(w)} \left(\tfrac{1}{x}\right)^2 f(\tfrac{1}{x} \mid z)\, F_\mu\!\left(\frac{K}{x^{2/\rho}} \mid z\right) dx. \tag{3.6}$$

Since $f(1/x \mid z)$ and $F_0(K/x^{2/\rho} \mid z)$ are assumed continuous the integrand is continuous. Furthermore, since $F_0(K/x^{2/\rho} \mid z)$ is monotone and $f(1/x \mid z)$ has connected support, the integrand has connected support. Hence we can apply Lemma A.2 and conclude that $g_{\gamma,\mu}(w)$ is a differentiable function.

Let γ_1 and γ_2 satisfy $g_{\gamma_1,0}(\psi_1) = g_{\gamma_2,\mu}(\psi_1) = \psi_2$ and $G(w) = g_{\gamma_1,0}(w) - g_{\gamma_2,\mu}(w)$, and note that $G(\phi_1(K)) = \phi_2(K) - g_{\gamma_2,\mu}(\phi_1(K))$ and $G(\psi_1) = 0$. We can establish (3.5) by showing that G satisfies the assumptions of Lemma A.1, which implies $\phi_2 > g_{\gamma_2,\mu}(\phi_1)$.

Let x_0 be a point such that $G(x_0)=0$ and let $y_0 = g_{\gamma_1,0}(x_0) = g_{\gamma_2,\mu}(x_0)$. Since γ_1 and γ_2 are fixed, by Lemma A.2

$$\left.\frac{dG(w)}{dw}\right|_{w=x_0} = \frac{(\frac{1}{x_0})^2 f(\frac{1}{x_0} \mid z)}{(\frac{1}{y_0})^2 f(\frac{1}{y_0} \mid z)} \left\{ \frac{F_0\!\left(\frac{K}{x_0^{2/\rho}} \mid z\right)}{F_0\!\left(\frac{K}{y_0^{2/\rho}} \mid z\right)} - \frac{F_\mu\!\left(\frac{K}{x_0^{2/\rho}} \mid z\right)}{F_\mu\!\left(\frac{K}{y_0^{2/\rho}} \mid z\right)} \right\}. \tag{3.7}$$

The monotonicity of F_μ / F_0 establishes that the term in braces is negative, hence by Lemma A.1, $G(\phi_1(K))$ is positive and (3.5) is established. □

REMARK. If y^2 and s are not conditionally independent, the probability $P_\mu(y^2 \le Kw^{2/\rho} \mid s^\rho = w, z)$ may not have a tractable form. However, making the conditional independence assumption we have

$$P_\mu(y^2 \le Kw^{2/\rho} \mid s^\rho = w, z) = F_\mu(Kw^{2/\rho} \mid z) \tag{3.8}$$

which justifies (3.4). The assumption is used only to ensure that the argument of F_μ is a monotone function of w. Our construction would work if we simply require the function

$$\frac{P_\mu(y^2 \le Kw^{2/\rho} \mid s^\rho = w, z)}{P_0(y^2 \le Kw^{2/\rho} \mid s^\rho = w, z)} \tag{3.9}$$

to be increasing in w.

We saw that for every $\phi_1(K)$ and $\phi_2(K)$ satisfying (3.2) and $\phi_1(K) \leq \psi_1$ the coverage probability of $I_1(s,t,z,K)$ is at least $1-\alpha$. However, in order to gain in length we need some additional restrictions on $\phi_1(K)$. When $t \leq K$, the length of the interval is equal to $(\phi_2(K)-\phi_1(K))s^\rho$. Subject to (3.2) the length is decreasing as a function of the lower limit of integration if $dg_{\gamma_1,0}(w)/dw > 1$. Using the formula for the derivative of $g_{\gamma_1,0}(w)$ derived from Lemma A.2, we obtain the expression

$$\frac{dg_{\gamma_1,0}(w)}{dw}\bigg|_{w=\psi_1} = \frac{F_0\left(\frac{K}{\psi_1^{2/\rho}} \mid z\right)}{F_0\left(\frac{K}{\psi_2^{2/\rho}} \mid z\right)}. \tag{3.10}$$

