The Annals of Statistics
1990, Vol. 18, No. 2, 972–980

IMPROVED CONFIDENCE INTERVALS FOR A NORMAL VARIANCE

BY GLENN SHORROCK

Université du Québec à Montréal

The usual confidence interval for the variance σ^2 of a normal distribution is a function of the sample variance alone. In this paper, we construct confidence intervals for σ^2 that also depend on the sample mean. These intervals have the same length as the shortest interval depending only on the sample variance and have uniformly higher probability of coverage. The coverage probabilities of these intervals and others are compared.

1. Background. Stein (1964) constructed a point estimator for the normal variance σ^2 with smaller mean squared error than estimators based solely on $S^2 = \Sigma(X_i - \bar{X})^2$ [where X_i, $i = 1, 2, \dots, n$, are iid $N(\mu, \sigma^2)$]. His estimator effectively pooled the sample mean $\bar{X}$ and S^2 together whenever the small relative size of $\bar{X}^2$ indicated that the mean of the population was close to zero. In this way, he was sometimes able to gain the equivalent of another degree of freedom and so improve on the usual estimator.

More formally, Stein chose $\varphi = \varphi_0(\bar{X}^2/S^2)$ to minimize $E[(\varphi S^2 - \sigma^2)^2 | (\bar{X}^2/S^2), \mu = 0]$ for each value of $\bar{X}^2/S^2$. He showed that the point estimator $\min[\varphi_0(\bar{X}^2/S^2)S^2, S^2/(n+1)]$ has smaller expected squared error than $S^2/(n+1)$, uniformly in (μ, σ^2).

Brown (1968) extended Stein's result to a larger class of estimators and loss functions. In the squared error case, he chose $\varphi = \varphi_c$ to minimize, for fixed $c > 0$, the conditional expected loss $E[(\varphi S^2 - \sigma^2)^2 | (\bar{X}^2/S^2) \le c, \mu = 0]$ and proved that

$$\varphi_c(\bar{X}^2/S^2)S^2 = \begin{cases} \varphi_c S^2, & \text{if } \bar{X}^2/S^2 \le c, \\ S^2/(n+1), & \text{otherwise}, \end{cases}$$

has smaller expected squared error loss than $S^2/(n+1)$, uniformly in (μ, σ^2).

Because admissible estimators are usually Bayes rules or the limits of Bayes rules [see, for example, Brown (1971)], and these are usually analytic functions, it appears unlikely that Stein's point estimator is admissible. The inadmissibility of Brown's point estimator was shown in Brewster and Zidek (1974), who consider a "smooth" version of Brown's estimator: They chose $\varphi = \varphi(y)$ to minimize $E(L(\sigma^2, \varphi S^2) | \bar{X}^2/S^2 \le y, \mu = 0)$, where $L(\cdot, \cdot)$ is one of a class of bowl-shaped loss functions and y is the observed value of $\bar{X}^2/S^2$. This smooth estimator $\varphi(y)S^2$ is generalized Bayes and, when the loss function is squared error, it is admissible [Proskin (1985)].

For the interval estimator problem, Tate and Klett (1959) tabulated the endpoints of the shortest confidence interval for σ^2 that depends only on S^2.

Received August 1987; revised May 1989.
AMS 1980 *subject classifications*. Primary 62C99; secondary 62F25.
Key words and phrases. Generalized Bayes.

Cohen (1972), in a more general formulation that allowed the presence of several unknown means, used Brown-type estimators to construct confidence intervals with the same length as Tate and Klett's interval and uniformly higher probability of coverage. His intervals shifted the shortest interval slightly closer to zero whenever $\overline{X}^2/S^2 < c$, for some constant c.

The general form of loss function considered in Brewster and Zidek (1974), while not of the form considered by Cohen (1972), also applies to interval-estimation problems. Brewster (1972) illustrated this by constructing generalized-Bayes confidence intervals for two of these problems.

