Joint Confidence Sets for the Mean and Variance of a Normal Distribution

Barry C. ARNOLD and Robert M. SHAVELLE

The standard example of an exact joint confidence set provided in elementary textbooks involves normal data. In addition textbooks provide a variety of asymptotic approximate joint confidence sets for (μ, σ^2) in such a setting. What is lacking, and what is provided in this article, is a comparative study of the size and actual performance of these confidence sets. A mild surprise arises in the comparison. If the sample size is 100 or more, the asymptotic confidence regions, in addition to being more robust to violations of distributional assumptions, actually outperform the exact region in terms of expected area of the regions.

KEY WORDS: Confidence region; Likelihood ratio test.

1. INTRODUCTION

Suppose that $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma^2)$, where $\mu \in \mathbf{R}$ and $\sigma^2 \in \mathbf{R}^+$. A $100(1-\gamma)\%$ confidence set (or region) for (μ, σ^2) is a random set $\mathcal{R}(\underline{X})$ such that

$$P((\mu, \sigma^2) \in \mathcal{R}(\underline{X})) = 1 - \gamma, \quad \forall (\mu, \sigma^2) . \qquad (1)$$

One of the earliest references to such multiparameter confidence regions is Cramer (1951). Specific examples in which exact regions are obtainable typically involve pivotal quantities (functions of data and of unknown parameters whose distributions are known and parameter free). The earliest precise description of an exact confidence set for (μ, σ^2) based on normal data appears in Mood (1950, p. 227). Mood gives no reference for the result, but also gives no indication that it was new to his textbook. Subsequently the exact confidence region appears quite regularly in elementary statistics textbooks; prior to 1950, it does not. It is not then unreasonable to christen it the Mood exact region.

A variety of large sample confidence sets may also be constructed. Most are based on the approximate multivariate normality of maximum likelihood estimates. A notable exception is the likelihood ratio test confidence region (see, for example, Meeker and Escobar 1995), which is based on an asymptotic chi-squared distribution. The Mood regions, although exact, make no pretense of being optimal with regards to expected area. Conversely, the maximum likelihood based approximate regions are asymptotically of smallest expected area (Kendall and Stuart 1979, p. 139). For fixed relatively small sample sizes, it is not a priori obvious which approach should be used. This article investigates this issue.

2. MOOD EXACT REGIONS

If $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma^2)$, then there are natural independent pivotal quantities

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \qquad (2)$$

and

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 / \sigma^2 \sim \chi^2_{n-1} . \qquad (3)$$

From standard tables we can find numbers a, b, and c such that

$$P(-a < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < a) = 1 - \alpha_1, \qquad (4)$$

and

$$P(b < \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} < c) = 1 - \alpha_2 . \qquad (5)$$

Specifically a will be the upper $\alpha_1/2$ percentile of a standard normal distribution and b and c will be lower and upper $\alpha_2/2$ percentiles of a χ^2_{n-1} distribution, respectively. [The flexibility in choice of α_1 and α_2 and subsequent allocation to upper and lower tails (which amounts to sliding or stretching the region) will be considered in Section 5.3.] Because the two pivotal quantities are independent we may write

$$\begin{aligned} (1-\alpha_1)(1-\alpha_2) &= P\left(-a < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < a, b < \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} < c\right) \\ &= P\left(\bar{X} - a\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + a\frac{\sigma}{\sqrt{n}}, \frac{\Sigma (X_i - \bar{X})^2}{c} < \sigma^2 < \frac{\Sigma (X_i - \bar{X})^2}{b}\right) . \end{aligned} \qquad (6)$$

Barry C. Arnold is Professor, and Robert M. Shavelle is Visiting Assistant Professor, Department of Statistics, University of California-Riverside, Riverside, CA 92521-0138 (E-mail: barnold@ucrstat2.ucr.edu). The authors thank the associate editor and three anonymous referees for unanimously suggesting that the likelihood ratio test confidence region be added to our original list. The manuscript owes much to their helpful and constructive comments and criticisms.

© 1998 American Statistical Association

Thus, a $100(1-\alpha_1)(1-\alpha_2)\%$ confidence set for (μ, σ^2) is the region $\mathcal{R}(\underline{X})$, where

$$\mathcal{R}(\underline{X}) = \left\{ (\mu, \sigma^2) : \bar{X} - a\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + a\frac{\sigma}{\sqrt{n}}, \frac{\Sigma(X_i - \bar{X})^2}{c} < \sigma^2 < \frac{\Sigma(X_i - \bar{X})^2}{b} \right\} . \quad (7)$$

Figure 1 shows the (roughly) trapezoidal shape of the Mood exact region. This region, and all the regions of the article, have been centered and scaled by setting $\bar{x} = 0$ and $s^2 = 1$. This may be done without loss of generality as the area of each region is independent of $\bar{x}$, and is a (constant) multiple of s^3 (see Section 5.3).

Note 1: The joint confidence region may be projected onto the vertical σ^2 axis to obtain a conservative $100(1-\alpha_1)(1-\alpha_2)\%$ confidence interval for σ^2, or similarly onto the horizontal axis for μ (see Nickerson [1994], where this is done for regression coefficients, or Hochberg and Tamhane [1987] for a general approach assuming normality and known variance for large sample estimates).