The last ratio is always greater than or equal to one since $\psi_1 < \psi_2$ and F_0 is a nondecreasing function. Therefore we cannot have an increase on the length for any K. Lemma A.3 shows that we can always chose an appropriate K such that

$$F_0\left(\frac{K}{\psi_1^{2/\rho}} \mid z\right) > F_0\left(\frac{K}{\psi_2^{2/\rho}} \mid z\right) \tag{3.11}$$

that is, the derivative of the length, as a function of the lower limit, is strictly negative in some neighborhood of ψ_1. Since from Lemma A.2 we also know that the derivative is continuous, for every $\phi_1(K)$ sufficiently close to ψ_1 we will have some gain in length. We now make more precise what we mean by "sufficiently close".

Define $\phi_1^0(K)$ and $\phi_2^0(K)$ to satisfy

$$\int_{\phi_1^0(K)}^{\phi_2^0(K)} \left(\tfrac{1}{x}\right)^2 f\left(\tfrac{1}{x} \mid z\right) F_0\left(\frac{K}{x^{2/\rho}} \mid z\right) dx = \int_{\psi_1}^{\psi_2} \left(\tfrac{1}{x}\right)^2 f\left(\tfrac{1}{x} \mid z\right) F_0\left(\frac{K}{x^{2/\rho}} \mid z\right) dx \tag{3.12}$$

and

$$\left(\frac{1}{\phi_1^0(K)}\right)^2 f\left(\frac{1}{\phi_1^0(K)} \mid z\right) F_0\left(\frac{K}{\{\phi_1^0(K)\}^{2/\rho}} \mid z\right)$$

$$= \left(\frac{1}{\phi_2^0(K)}\right)^2 f\left(\frac{1}{\phi_2^0(K)} \mid z\right) F_0\left(\frac{K}{\{\phi_2^0(K)\}^{2/\rho}} \mid z\right). \tag{3.13}$$

Since we do not assume unimodality of the function $(1/x)^2 f(1/x \mid z)\ F_0(K/x^{2/\rho} \mid z)$ there may be more than one solution to (3.12) and (3.13). If there are no solutions with $\phi_1^0(K) \leq \psi_1$, then any $\phi_1(K) \leq \psi_1$ and $\phi_2(K)$ satisfying (3.2) defines a confidence interval $\mathbf{I_1}(s, t, z, K)$ shorter than $(\psi_1 s^\rho,\ \psi_1 s^\rho)$. Otherwise we take $\phi_1(K)$ greater than every solution to equations (3.12) and (3.13) with $\phi_1^0(K) \leq \psi_1$. Then $dg_{\gamma_1,0}(w)/dw > 1$ for every w in the interval $(\phi_1(K), \psi_1)$ which implies that we have some gain in terms of length. Thus we need to take $\phi_1(K)$ close enough to ψ_1 so that

$$\left(\frac{1}{w}\right)^2 f\left(\frac{1}{w} \mid z\right) F_0\left(\frac{K}{w^{2/\rho}} \mid z\right) > \left(\frac{1}{g_{\gamma_1,0}(w)}\right)^2 f\left(\frac{1}{g_{\gamma_1,0}(w)} \mid z\right) F_0\left(\frac{K}{\{g_{\gamma_1,0}(w)\}^{2/\rho}} \mid z\right)$$

$$\forall\, w \in [\,\phi_1(K),\ \psi_1\,], \tag{3.14}$$

The last requirement is intuitively expected. If the value of the integrand at the lower limit of the interval is smaller that the value at the upper limit then by keeping the area constant and shifting the lower limit towards zero we would increase the distance between the endpoints.