In this paper, we construct confidence intervals which improve upon the shortest interval for σ^2 that depends only on S^2. These intervals are analogues of the various point estimators previously mentioned; in each case, we use the same formulation of the problem, form of interval and loss function as Cohen (1972) and the same strategy for improvement: We shift the interval depending only on S^2 slightly closer to zero whenever the sample evidence indicates that μ itself is close to zero. The confidence interval analogue of the point estimator in Brewster and Zidek (1974) that we find is generalized Bayes among scale and orthogonal invariant intervals with the same length.

2. Introduction. Let

$$\mathbf{X}_1 = (X_1, X_2, \dots, X_n) \quad \text{and} \quad \mathbf{X}_2 = (X_{n+1}, X_{n+2}, \dots, X_{n+p})$$

be two independent random vectors where $\mathbf{X}_1 \sim \mathrm{MVN}_n(\mathbf{0}, \sigma^2 I_n)$ and $\mathbf{X}_2 \sim \mathrm{MVN}_p(\boldsymbol{\mu}, \sigma^2 I_p)$ with mean vector $\boldsymbol{\mu} = (\mu_{n+1}, \mu_{n+2}, \dots, \mu_{n+p})$.

Let $S^2 = \sum_{i=1}^{n} X_i^2$ and $T^2 = \sum_{i=n+1}^{n+p} X_i^2$. Then $S^2 \sim \sigma^2\chi_n^2$ and $T^2 \sim \sigma^2\chi_p^2(\eta)$ where $\eta = \|\boldsymbol{\mu}\|^2/\sigma^2$ and where χ_n^2 and $\chi_p^2(\eta)$ represent, respectively, a central chi-square distribution with n degrees of freedom and a noncentral chi-square with p degrees of freedom and noncentrality parameter η. A sufficient statistic for the unknown parameter $(\sigma^2, \boldsymbol{\mu})$ is $(S^2, \mathbf{X}_2)$.

The problem of finding good confidence intervals for σ^2 can be considered as a decision problem. We let $I(S^2, \mathbf{X}_2) = (\varphi_1(S^2, \mathbf{X}_2), \varphi_2(S^2, \mathbf{X}_2))$ be a confidence interval for σ^2 and say that the risk we incur in using $I(S^2, \mathbf{X}_2)$ is the probability of not covering the true value of σ^2, i.e.,

$$R((\sigma^2, \boldsymbol{\mu}), I) = P(\sigma^2 \notin I(S^2, \mathbf{X}_2)).$$

This decision problem remains invariant under the group G of transformations (k, Γ) that map

$$(S^2, \mathbf{X}_2) \to (k^2 S^2, k\Gamma\mathbf{X}_2),$$
$$(\sigma^2, \boldsymbol{\mu}) \to (k^2\sigma^2, k\Gamma\boldsymbol{\mu}),$$
$$\varphi_i(S^2, \mathbf{X}_2) \to k^2\varphi_i(S^2, \mathbf{X}_2), \qquad i = 1, 2,$$

where $k > 0$ is any positive real number and Γ is any $p \times p$ real orthogonal matrix. The invariant confidence intervals under this group have the form $(\varphi_1(Z^2)S^2, \varphi_2(Z^2)S^2)$, where φ_1 and φ_2 are positive functions and $Z^2 = T^2/(S^2 + T^2)$.

We will restrict attention to invariant confidence intervals for σ^2, depending jointly on S^2 and Z^2, that have the same length as the shortest confidence interval for σ^2 that is a function of S^2 alone. For a normal variance, the form of the shortest level $(1-\alpha)$ confidence interval that depends only on S^2 is $(S^2/a_n, S^2/b_n)$, where a_n and b_n are such that $\int_{a_n}^{b_n} f_n(x)\,dx = 1-\alpha$ and $f_{n+4}(a_n) = f_{n+4}(b_n)$ [$f_n(\cdot)$ denotes a chi-square density with n degrees of freedom]. Thus we consider intervals of the form $(\varphi(Z^2)S^2, (\varphi(Z^2)+c)S^2)$, where $c = 1/b_n - 1/a_n$ and φ is some positive function of Z^2.