Note 2: For purposes of completeness, another exact method (Wilks 1962, p. 383) deserves mention. Because $W = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2_n$, we have that $\{(\mu, \sigma^2) : W \leq \chi^2_{n,1-\gamma}\}$ is a $100(1-\gamma)\%$ confidence region for (μ, σ^2). See Figure 2, where again $\bar{x} = 0$ and $s^2 = 1$. As pointed out by Wilks, however, the region is unsatisfactory because it is unbounded, and will not be considered further.

A caveat is in order regarding the use of such exact confidence regions. The analysis justifying their use is highly dependent on the normality assumption. If normality is doubtful, or the sample size is sufficiently large to dispense with the assumption, we might try to approximate the joint confidence set by using the large sample properties of the maximum likelihood estimators, $\bar{X}$ and S^2, estimates which can be expected to be reasonably robust to minor violations of the normality assumption. In addition of course, as mentioned in the introduction, even under an assumption of normality, there is no optimality claim advanced for Mood exact regions and the large sample regions may still fare well in comparison. This article will compare the shape, area, and performance (true confidence level) of four such large sample approximations to the corresponding aspects of the Mood exact region.

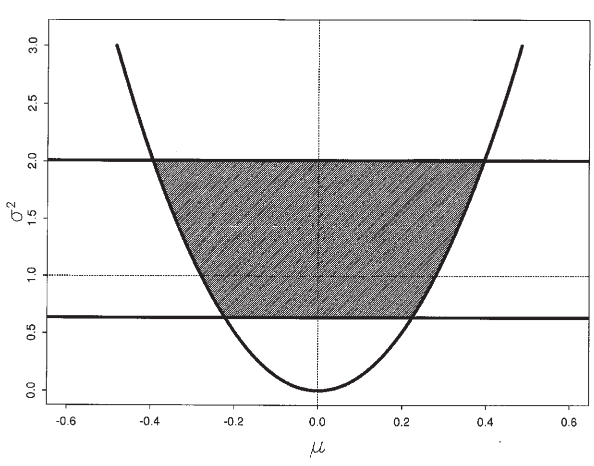

Figure 1. The Mood Exact Region for γ = :10 and n = 25. Without loss of generality, $\bar{x}$= 0 and s^2 = 1. Note that the vertical and horizontal scales are not equal.

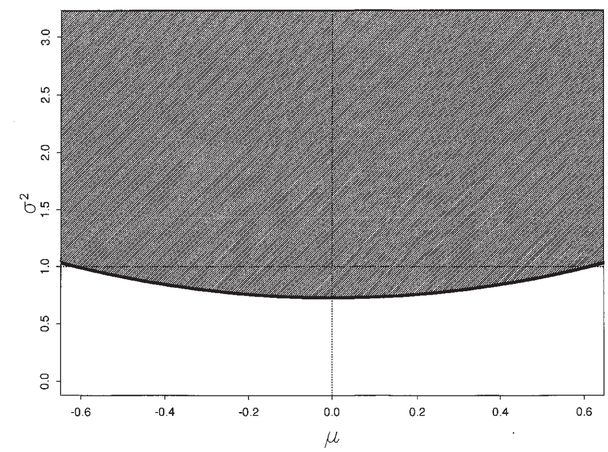

Figure 2. Wilks' Unbounded Exact Region for γ = :10 and n = 25. Without loss of generality, $\bar{x}$= 0 and s^2 = 1. Note that the vertical and horizontal scales are not equal.

3. APPROXIMATE REGIONS IN GENERAL

It is well known (see, for example, Mood 1950, p. 211) that the multivariate MLE, $\hat{\underline{\theta}}_{(n)}$, for $\underline{\theta}$ a $(k \times 1)$ vector of parameters, is asymptotically normal in the sense that

$$\hat{\underline{\theta}}_{(n)} \dot{\sim} N^{(k)}\left(\underline{\theta}, \frac{1}{n}\Sigma(\underline{\theta})\right), \quad (8)$$

where $\Sigma^{-1}(\underline{\theta}) = (\sigma^{ij}(\underline{\theta}))$ in which $\sigma^{ij}(\underline{\theta}) \equiv -E[\frac{\partial^2}{\partial\theta_i\partial\theta_j}(\log\ f(\underline{X};\underline{\theta}))]$. In (8) and henceforth the symbol $\dot{\sim}$ denotes "is approximately distributed as." From (8) it follows that the associated quadratic form

$$U = \Sigma_{i=1}^{k}\Sigma_{j=1}^{k} n\sigma^{ij}(\underline{\theta})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \quad (9)$$

has, for large n, an approximate chi-square distribution with k degrees of freedom, i.e. $U \dot{\sim} \chi^2_k$. Expression U in (9) is thus an approximate pivotal quantity for $\underline{\theta}$, for large n. [The optimality properties of such large sample regions were first investigated by Wilks and Daly (1939) and Bartlett (1955), and later detailed by Wilks (1962).]