If $\phi_1(K)$ is close enough to ψ_1 to satisfy both (3.2) and (3.14), we can further improve upon the interval $\mathbf{I_1}(s, t, z, K)$ by taking another cutoff point. By further partitioning the set of possible values of t, we can construct confidence intervals that are based on more cutoff points, and eventually fill the interval $(0, +\infty)$ with points. In a manner similar to Brewster and Zidek, we create a triangular array $\{\underset{\sim}{K}_m\}$ array as follows: For each m, define $\underset{\sim}{K}_m = (K_{m,1}, \ldots, K_{m,m-1}, K_{m,m})$, where $0 < K_{m,1} < \ldots < K_{m,m-1} < K_{m,m} < +\infty$. Furthermore, we require $\lim_{m\to\infty} K_{m,1} = 0$ and $\lim_{m\to\infty} K_{m,m} = +\infty$ and $\lim_{m\to\infty} \max_i (K_{m,i} - K_{m,i-1}) = 0$.

The intervals based on $\underset{\sim}{K}_1, \underset{\sim}{K}_2, \ldots$ will be denoted by $\mathbf{I_m}(s, t, z, \underset{\sim}{K}_m)$ and the endpoints satisfy:

For $i = 1, 2, \ldots, m-1$,

$$\int_{\phi_1(K_{m,i})}^{\phi_2(K_{m,i})} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,i}}{x^{2/\rho}} \mid z\right) dx$$

$$= \int_{\phi_1(K_{m,i+1})}^{\phi_2(K_{m,i+1})} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,i}}{x^{2/\rho}} \mid z\right) dx. \tag{3.15}$$

For $i = m$,

$$\int_{\phi_1(K_{m,m})}^{\phi_2(K_{m,m})} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,m}}{x^{2/\rho}} \mid z\right) dx$$

$$= \int_{\psi_1}^{\psi_2} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,m}}{x^{2/\rho}} \mid z\right) dx, \tag{3.16}$$

where $\phi_1(K_{m,1}) \le \phi_1(K_{m,2}) \le \cdots \le \phi_1(K_{m,m}) \le \psi_1$ and also for each $i = 1, \ldots, m$

$$\left(\frac{1}{w}\right)^2 f\left(\frac{1}{w} \mid z\right) F_0\left(\frac{K_{m,i}}{w^{2/\rho}} \mid z\right) > \left(\frac{1}{g_{\gamma_i,0}(w)}\right)^2 f\left(\frac{1}{g_{\gamma_i,0}(w)} \mid z\right) F_0\left(\frac{K_{m,i}}{\{g_{\gamma_i,0}(w)\}^{2/\rho}} \mid z\right)$$

$$\forall\, w \in [\, \phi_1(K_{m,i}), \phi_1(K_{m,i+1})\,] \tag{3.17}$$

where

$$\gamma_i = \int_{\phi_1(K_{m,i+1})}^{\phi_2(K_{m,i+1})} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,i}}{x^{2/\rho}} \mid z\right) dx. \tag{3.18}$$

If there is no $w \in [\, \phi_1(K_{m,i}), \phi_1(K_{m,i+1})\,]$ such that (3.17) is satisfied then we take $\phi_1(K_{m,i}) = \phi_1(K_{m,i+1})$.

The intervals $I_m(s, t, z, \underset{\sim}{K}_m)$ have minimum coverage probability equal to $1-\alpha$. It is not guaranteed that for a given m, the intervals are strictly shorter than $C_U(s, z)$ since we may have

$$P\{\, t \le K_{m,m} \,\} = 0, \tag{3.19}$$

or

$$F_0\left(\frac{K_{m,i}}{\psi_1^{2/\rho}} \mid z\right) = F_0\left(\frac{K_{m,i}}{\psi_2^{2/\rho}} \mid z\right) \tag{3.20}$$