Because the upper and lower limits of the invariant confidence interval depend on $(\sigma^2, \boldsymbol{\mu})$ only through the noncentrality parameter $\eta = \|\boldsymbol{\mu}\|^2/\sigma^2$, the risk of the invariant confidence interval also depends only on η and we may assume, without loss of generality, that $\sigma^2 = 1$. In order to reduce the problem to the consideration of central chi-square distributions only, we take advantage of a well-known property of the noncentral chi-square distribution and postulate the existence of a Poisson random variable K with mean $\eta/2$, such that the distribution of T^2 conditional on K is chi-square with $p+2K$ degrees of freedom. Because $Z^2 = T^2/(S^2+T^2)$, we have, conditional on K, that (T^2+S^2) and Z^2 are independent, that (T^2+S^2) is distributed as a central chi-square random variable with $n+p+2K$ degrees of freedom and that Z^2 is distributed as a beta random variable with parameters $p/2+K$ and $n/2$.

3. Construction of a Stein-type confidence interval. In this section, we see that if we choose the shifted interval that maximizes the conditional probability of covering σ^2, given $Z^2 = z^2$ and $K = k$, we are led to a confidence-interval analogue of the point estimator in Stein (1964).

We seek, for each value of Z^2 and K, that φ which maximizes

$$P\big(\varphi S^2 \le \sigma^2 \le (\varphi + c)S^2 \big| Z^2 = z^2, K = k\big). \tag{3.1}$$

This probability is proportional to

$$\int_{\varphi}^{\varphi+c} f_{n+p+2k+4}\left(\frac{1}{1-z^2}\,\frac{1}{y}\right) dy$$

and because the integrand of this last expression is a unimodal function of y, the maximizing $\varphi = \varphi(k, z^2)$ is the unique solution of

$$f_{n+p+2k+4}\left(\frac{1}{1-z^2}\,\frac{1}{\varphi}\right) = f_{n+p+2k+4}\left(\frac{1}{1-z^2}\,\frac{1}{\varphi+c}\right).$$

The Stein-like confidence interval is $I_S(S^2, Z^2)$, where $I_S(S^2, Z^2) = (\varphi_s(Z^2)S^2, (\varphi_s(Z^2)+c)S^2)$ and $\varphi_s(z^2) = \min[\varphi(0, z^2), 1/b_n]$.

THEOREM 3.1. *The coverage probability of the interval $I_S(S^2, Z^2)$ is uniformly greater than that of $(S^2/b_n, S^2/a_n)$, for all values of the noncentrality parameter, η.*

The theorem will be proved with the help of the following lemma, which is due to Brewster (personal communication).

LEMMA 3.1. *Let f, g be two unimodal densities and let $\varphi = \varphi_f$ maximize $\int_\varphi^{\varphi+c} f(x)\,dx$ and $\varphi = \varphi_g$ maximize $\int_\varphi^{\varphi+c} g(x)\,dx$. Then, if f/g is an increasing function, $\varphi_f > \varphi_g$.*

PROOF OF THEOREM 3.1. By Lemma 3.1 and the monotone likelihood property of the chi-square density, $\varphi((k, z^2)$ is a decreasing function of k for each fixed $z^2 > 0$ and so $\varphi(k, z^2) < \varphi(0, z^2)$. Also, by the unimodality of $f_{n+p+2k+4}(1/((1-z^2)y))$, the probability (3.1) increases as φ decreases to $\varphi(k, z^2)$. Using these facts and conditioning on K and Z^2, we see that

$$P(\sigma^2 \in I_S(S^2, Z^2)|Z^2 = z^2, K = k) > P(\sigma^2 \in (S^2/b_n, S^2/a_n)|Z^2 = z^2, K = k).$$