Because $\hat{\underline{\theta}}$ is a strongly consistent estimate of $\underline{\theta}$, substitution of $\hat{\underline{\theta}}$ for $\underline{\theta}$ in the expression $\sigma^{ij}(\underline{\theta})$ in (9) will still yield a quantity that has an asymptotic chi-square distribution. The approximation is a little cruder but the resulting approximate pivotal quantity is more easily "inverted" to yield ellipsoidal confidence sets. This more crude approximate pivotal is then

$$V = \sum_{i=1}^{k}\sum_{j=1}^{k} n\sigma^{ij}(\hat{\underline{\theta}})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \dot{\sim} \chi^2_k . \quad (10)$$

Another approximate confidence region is one based on the likelihood ratio test statistic. If we have n observations

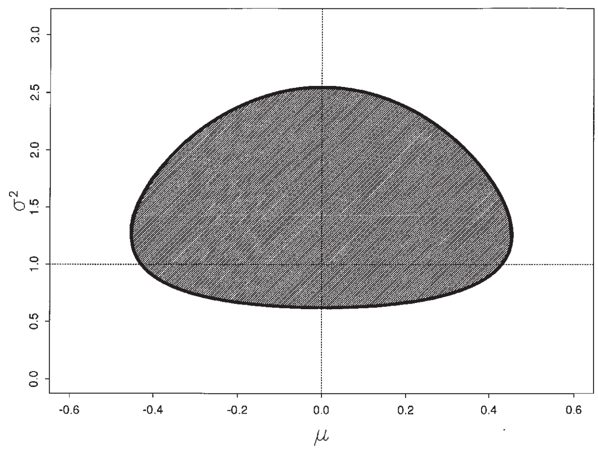

Figure 3. The Large Sample Region for γ =.10 and n = 25. Without loss of generality, $\bar{x}$= 0 and s^2 = 1. Note that the vertical and horizontal scales are not equal.

$\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ with common density $f(\underline{x};\underline{\theta})$ in which $\underline{\theta}$ is k-dimensional, we may test $H : \underline{\theta} = \underline{\theta}_o$ by considering the likelihood ratio test statistic

$$R_n(\underline{\theta}_o) = \frac{L_n(\underline{\theta}_o)}{L_n(\hat{\underline{\theta}})}, \tag{11}$$

where $L_n(\underline{\theta}) = \Pi_{i=1}^n f(\underline{X}_i, \underline{\theta})$ denotes the likelihood function and $\hat{\underline{\theta}}$ denotes the maximum likelihood estimate of $\underline{\theta}$ based on $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$. It is well known (see, for example, Kendall and Stuart 1979) that when $\underline{\theta} = \underline{\theta}_o$, the quantity $-2\log R_n(\underline{\theta}_o)$ has, for large n, an approximate χ^2 distribution with k degrees of freedom. It then follows that the quantity $-2\log R_n(\underline{\theta})$ is asymptotically a pivotal quantity, and consequently an approximate $100(1-\gamma)\%$ confidence region for $\underline{\theta}$ is given by $\{\underline{\theta} : -2\log R_n(\underline{\theta}) < \chi^2_{k,1-\gamma}\}$. Meeker and Escobar (1995) reported that such approximate confidence regions perform well in a variety of contexts. As we shall see in the following, the technique is quite effective in the normal case that is the focus of the present article.

4. APPROXIMATE REGIONS FOR THE MEAN AND VARIANCE OF A NORMAL DISTRIBUTION

4.1 The Basic Large Sample Region

In the case of univariate normal data there are $k = 2$ parameters. Let μ and σ^2 denote the mean and variance respectively. Then, as described in Section 3,

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} \dot{\sim} N^{(2)} \left[\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right], \tag{12}$$

and

$$U = \frac{n}{\sigma^2}(\hat{\mu} - \mu)^2 + \frac{n}{2\sigma^4}(\hat{\sigma}^2 - \sigma^2)^2 \dot{\sim} \chi^2_2 . \tag{13}$$

Or, writing this in terms of more transparent notation for the MLEs,

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = S^2 = \frac{1}{n}\Sigma(X_i - \bar{X})^2 , \tag{14}$$

we have

$$U = \frac{n}{\sigma^2}(\bar{X} - \mu)^2 + \frac{n}{2\sigma^4}(S^2 - \sigma^2)^2 \dot{\sim} \chi^2_2 . \tag{15}$$

Then, for large n, $\{(\mu, \sigma^2) : U < \chi^2_{2,1-\gamma}\}$ is an approximate $100(1-\gamma)\%$ confidence set for (μ, σ^2). Figure 3 shows the ocular shape of this approximate joint confidence region. This is the first approximate confidence set to be considered; the following two are simple variations on it.

4.2 The Large Sample Region with Plug-in Values for the Asymptotic Variances

The second approximate confidence region uses plug-in values (S^2 and $2S^4$, respectively) for the asymptotic variances, σ^2 and $2\sigma^4$, obtaining

$$V = \frac{n}{S^2}(\bar{X} - \mu)^2 + \frac{n}{2S^4}(S^2 - \sigma^2)^2. \tag{16}$$

Then, as discussed in Section 3, since $S^2 \xrightarrow{\text{a.s.}} \sigma^2$, it follows that $V \dot{\sim} \chi^2_2$. Thus, for large enough n, $\{(\mu, \sigma^2) : V < \chi^2_{2,1-\gamma}\}$ is an approximate $100(1-\gamma)\%$ confidence set for (μ, σ^2).