for $i = 1, 2, \ldots, m$. In the first case the intervals $C_U(s, z)$ and $I_m(s, t, z, \underset{\sim}{K}_m)$ coincide, whereas in the second case the intervals have different endpoints but the lengths are the same. However, equation (3.19) cannot hold for every m, because it would mean that $P\{\ t < +\infty\ \} = 0$, since $\lim_{m\to\infty} K_{m,m} = +\infty$. On the other hand Lemma A.3 shows that we cannot have

$$F_0\left(\frac{K}{\psi_1{}^{2/\rho}} \mid z \right) = F_0\left(\frac{K}{\psi_2{}^{2/\rho}} \mid z \right) \tag{3.21}$$

for every K. Therefore, by filling $(0, +\infty)$ with cutoff points, we know that, eventually, the interval $I_m(s, t, z, \underset{\sim}{K}_m)$ will improve upon $C_U(s, z)$.

As $m \to \infty$ the endpoints of $I_m(s, t, z, \underset{\sim}{K}_m)$ tend to some functions $\phi_1(t)$ and $\phi_2(t)$. In order to determine $\phi_1(t)$ and $\phi_2(t)$ we work as in Goutis and Casella (1990), assuming that the function $F_0(x \mid z)$ is twice differentiable with second derivative bounded in finite intervals. We define

$$K_{m,i(t)} = \inf\{\ K \in \underset{\sim}{K}_m : K \geq t\ \}. \tag{3.22}$$

Then, for given t and s, the confidence interval at the mth stage is $(\phi_1(K_{m,i(t)})s^\rho,\ \phi_2(K_{m,i(t)})s^\rho)$ where the endpoints $\phi_1(K_{m,i(t)})$ and $\phi_2(K_{m,i(t)})$ satisfy

$$\int_{\phi_1(K_{m,i(t)})}^{\phi_2(K_{m,i(t)})} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,i(t)}}{x^{2/\rho}} \mid z \right) dx$$

$$= \int_{\phi_1(K_{m,i(t)+1})}^{\phi_2(K_{m,i(t)+1})} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{K_{m,i(t)}}{x^{2/\rho}} \mid z \right) dx. \tag{3.23}$$

Now use Taylor's series expansion of $F_0(\ K_{m,i(t)}/x^{2/\rho} \mid z\ )$ around $K_{m,i(t)+1})$, keeping the first two terms. By bringing the first term of the sum to the left, dividing both sides by $K_{m,i(t)}) - K_{m,i(t)+1})$ and taking the limit as $m \to \infty$. Equation (3.23) becomes

$$\frac{d}{dt}\left[\int_{\phi_1(t)}^{\phi_2(t)} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{t}{x^{2/\rho}} \mid z \right) dx \right]$$

$$= \int_{\phi_1(t)}^{\phi_2(t)} \left(\frac{1}{x}\right)^2 f\left(\frac{1}{x} \mid z\right) F_0\left(\frac{t}{x^{2/\rho}} \mid z\right) dx. \tag{3.24}$$

where f_0 denotes the derivative of F_0. Using Leibniz' formula equation (3.24) becomes

$$\frac{d\phi_1(t)}{dt}\left(\frac{1}{\phi_1(t)}\right)^2 f\left(\frac{1}{\phi_1(t)} \mid z\right) F_0\left(\frac{t}{\{\phi_1(t)\}^{2/\rho}} \mid z\right)$$

$$= \frac{d\phi_2(t)}{dt}\left(\frac{1}{\phi_2(t)}\right)^2 f\left(\frac{1}{\phi_2(t)} \mid z\right) F_0\left(\frac{t}{\{\phi_2(t)\}^{2/\rho}} \mid z\right). \tag{3.25}$$

Note that the last relation does not determine $\phi_1(t)$ and $\phi_2(t)$. In general we cannot specify another equation that uniquely defines the endpoints because requirement (3.17) does not uniquely determine $\mathbf{I_m}(s, t, z, \underset{\sim}{K}_m)$.