But $\varphi = 1/b_n$ maximizes $\int_\varphi^{\varphi+c} f_{n+4}(1/y)\,dy$, while $\varphi(0,0)$ maximizes $\int_\varphi^{\varphi+c} f_{n+p+4}(1/y)\,dy$. Thus, by Lemma 3.1, we have $\varphi(0,0) < 1/b_n$. However $\varphi(0, z^2)$ is continuous in z^2, so $\varphi(0, z^2) < 1/b_n$, for some z^2 lying in an interval to the right of zero. The result follows when we take expectations over Z^2 and K. □

4. A generalized-Bayes confidence interval. If we choose to find the interval that maximizes the conditional probability of covering σ^2, given $Z^2 \le r$ and $K = k$ [instead of the conditional probability (3.1)], we are led to a confidence-interval analogue of the point estimator in Brown (1968).

We let $\varphi = \varphi_0(r)$ maximize the conditional probability of coverage,

$$P\big(\varphi S^2 \le \sigma^2 \le (\varphi + c)S^2 | Z^2 \le r, K = 0\big), \tag{4.1}$$

for all r. This probability is proportional to

$$\int_\varphi^{\varphi+c} f_{n+4}\left(\frac{1}{y}\right) F_p\left(\frac{r}{1-r}\frac{1}{y}\right) dy \tag{4.2}$$

and the integrand of this last expression is a unimodal function of y. Thus, $\varphi_0(r)$ is the unique solution of the equation

$$f_{n+4}\left(\frac{1}{\varphi}\right) F_p\left(\frac{r}{1-r}\frac{1}{\varphi}\right) = f_{n+4}\left(\frac{1}{\varphi+c}\right) F_p\left(\frac{r}{1-r}\frac{1}{\varphi+c}\right).$$

The interval analogue of Brown's estimator is

$$I_1(S^2, Z^2; r) = \begin{cases} (S^2/b_n, S^2/a_n), & \text{if } Z^2 > r, \\ (\varphi_0(r)S^2, (\varphi_0(r) + c)S^2), & \text{if } Z^2 \le r. \end{cases}$$

When $r = 1$, it can be shown that this interval coincides with $(S^2/b_n, S^2/a_n)$. By conditioning on K and arguing similarly to Theorem 3.1, we can demonstrate the following result [full technical details of the proofs in this section can be found in Shorrock (1988)]:

THEOREM 4.1. *The coverage probability of the interval* $I_1(S^2, Z^2; r)$ *is uniformly greater than that of* $(S^2/b_n, S^2/a_n)$, *for all* $r \in (0, 1)$.

Both Theorems 3.1 and 4.1 have demonstrated the inadmissibility of the usual confidence interval for the normal variance $(S^2/b_n, S^2/a_n)$. Following the strategy of Brewster and Zidek (1974), the interval $I_1(S^2, Z^2; r)$ can in turn be improved upon by an interval with a second cutoff point, say $r_2 > r_1$, which takes different action depending on whether $Z^2 \le r_1$, $r_1 < Z^2 \le r_2$, or $Z^2 \ge r_2$.

Similarly, an interval with m cutoff points, say $I_m(S^2, Z^2; r_1, r_2, \dots, r_m)$, can be improved upon by adding one more cutoff point. This suggests taking the limit of confidence intervals $I_m(S^2, Z^2; r_1, r_2, \dots, r_m)$ as we fill up the interval $[0, 1]$ with cutoff points. The limiting interval that we obtain is $I_{\mathrm{BZ}}(S^2, Z^2) = [\varphi_0(Z^2)S^2, (\varphi_0(Z^2) + c)S^2]$, an analogue of the generalized-Bayes point estimator in Brewster and Zidek (1974). The following result is stronger than that given in Brewster and Zidek (1974); they show that the generalized-Bayes point estimator dominates the usual one but not that it dominates it strictly.