The payoff for using plug-ins (i.e., using V instead of U) is that the awkward region obtained using (15) is replaced by a region whose boundary is the equation of an ellipse (16) centered at $(\bar{X}, S^2)$ with major axes parallel to the coordinate axes (see Fig. 4).

4.3 The Large Sample Region with Plug-in Values for the Asymptotic Variances and an F Distribution

This third approximation to the exact region is merely a modification of the second. If plug-ins are used for the asymptotic variances, it might be prudent to change the asymptotic distribution from χ^2_2 to $2F_{2,n-2}$ (Douglas 1993). The intent is to balance out the loss of accuracy due to the substitution (using plug-ins) with a corresponding increase in the size of the region. The resulting region will always be larger than the second approximation as $kF_{k,v,1-\gamma} \geq \chi^2_{k,1-\gamma}$ for all k and v, and for $\gamma \geq .50$. For comparison, this confidence set is the larger of the two ellipses of Figure 4.

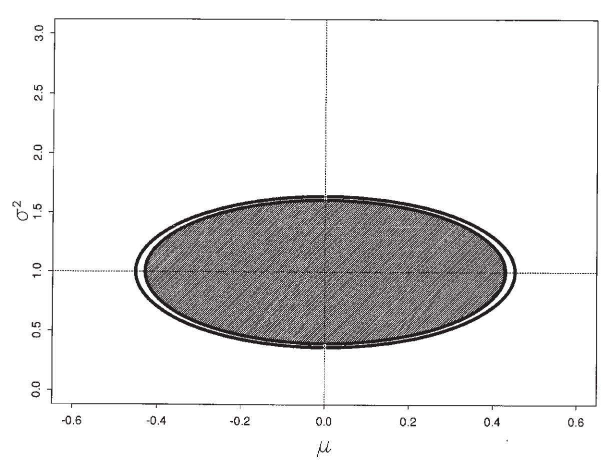

Figure 4. The Large Sample Regions with Plug-ins for γ = .10 and n = 25. Without loss of generality, $\bar{x}$= 0 and s^2 = 1. The smaller ellipse is based on the critical value $\chi^2_{2;.90}$, and the larger on the critical value $2F_{2;23;.90}$. Note that the vertical and horizontal scales are not equal.

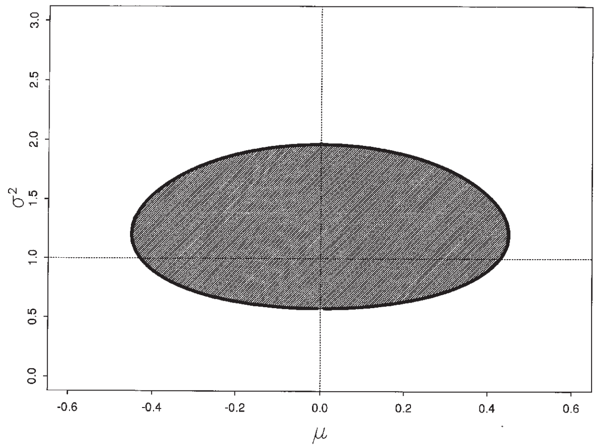

Figure 5. The Likelihood Ratio Test Confidence Region for γ = .10 and n =25. Without loss of generality, $\bar{x}$=0 and s^2 =1. Note that the vertical and horizontal scales are not equal.

4.4 The Likelihood Ratio Test Confidence Region

For the case in which our sample is from a normal(μ, σ^2) distribution, the likelihood ratio test statistic, equation (11), simplifies to yield

$$R_n(\mu, \sigma^2) = (\frac{\hat{\sigma}^2}{\sigma^2})^{n/2}\exp\left[\frac{1}{2\hat{\sigma}^2}\Sigma_{i=1}^n(X_i - \hat{\mu})^2 - \frac{1}{2\sigma^2}\Sigma_{i=1}^n(X_i - \mu)^2\right] . \quad (17)$$

Because $\hat{\sigma^2} = S^2$ and $\hat{\mu} = \bar{X}$, our asymptotic pivotal quantity is expressed as

$$-2\log R_n(\mu, \sigma^2) \; = \; n\log\frac{\sigma^2}{S^2} + \frac{nS^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} - n \; . \quad (18)$$

As discussed in Section 3, $-2\log R_n(\mu, \sigma^2) \dot{\sim} \chi^2_2$. It follows that for large n, $\{(\mu, \sigma^2) : n\log\frac{\sigma^2}{S^2} + \frac{nS^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} - n < \chi^2_{2,1-\gamma}\}$ is an approximate $100(1 - \gamma)\%$ confidence set for (μ, σ^2). An illustration of this region is given in Figure 5.

Note: Likelihood ratio test confidence regions are asymptotically identical to a Bayesian credibility region (Press 1989, p. 29) obtained using a uniform prior $(P(\mu, \sigma^2) \propto 1)$ for (μ, σ^2).