By the Lebesgue Dominated Convergence Theorem we know that the confidence coefficient of any interval constructed in this way is $1-\alpha$. We also saw that for every finite step m, possibly for m greater than some m_0, the interval $\mathbf{I_m}(s, t, z, \underset{\sim}{K}_m)$ is shorter than $C_U(s, z)$. However it is not clear what happens with the limiting interval. Only in special cases we can specify another equation that defines the limiting interval and make statements about its length.

4. An Example. We now illustrate the above method by constructing a location scale invariant confidence set and then improving upon it by no longer requiring scale invariance. Consider $X_1, X_2, \ldots, X_n$ identical independently distributed random variables, having exponential density with location parameter μ and scale parameter σ, that is

$$f_X(x) = \frac{1}{\sigma}\exp\{-(x-\mu)/\sigma\} \qquad \text{for } x \geq \mu. \tag{4.1}$$

By sufficiency the data reduce to $Y = \min\{X_1, \ldots X_n\}$ and $S = \sum_{i=2}^{n}(X_i - Y)$. The random variables Y and S are independent and the distribution of S does not depend on μ (Govindarajulu (1966)). It is straightforward to see that the distribution of Y is exponential with parameters μ and σ / n. However, the distribution of S is intractable for $n > 2$, so for illustration purposes we will consider only the case $n = 2$. Then $S = |X_1 - X_2|$ has also an exponential density with parameters 0 and σ. Hence the shortest $1-\alpha$ confidence interval for σ, based on s, alone has endpoints $\psi_1 s$ and $\psi_2 s$ determined by equations (2.5) and (2.7), which in this case are

$$\exp(-\frac{1}{\psi_2}) - \exp(-\frac{1}{\psi_1}) = 1-\alpha \tag{4.2}$$

and

$$\left(\frac{1}{\psi_2}\right)^2 \exp(-\frac{1}{\psi_2}) = \left(\frac{1}{\psi_1}\right)^2 \exp(-\frac{1}{\psi_1}) \tag{4.3}$$

When $t \leq K$ the interval $I_1(s,t,K)$ has endpoints which satisfy equation (3.2). We can calculate the integral in a closed form showing that the endpoints satisfy

$$\exp(-\frac{1}{\phi_2(K)}) - \frac{1}{1+2\sqrt{K}}\exp(-\frac{1+2\sqrt{K}}{\phi_2(K)}) - \exp(-\frac{1}{\phi_1(K)}) + \frac{1}{1+2\sqrt{K}}\exp(-\frac{1+2\sqrt{K}}{\phi_1(K)}) \tag{4.4}$$

$$= \exp(-\frac{1}{\psi_2}) - \frac{1}{1+2\sqrt{K}}\exp(-\frac{1+2\sqrt{K}}{\psi_2}) - \exp(-\frac{1}{\psi_1}) + \frac{1}{1+2\sqrt{K}}\exp(-\frac{1+2\sqrt{K}}{\psi_1})$$

and $\phi_1(K) \leq \psi_1$. Since in this case $F_\mu(y)/F_0(y)$ is increasing in y for $\mu > 0$, hence the interval $I_1(s,t,K)$ dominates $C_U(s)$ only for positive values of the location parameter.

The points $\phi_1^0(K)$ and $\phi_2^0(K)$, defined by (3.12), satisfy (4.4) and

$$\left(\frac{1}{\phi_1^0(K)}\right)^2 \exp(-\frac{1}{\phi_1^0(K)})\left(1-\exp(-\frac{2\sqrt{K}}{\phi_1^0(K)})\right) = \left(\frac{1}{\phi_2^0(K)}\right)^2 \exp(-\frac{1}{\phi_2^0(K)})\left(1-\exp(-\frac{2\sqrt{K}}{\phi_2^0(K)})\right) \tag{4.5}$$

Since the function $\left(\frac{1}{x}\right)^2 \exp(-\frac{1}{x})\left(1-\exp(-\frac{2\sqrt{K}}{x})\right)$ is unimodal $\phi_1^0(K)$ and $\phi_2^0(K)$ are unique, and for any $\phi_1(K) \geq \phi_1^0(K)$ the interval $I_1(s,t,K)$ has shorter length with positive probability.