THEOREM 4.2. *The coverage probability of the interval* $I_{\mathrm{BZ}}(S^2, Z^2)$ *is strictly greater than that of* $[S^2/b_n, S^2/a_n]$, *for all* $\eta > 0$.

The proof of the theorem requires the following lemma:

LEMMA 4.1. *Let* $g(x)$ *be a bounded function of* x *such that* $|g(x)| \le B$ *and* $\int_0^\infty g(x) f_{p+2k}(x)\,dx = 0$, $\forall\, k \ge 0$. *Then* $g(x) \equiv 0$ *a.e. on* $(-\infty, \infty)$.

PROOF OF THEOREM 4.2. The endpoints of $I_m(S^2, Z^2; r_1, r_2, \dots, r_m)$ tend to those of $I_{\mathrm{BZ}}(S^2, Z^2)$ and so, by the bounded convergence theorem, for each k,

$$P\big(\sigma^2 \in I_{\mathrm{BZ}}(S^2, Z^2) | K = k\big) \ge P\big(\sigma^2 \in [S^2/b_n, S^2/a_n] | K = k\big). \tag{4.3}$$

Taking expectations over K, we have

$$P\big(\sigma^2 \in I_{\mathrm{BZ}}(S^2, Z^2) | \eta\big) \ge P\big(\sigma^2 \in [S^2/b_n, S^2/a_n] | \eta\big).$$

Suppose this inequality is not strict. Suppose there exists $\eta_0 > 0$ such that

$$P\big(\sigma^2 \in I_{\mathrm{BZ}}(S^2, Z^2) | \eta_0\big) = P\big(\sigma^2 \in [S^2/b_n, S^2/a_n] | \eta_0\big) = 1 - \alpha,$$

where $0 < \alpha < 1$. Now, for $\eta > 0$, $P(K = k|\eta) > 0$, $\forall\, k \ge 0$. Thus, by (4.3), $P(\sigma^2 \in I_{\mathrm{BZ}}(S^2, Z^2)|K = k) = 1 - \alpha$, $\forall\, k \ge 0$. But, using the conditional independence of S^2 and Z^2, we have

$$P\big(\sigma^2 \in I_{\mathrm{BZ}}(S^2, Z^2) | K = k\big) = \int \left\{ \int_{1/(u+c)}^{1/u} f_n(s^2)\,ds^2 \right\} f_{p+2k}(t^2)\,dt^2,$$

where $u = \varphi_0(t^2/(t^2 + s^2))$ and by Lemma 4.1, this implies that

$$\int_{1/(u+c)}^{1/u} f_n(s^2)\,ds^2 \equiv 1 - \alpha, \qquad \forall\, t^2 > 0.$$

Because $\varphi = 1/b_n$ is the unique solution of the equation

$$\int_{1/(\varphi+c)}^{1/\varphi} f_n(s^2)\,ds^2 = 1 - \alpha,$$

we have that $b_n \equiv \varphi_0(T^2/(T^2 + S^2)) = \varphi_0(Z^2)$. But this is a contradiction, because φ_0 is an increasing function. □

The confidence interval $I_{BZ}(S^2, Z^2)$ is also generalized Bayes among scale-and-orthogonal-invariant confidence intervals for σ^2 of the same length.

THEOREM 4.3. *Among all confidence intervals of the form* $[\varphi(Z^2)S^2, (\varphi(Z^2) + c)S^2]$, $I_{BZ}(S^2, Z^2)$ *is generalized Bayes with respect to the improper prior density*

$$\pi(\eta) = \frac{p}{4}\int_0^\infty e^{-\eta z/2}(1+z)^{-p/2}\,dz.$$

PROOF. By a change of variables, we have

$$\pi(\eta) = \beta\int_0^1 y^{\beta-2}e^{-t/y}\,dy, \quad \text{where } \beta = p/2. \tag{4.4}$$