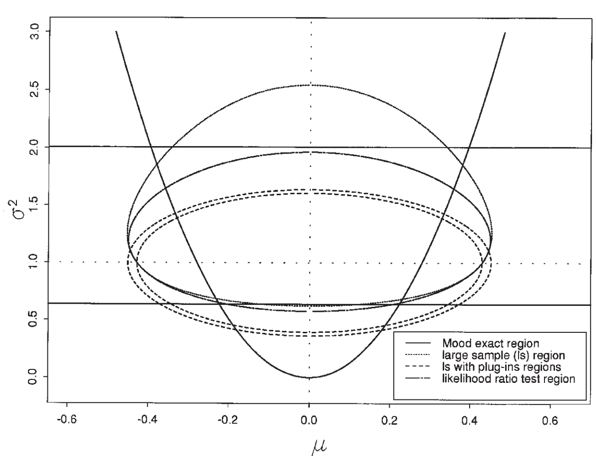

Figure 6. All Five Confidence Regions for γ = .10 and n = 25. Without loss of generality, $\bar{x}$= 0 and s^2 = 1. The horizontal reference lines are located at $\Sigma(x_i - \bar{x})^2/\chi^2_{24;:026}$ and $\Sigma(x_i - \bar{x})^2/\chi^2_{24;:974}$. Note that the vertical and horizontal scales are not equal.

5. COMPARISON OF THE REGIONS

The shape, confidence level, and areas of the previously discussed five joint confidence regions are now compared. This is done for sample sizes of 10, 25, and 100 and for overall nominal level of confidence .90.

5.1 Shape

For a sample of size $n = 25$ the methods outlined previously were used to produce the five nominal 90% confidence regions shown in Figure 6. Without loss of generality, we have let $\bar{x} = 0$ and $s^2 = 1$, and preliminarily have let $1 - \alpha_1 = 1 - \alpha_2 = \sqrt{(1 - .10)}$. In view of the graph, a few observations are in order. First, the large sample region seems to coincide with the lower horizontal line because of the approximate identity

$$\frac{\chi^2_{n-1,.974}}{n} \approx 1 + \sqrt{\frac{2\chi^2_{2,.90}}{n}}$$

as can be verified by a table scan of the χ^2 distribution. Second, the horizontal and vertical locations of all five regions are independently determined by the realized values of the MLE's $\bar{X}$ and S^2, respectively. Lastly, all five regions are horizontally centered about $\bar{x}$, and the large sample regions containing plug-ins for the asymptotic variances (the two ellipses) are vertically centered about s^2. The Mood exact region is not vertically centered about s^2 due to the skewness of the χ^2 distribution for small degrees of freedom.

To demonstrate the effect of sample size, the five 90% confidence regions for $n = 100$ are shown in Figure 7. Again, without loss of generality, we have taken $\bar{x} = 0$ and $s^2 = 1$, and preliminarily have let $1 - \alpha_1 = 1 - \alpha_2 = \sqrt{(1 - .10)}$. The four large sample regions are remarkably similar in size and shape, and closely approximate the Mood exact region; such could be expected knowing the large sample properties of the MLE's. For sample sizes of 200 or more, the four approximate regions are nearly indistinguishable from one another, and cover all but the "corners" of the Mood exact region.

5.2 Confidence

Although it is true that the nominal confidence level for each of the four approximate regions was set at .90, the actual confidence level obtained in using them is of more interest. Naturally, as the sample size becomes "large," the large sample approximations will become, for all practical purposes, exact. However, for small to moderate sample sizes, the actual and nominal confidence levels of the asymptotic regions may differ substantially.

We construct each of the four approximate regions for samples of size 10, 25, and 100 from simulated normal data. For reference purposes, the Mood exact region is also

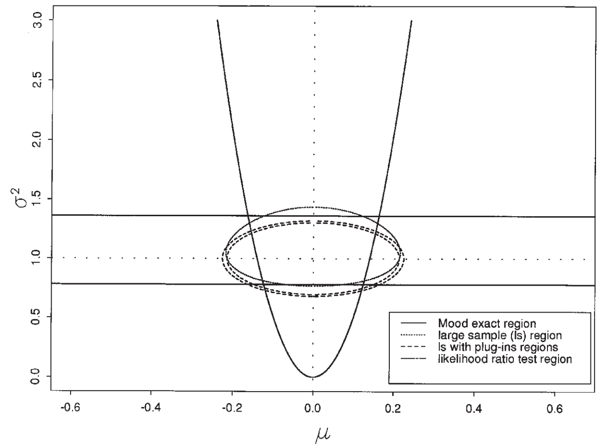

Figure 7. All Five Confidence Regions for γ = .10 and n = 100. Without loss of generality, $\bar{x}=0$ and s^2 =1. Note that the vertical and horizontal scales are not equal.

constructed. Because for simulated data the true mean and variance are known, it can be checked if each region "covers" this point. Then, the procedure may be repeated and the observed fraction of times that each region contains the true value may be compared to the expected (theoretical) fraction. Although the "confidence" is on the procedure, and not on the (fixed) parameter pair (μ, σ^2), the view here is clearly to long-run probabilities (though, see Kalbfleisch (1985) or Fieller (1954) for a fiducial argument).

Preliminarily, we let $1-\alpha_1 = 1-\alpha_2 = \sqrt{(1-.10)}$ and construct the $100(1-\alpha_1)(1-\alpha_2)\% = 90\%$ Mood exact region and four nominal 90% approximate regions. Each sample of size 10, 25, or 100 consisted of double precision pseudorandom normal deviates with mean 0 and variance 1 (generated using subroutine DRNNOR of the IMSL Statistics/Library, which has a period of over 2 billion). The standard error of the fraction of successes (coverage percentage) in one billion Bernoulli trials is approximately .000009, which (by taking ± 2 SE's) shows that four-decimal-place accuracy should be obtained. Results of the simulations are shown in Table 1.