Figure 1 exhibits the relative gain in length using only one cutoff point. We used the endpoints $\phi_1^0(K)$ and $\phi_2^0(K)$ which give the minimum length when only one cutoff point is used. For several values of the constant K, the coverage probability remained virtually the same. Previous calculations of relative risk for point estimation of a normal variance (Rukhin 1987) or numerical results for gains of intervals for a normal variance (Shorrock 1990, Goutis and Casella 1990) suggest that the improvement cannot be substantial. Recall, however, that we are only illustrating the case $n = 2$, and that somewhat greater improvement should be possible in other cases.

APPENDIX

LEMMA A.1. If a differentiable function f(x) defined on the real line has $f'(x) < 0$ whenever $f(x) = 0$ and there is an x_0 such that $f(x_0) = 0$, then f(x) is positive for $x < x_0$ and negative for $x > x_0$.

LEMMA A.2. Let f(x) be a nonnegative integrable function on the real line and γ a fixed constant smaller than $\int_{-\infty}^{\infty} f(x)\, dx$. Let

$$D = \{ w \mid \int_{w}^{\infty} f(x)\, dx > \gamma \} \tag{A.1}$$

and $g_\gamma(w) : D \rightarrow \Re$ be defined as the solution to the equation

$$\gamma = \int_{w}^{g_\gamma(w)} f(x)\, dx. \tag{A.2}$$

If f(x) is continuous and the set $E = \{ x \mid f(x) \neq 0 \}$ is connected then the function g_γ is differentiable and its derivative is equal to $f(w)/f\{g_\gamma(w)\}$. Furthermore, the derivative is continuous.

PROOF. Observe that the function $g_\gamma(w)$ is continuous. If it were not, since it is increasing, it must have a jump, that is

$$U = \lim_{w \downarrow w_0} g_\gamma(w) > \lim_{w \uparrow w_0} g_\gamma(w) = L \tag{A.3}$$

which implies that

$$\int_{L}^{U} f(x)\, dx = 0. \tag{A.4}$$

But the last relation contradicts the assumption that E is connected.

Fix $w_0 \in D$ and let $w < w_0$ Then

$$\int_{w_0}^{g_\gamma(w_0)} f(x)\, dx = \int_{w}^{g_\gamma(w)} f(x)\, dx = \gamma \tag{A.5}$$

implies

$$\int_{w}^{w_0} f(x)\,dx = \int_{g_\gamma(w)}^{g_\gamma(w_0)} f(x)\,dx. \tag{A.6}$$

If $M_1 = \sup\{ f(x) \mid w \le x \le w_0 \}$ and

$N_1 = \inf\{ f(x) \mid g_\gamma(w) \le x \le g_\gamma(w_0) \}$ then we have

$$\int_{w}^{w_0} f(x)\,dx \le M_1 (w_0 - w) \tag{A.7}$$

and

$$\int_{g_\gamma(w)}^{g_\gamma(w_0)} f(x)\,dx \ge N_1 \{ g_\gamma(w_0) - g_\gamma(w) \}. \tag{A.8}$$

The last two equations, together with (A.6), imply that

$$\frac{M_1}{N_1} \ge \frac{g_\gamma(w_0) - g_\gamma(w)}{w_0 - w}. \tag{A.9}$$

Note that N_1 is strictly positive because it is the infimum of the function f(x) over the closed interval $[\, g_\gamma(w), g_\gamma(w_0) \,]$ and f(x) is strictly positive for every $x \in [\, g_\gamma(w), g_\gamma(w_0) \,]$. If $f(x_0) = 0$ for some x_0, then, since the set E is connected, it would be zero for every $x > x_0$. That would imply that

$$\int_{x_0}^{\infty} f(x)\,dx = \int_{g_\gamma(w_0)}^{\infty} f(x)\,dx = 0 \tag{A.10}$$

which contradicts $w_0 \in D$. Hence the LHS of (A.9) is finite.