Let X and W be iid exponential random variables with mean 1 and let $Y = \exp(-W/\beta)$. Then (4.4) is also the density of the product XY. Because the kth moment of XY is $k!/(1 + k/\beta)$, the same is true for $\pi(\eta)$. Using this, the marginal distribution $\pi(k)$ of K can be simplified to

$$\pi(k) = \frac{p/2}{k + p/2}.$$

While $\pi(k)$ is clearly improper, the posterior distribution of K given Z^2 is well defined, because

$$\sum_{k=0}^\infty f_{(Z^2|K)}(Z^2|k)\pi(k) = \frac{p}{2}\sum_{k=0}^\infty \frac{\Gamma((n+p)/2+k)}{\Gamma(p/2)}\,\frac{(Z^2)^{p/2+k-1}(1-Z^2)^{n/2-1}}{k+p/2},$$

which, by the ratio test, converges for $Z^2 < 1$.

Using formula 26.4.6 of Abramowitz and Stegun (1965) to expand the $F_p(\cdot)$ in (4.2), we can show that (4.2) equals

$$\int_{1/(\phi+c)}^{1/\phi}\sum_{k=0}^\infty f_{(S^2|Z^2,K)}(S^2|Z^2,K)\,f(Z^2|K=k)\pi(k)\,ds^2,$$

which, in turn, is equal to the posterior probability of coverage, given Z^2, of an interval of the form $(\varphi S^2, (\varphi + c)S^2)$. Thus, because I_{BZ} maximizes (4.2), it also maximizes the posterior probability of coverage and so it is generalized Bayes with respect to $\pi(\eta)$. □

5. Numerical results. Some sample probabilities of covering σ^2 were computed for $I_S(S^2, Z^2)$, the Stein-type confidence interval, and for