It is indeed remarkable that the large sample region has true confidence level very close to its nominal level for sample sizes 25 and 100; and, even for a sample size as low as 10, the true confidence level is within 2% of the nominal level. Furthermore, for all three sample sizes the true confidence level of the large sample region actually exceeds the nominal level.

Table 1. Fraction of Successes (the true parameter pair lies in the joint confidence region) in One Billion Trials

	Sample size (n)		
Region type	*10*	*25*	*100*
Mood exact	.9000	.9000	.9000
large sample (ls)	.9170	.9079	.9020
ls with plug-ins (lspi)	.7496	.8321	.8819
lspi and *F* statistic (lspif)	.8040	.8521	.8870
likelihood ratio test	.8760	.8911	.8978
nominal confidence level	.9000	.9000	.9000

Turning to consider the large sample plug-ins region, we find that, despite our glib statement that plugging in a strongly consistent estimate should not cause problems, the true situation for sample sizes 10 and 25 is disappointing. The true confidence levels are markedly smaller than the nominal levels; that is, the χ^2 critical values used were too small. It is clear that use of the F statistic does improve matters, but there is still a displeasingly low true confidence level associated with such regions.

The likelihood ratio test region is seen to be not much worse than the large sample region, though it has a true confidence level somewhat lower than the nominal level.

5.3 Area

For sample size n and desired joint confidence 90%, areas of the five regions may be computed. For simplicity, let $1-\alpha_1 = 1-\alpha_2 = \sqrt{(1-\gamma)}$, where $\gamma = .10$. It can be shown that the areas are as follows:
Mood exact region:

$$\text{area} = \frac{4nz_{.026}}{3}\left[\left(\frac{1}{\chi^2_{n-1,.026}}\right)^{3/2} - \left(\frac{1}{\chi^2_{n-1,.974}}\right)^{3/2}\right] S^3. \tag{19}$$

Large sample (ocular) region:

$$\text{area} = \left[\int_a^b \sqrt{\frac{4\chi^2_{2,.90}u^2 - 2n(1-u)^2}{nu}}\, du\right] S^3,$$
$$\text{where} \quad a, b = \frac{n \mp \sqrt{2n\chi^2_{2,.90}}}{n - 2\chi^2_{2,.90}} \tag{20}$$

Large sample region with plug-ins for the asymptotic variances (ellipse):

$$\text{area} = \frac{\sqrt{2}\pi\chi^2_{2,.90}}{n} S^3. \tag{21}$$

Large sample region with plug-ins for the asymptotic variances and using an F statistic (larger ellipse):

$$\text{area} = \frac{\sqrt{2}\pi 2F_{2,n-2,.90}}{n} S^3. \tag{22}$$

Likelihood ratio test confidence region:

$$\text{area} = \left[\int 2\sqrt{\frac{\chi^2_{2,.90}u}{n} + u - 1 - u ln(u)}\, du\right] S^3, \tag{23}$$

where the integral is over values for which the expression under the radical is positive.

It is evident, and retrospectively obvious, that the areas of the above five regions do not depend on the realized values of $\bar{X}$ (just a location parameter) and that all five areas can be expressed as a constant multiple (which depends only on n and the nominal confidence level) of S^3.

Table 2. Areas (area/S^3) of Confidence Regions of Nominal Level .90

Region type	Sample size (n) 10	25	100
Mood exact	5.47	1.21	.2335
large sample (ls)	29.90	1.35	.2272
ls with plug-ins (lspi)	2.05	.82	.2046
lspi and F statistic (lspif)	2.77	.91	.2095
likelihood ratio test	3.24	.98	.2139

Thus, conveniently, it is sufficient to consider the ratio, area/S^3, for each of the five regions. Table 2 does so for samples of size 10, 25, and 100.

Given the discrepancy between the true and nominal confidence levels for the large sample regions detailed in Section 5.2, a comparison of the areas in Table 2 is misleading. The requisite comparison is of the sizes of true 90% confidence regions. By experimentation, one may find the critical values for each of the large sample approximations such that the true confidence in using them is 90%. To obtain the desired 90% confidence using the large sample region, corresponding χ^2_2 percentiles of 88.0, 89.2, and 89.7 for sample sizes 10, 25, and 100 respectively, would need to be used. For the large sample region with plug-ins, χ^2_2 percentiles of 99.8, 97.0, and 92.0, respectively, are required (or the analogous F statistic percentiles). Last, for the likelihood ratio test region, the χ^2_2 percentiles are 92.1, 90.9, and 90.2, respectively.

Note: It may be observed that the statistics used in constructing of large sample regions appearing in Equations (15), (16), and (18) are all pivotal quantities; that is, their distributions for a given true value of (μ, σ^2) are independent of (μ, σ^2). They do not have χ^2_2 distributions; their exact distributions depend on the sample size, n. Percentiles of the true distributions of pivotal quantities (15), (16), and (18) could be obtained via simulation for a given n. In the previous paragraph we presented results equivalent to such a determination for $n = 10$, 25, and 100 and for a 90% level of confidence.