On the other hand if we define

$M_2 = \inf\{ f(x) \mid w \le x \le w_0 \}$ and

$N_2 = \sup\{ f(x) \mid g_\gamma(w) \le x \le g_\gamma(w_0) \}$ then

$$N_2 \{ g_\gamma(w_0) - g_\gamma(w) \} \ge \int_{g_\gamma(w)}^{g_\gamma(w_0)} f(x)\,dx = \int_{w}^{w_0} f(x)\,dx \ge M_2 (w_0 - w) \tag{A.11}$$

which implies

$$\frac{M_2}{N_2} \leq \frac{g_\gamma(w_0)-g_\gamma(w)}{w_0-w}. \tag{A.12}$$

Putting the equations (A.9) and (A.12) together

$$\frac{M_2}{N_2} \leq \frac{g_\gamma(w_0)-g_\gamma(w)}{w_0-w} \leq \frac{M_1}{N_1}. \tag{A.13}$$

Letting $w \rightarrow w_0$, because of the continuity of f(x) and $g_\gamma(w)$, we have

$$\lim_{w\rightarrow w_0} M_1 = \lim_{w\rightarrow w_0} M_2 = f(w_0) \tag{A.14}$$

and

$$\lim_{w\rightarrow w_0} N_1 = \lim_{w\rightarrow w_0} N_2 = f\{g_\gamma(w_0)\}. \tag{A.15}$$

From equations (A.13) – (A.15) we conclude that

$$\lim_{w\rightarrow w_0} \frac{g_\gamma(w_0)-g_\gamma(w)}{w_0-w} = \frac{f(w_0)}{f\{g_\gamma(w_0)\}}. \tag{A.16}$$

Repeating the argument for $w > w_0$ we obtain the same limit. Therefore the function $g_\gamma(w)$ is differentiable and its derivative is equal to $f(w)/f\{g_\gamma(w)\}$. The derivative is continuous as a ratio of two continuous functions. □

LEMMA A.3. Let F be a nondegenerate cumulative distribution function such that $F(0) = 0$. Then for every $x_1 < x_2$, there is a K such that $F(Kx_1) < F(Kx_2)$.

PROOF. Suppose that $F(Kx_1) = F(Kx_2)$ for every K. By letting $K = 1$ we have $F(x_1) = F(x_2)$. Since F is nondecreasing, $F(x) = F(x_1) = F(x_2)$ for every $x \in (x_1, x_2)$. Letting $K = x_1/x_2$ we get $F(x_1) = F(Kx_2) = F(Kx_1)$ therefore $F(x) = F(x_1) = F(x_2)$ for every $x \in (x_1^2 x_2, x_2)$. Repeating the same argument for $K = (x_1/x_2)^2, (x_1/x_2)^3 \ldots$ we can see that F must be constant for every $x \in ((x_1/x_2)^2, x_2), ((x_1/x_2)^3, x_2) \ldots$, hence for every $x < x_2$. In a similar way we can show that F must be constant for every $x > x_1$, that is, it is a degenerate distribution function. □

REFERENCES

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

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

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

GOUTIS, C. and CASELLA, G. (1990). Improved invariant confidence intervals for a normal variance. *Ann. Stat.* To appear.

GOVINDARAJULU, Z. (1966). Characterization of the exponential and power distributions. *Skand. Aktuarietidskr.* **49** 132-136.

MAATTA, J. M. and CASELLA, G. (1990). Developments in decision-theoretic variance estimation. *Stat. Science,* 5 90-120.

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

SHORROCK, G. (1990). Improved confidence intervals for a normal variance. *Ann. Statist.* 18 972-980.

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.

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.

Figure 1: Expected relative improvement in length for $1 - \alpha = 0.95$.