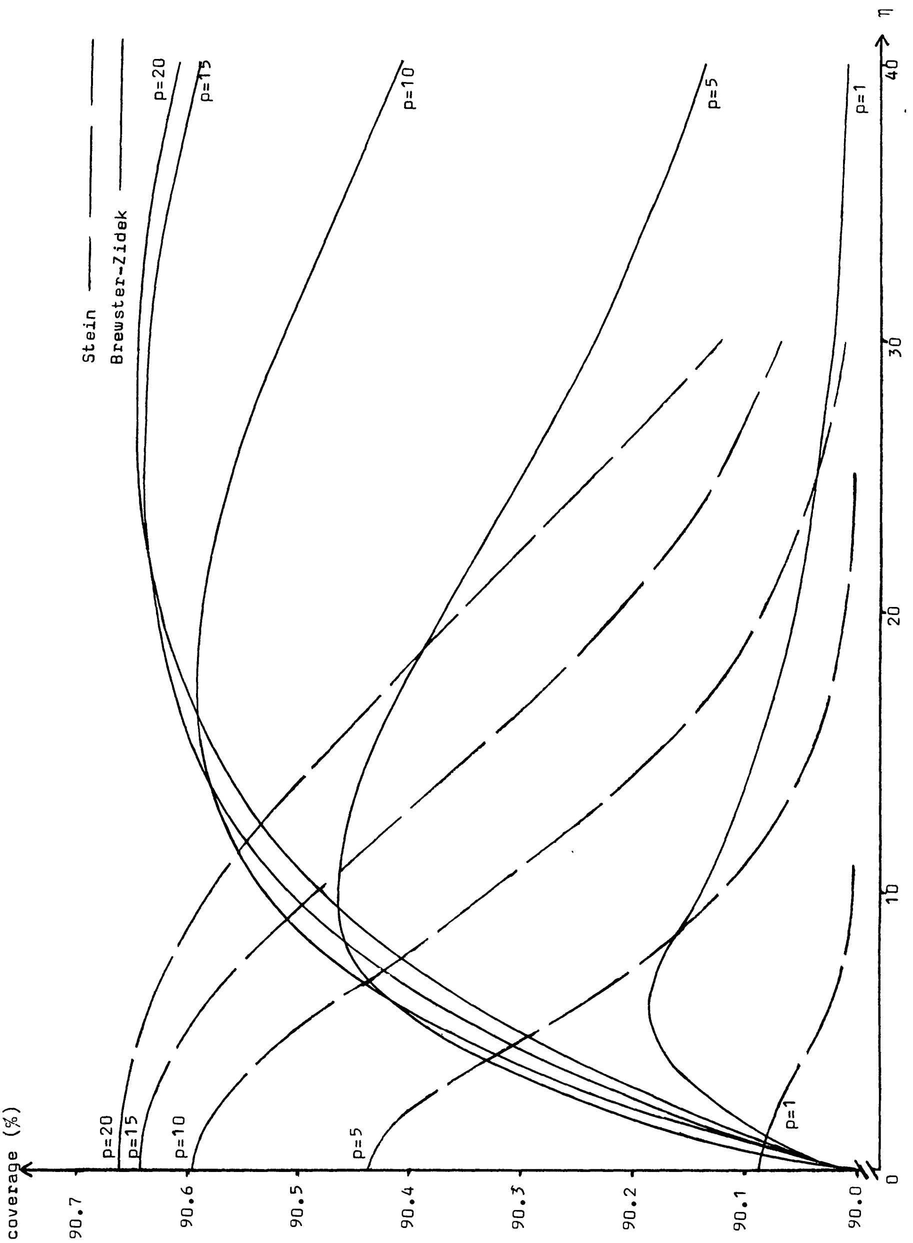

FIG. 1. *Probability of coverage for the Stein and Brewster–Zidek confidence intervals when* $(1 - \varepsilon) = 0.90$, $n = 10$ *and* $p = 5, 10, 15$ *and* 20.