In addition, it is possible that we have unfairly handicapped the Mood exact procedure by letting $\alpha_1 = \alpha_2$ and by using equal tail precisions in the χ^2_{n-1} cutoffs b and c (refer to (5)). The general area formula for the Mood exact region (of which (19) is a special case) is:

$$\text{area} = \frac{4nz_{\alpha_1/2}}{3}\left[\left(\frac{1}{\chi^2_{n-1,\delta}}\right)^{\frac{3}{2}} - \left(\frac{1}{\chi^2_{n-1,\alpha_2-\delta}}\right)^{\frac{3}{2}}\right] S^3, \tag{24}$$

where δ is the portion of α_2 which is put in the lower tail of the chi-squared distribution.

By an exhaustive search, the choices of α_1, α_2, and δ to produce a Mood exact 90% confidence set of smallest possible size may be found. Table 3 compares the smallest area based on optimal choice of α_1, α_2, and δ with the naive choice of $\alpha_1 = \alpha_2$ and equal tails. It is seen that a significantly smaller area may be obtained by using optimal values; and that the disparity between areas decreases with increasing sample size (due to the asymptotic normality of the χ^2 distribution).

Table 3. Comparison of Mood Exact Regions for $\gamma = .10$. The first entry (row) for each sample size is the equal allocation and equal tails region; the second is the optimal allocation and optimal tails region.

n	Area/S^3	α_1	α_2	δ
10	5.4745	.051	.051	.026
	3.8302	.027	.076	.074
25	1.2146	.051	.051	.026
	1.0594	.039	.064	.059
100	0.2335	.051	.051	.026
	0.2258	.048	.055	.042
1000	0.0217	.051	.051	.026
	0.0216	.051	.052	.031

For the benefit of the practitioner, we also include Table 4, which lists the optimal choices of α_1, α_2, and δ to produce smallest exact joint confidence sets of various confidence levels and sample sizes. What is immediately obvious from Tables 3 and 4 is that for $n = 1{,}000$ the optimal choices of α_1 and α_2 are approximately equal, but for $n \leq 100$ it is far from optimal to choose $\delta = \alpha_2/2$ (as is usually done).

Recall that previously we found the appropriate critical values to yield large sample regions with true 90% confidence. In addition, we have just found the optimal values to produce Mood exact regions of the smallest possible area. Thus, we are now in a position to fairly compare the areas of five true 90% confidence regions. Table 5 shows the results.

The first thing to be noticed is that the large sample region for sample size 10 is inordinately large; perhaps this is due to the inherent difficulty in estimating σ^2 from small

Table 4. Allocations of α_1, α_2, and δ to Produce Mood Exact Regions of Smallest Possible Area

$1-\gamma$	n	α_1	α_2	δ
80.0	10	.0621	.1470	.1433
80.0	25	.0847	.1260	.1133
80.0	100	.0998	.1113	.0828
80.0	1000	.1050	.1061	.0619
90.0	10	.0265	.0755	.0744
90.0	25	.0387	.0638	.0590
90.0	100	.0477	.0549	.0424
90.0	1000	.0510	.0516	.0307
95.0	10	.0117	.0388	.0384
95.0	25	.0180	.0326	.0307
95.0	100	.0231	.0275	.0219
95.0	1000	.0251	.0255	.0154
99.0	10	.0019	.0081	.0080
99.0	25	.0032	.0068	.0066
99.0	100	.0044	.0056	.0047
99.0	1000	.0049	.0051	.0032
99.9	10	.0001	.0009	.0008
99.9	25	.0002	.0008	.0007
99.9	100	.0004	.0006	.0005
99.9	1000	.0005	.0005	.0003

Table 5. Areas of True 90% Confidence Regions

Region type	Sample size (n) 10	25	100
Smallest Mood exact	3.83	1.06	.2258
Large sample (ls)	14.14	1.28	.2245
ls with plug-ins (lspi)	5.69	1.20	.2221
lspi and F-statistic (lspif)	5.69	1.20	.2221
Likelihood ratio test	3.76	1.03	.2161

samples. Second, for samples of size 100 or more, the regions are of relatively equal size. Third, the likelihood ratio test confidence regions are uniformly of smallest expected area, with the smallest Mood exact regions a close second for samples of size 10 and 25. Finally, it is seen that the (basic) large sample region is preferable to the Mood exact region only for sample size 100.

Table 5 also does not discredit Mood's (1950, p. 229) prescient remark that the optimal joint confidence regions would be analytically difficult to obtain but would be "roughly elliptical in shape."

In light of the excellent properties of the likelihood ratio test confidence region, we include a listing of the appropriate critical values to use with it for desired sample sizes and confidence levels. Table 6 shows approximate values obtained via one million simulations of $N(0,1)$ random variables. Values for other sample sizes may be obtained by interpolation.

6. ROBUSTNESS

It is of interest to examine the performance of the aforementioned confidence regions when the underlying random variables are not normally distributed. As an alternative, consider the t distribution with 1 (Cauchy), 5, 10, 50, 100, and 1,000 degrees of freedom (df). To again make a fair comparison among the regions, we have adjusted the critical values for the large sample regions, such that under normality they will each yield 90% confidence (as does the Mood exact region). Also, based on the results of Section 5.3, we have chosen to use the Mood exact region of smallest possible area (with optimal allocation and optimal tail precisions).