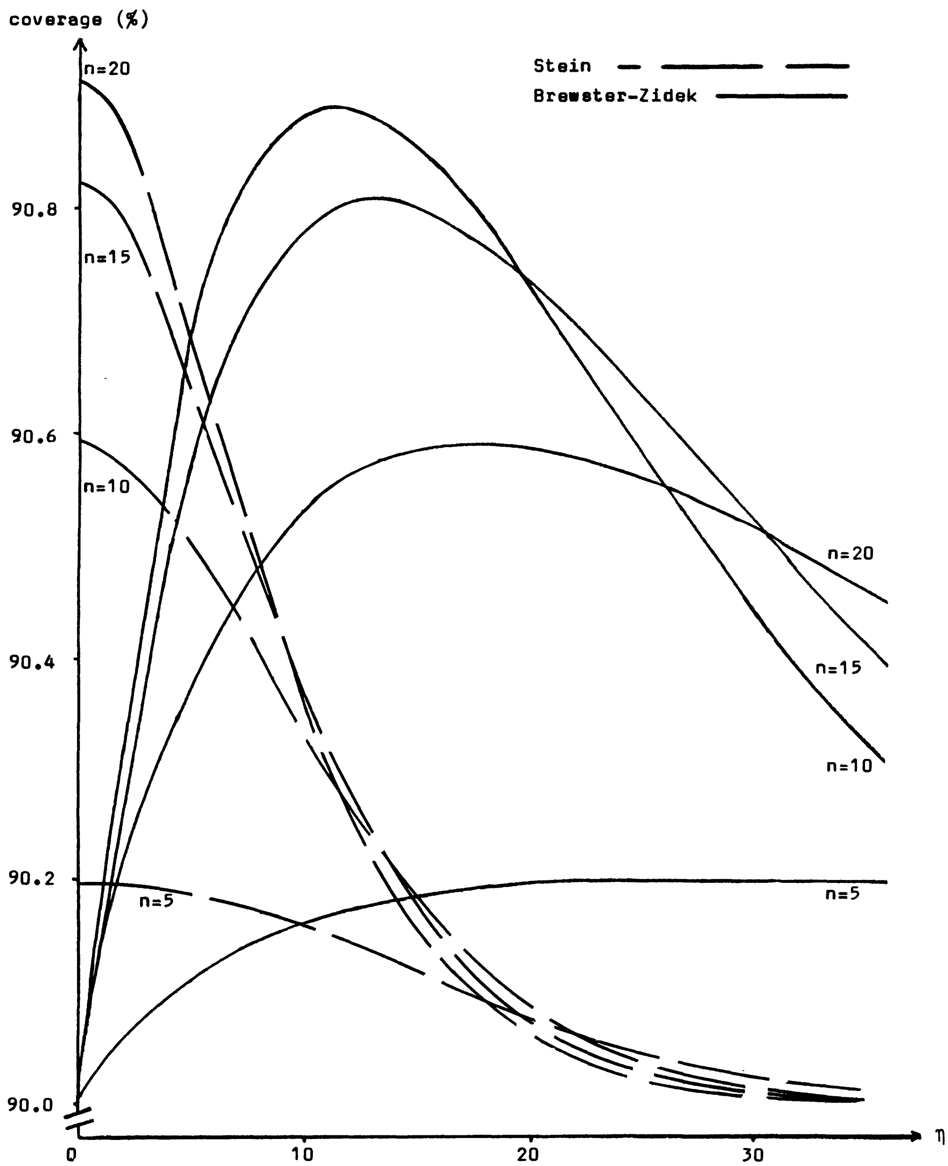

FIG. 2. *Probability of coverage for the Stein and Brewster–Zidek confidence intervals when* $(1 - \varepsilon) = 0.90$, $p = 10$ *and* $n = 5, 10, 15$ *and* 20.

$I_{BZ}(S^2, Z^2)$, the generalized-Bayes confidence interval, at a nominal $(1 - \varepsilon) = 0.90$ level. The results for $(1 - \varepsilon) = 0.95$ are qualitatively similar. Further comparisons of these intervals with the interval found in Cohen (1972) can be found in Shorrock (1988).

These computations were performed on a DEC VAX-11 using IMSL routines DCADRE and ZBRENT.

In Fig. 1, the probability of coverage of the intervals $I_S(S^2, Z^2)$ and $I_{BZ}(S^2, Z^2)$ is graphed for five different values of p, when $n = 10$.

Figure 2 is similar to Fig. 1, except that p is fixed at 10 and it is the central degrees of freedom n that are allowed to vary from 5 to 20 in steps of 5.

Comparing Fig. 1 with Fig. 2, it would appear that, while both the central and noncentral degrees of freedom play a role in determining the maximum improvement over $[S^2/b_n, S^2/a_n]$, the central degrees of freedom have a greater effect on the amount of improvement possible.

When $n = 29$ and $p = 29$, the maximum probability of coverage of the generalized-Bayes interval is 0.923, a 23% reduction in the probability of not covering the true variance. When $n = 29$ and $p = 29$, the Stein-type confidence interval, $I_S(S^2, Z^2)$, has a maximum probability of coverage of 0.917 and the interval given in Cohen (1972) is uniformly dominated by $I_S(S^2, Z^2)$.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs, and Math. Tables*. Dover, New York.

BREWSTER, J. F. (1972). On the admissibility of scale and quantile estimators. Ph.D. thesis, Dept. Math., Univ. of British Columbia.

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.

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

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

PROSKIN, H. M. (1985). An inadmissibility theorem with applications to the estimation of the variance of the normal distribution. Ph.D. dissertation, Dept. Statist., Rutgers Univ.

SHORROCK, G. (1988). Improved confidence intervals for a normal variance. Technical Report, Dept. Mathématiques et Informatique, Univ. Québec, Montréal.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distritution 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.

DÉPARTEMENT DE MATHÉMATIQUES
ET INFORMATIQUE
UNIVERSITÉ DU QUÉBEC À MONTRÉAL
C.P. 8888, SUCCURSALE "A"
MONTRÉAL, QUÉBEC
CANADA H3C 3P8