Table 6 shows the results of one million simulations. [It is not necessary to show both large sample regions which use plug-ins, as they are identical for adjusted critical values.] A distinct pattern emerges from the maze of numbers: the large sample region with plug-ins for the asymptotic variances is superior for all sample sizes and for all degrees of freedom. And, as expected, the differences between the five regions diminishes with increasing degrees of freedom. For example, when df$=100$ the actual confidences differ by less than 2%.

Recall from Table 5 that, for $n \leq 25$, the large sample regions with plug-ins (lspi) are substantially larger than the corresponding Mood exact or likelihood ratio test regions. However, the superiority of the lspi regions in terms of robustness here shows that the tradeoff may well be worth the additional area. In conclusion, if the normality assumption is to be doubted in favor of a t distribution alternative, the (highly) recommended joint confidence region for all samples of size greater than or equal to 10 is the large sample region (with plug-ins for the asymptotic variances and appropriately adjusted critical values). And, if the normality assumption is not quite as suspect, then the likelihood ratio test regions are to be preferred, as they have the smallest expected areas, and show a moderate degree of robustness (second only to the lspi regions).

Table 6. Critical Values for Various Sample Sizes (n) to be Used for the Likelihood Ratio Test Confidence Region in order to Yield the Desired (true) Confidence Level

n	Confidence level 90%	95%	99%
5	5.68	7.39	11.40
10	5.08	6.62	10.21
15	4.91	6.39	9.82
20	4.83	6.28	9.65
25	4.79	6.23	9.57
30	4.76	6.19	9.50
40	4.72	6.13	9.44
50	4.69	6.10	9.40
100	4.65	6.05	9.30
200	4.63	6.02	9.27
500	4.61	6.00	9.23
1000	4.61	5.99	9.22
∞	4.61	5.99	9.21

Table 7. Fraction of Successes ((0,1) Lies in the joint confidence region) for the Five Regions using Adjusted Critical Values in 1,000,000 Trials of t Distributed Random Variables

region	sample size (n) 10	25	100
df = 1			
Mood exact	.1445	.0068	.0000
large sample (ls)	.0844	.0036	.0000
ls with plug-ins (lspi)	.9930	.0415	.0000
likelihood ratio test	.1075	.0048	.0000
df = 5			
Mood exact	.8075	.6517	.1902
large sample (ls)	.6625	.5073	.1375
ls with plug-ins (lspi)	.9446	.9233	.2816
likelihood ratio test	.7323	.5730	.1620
df = 10			
Mood exact	.8813	.8411	.6432
large sample (ls)	.7973	.7436	.5491
ls with plug-ins (lspi)	.9252	.9420	.7562
likelihood ratio test	.8422	.7922	.5948
df = 50			
Mood exact	.9001	.8993	.8931
large sample (ls)	.8847	.8809	.8716
ls with plug-ins (lspi)	.9039	.9150	.9119
likelihood ratio test	.8939	.8911	.8832
df = 100			
Mood exact	.9001	.9004	.8985
large sample (ls)	.8927	.8914	.8880
ls with plug-ins (lspi)	.9006	.9097	.9080
likelihood ratio test	.8974	.8966	.8940
reference percentage	.9000	.9000	.9000

7. APPLICATIONS

Applications of the foregoing joint confidence regions include uses in testing and estimation. Those interested in the former may examine the two papers by Aitchison (1964, 1965) that propose "confidence-region tests" which would use the regions in the same way that confidence intervals are used in constructing tests of hypotheses involving a single parameter.

Estimation uses include such common practices as constructing confidence intervals for functions of the two parameters μ and σ^2. One novel application involves simultaneous confidence bands for a cdf (Cheng and Illes 1983). More traditionally, separate confidence intervals for μ, σ^2, $\mu + 2\sigma$, and $\frac{\sigma}{\mu}$ (see, for example, Vangel 1996) are constructed. However, having done so, one may not make a joint confidence statement about all four quantities. Whereas by first constructing a joint confidence region for (μ, σ^2), one may construct confidence intervals for μ, σ^2, $\mu + 2\sigma$, and $\frac{\sigma}{\mu}$ that are simultaneously of the desired joint confidence level.

For example, consider the 1907 cricket scores of Tunnicliffe as given by Elderton and Elderton (1920; originally from Wisden's Cricketers' Almanac). The data forms a roughly "normal" looking histogram, and may be summarized by $n = 45$, $\bar{x} = 30.2$, and $s^2 = 575.58$. We construct the 90% joint confidence large sample region with plug-ins for the asymptotic variances (lspi) using a suitably adjusted critical value (use 5.7 instead of $4.6052 = \chi^2_{2,0.90}$, so that under normality the region has exact 90% confidence). By graphing the resulting region, or by using simple grid search techniques, conservative 90% individual confidence intervals for the aforementioned four quantities (i.e., μ, σ^2, $\mu + 2\sigma$, and $\frac{\sigma}{\mu}$) are found to be (21.7,38.7), (286,865), (61.3,91.0), and (.52,1.18), respectively. Furthermore, we may assert with 90% confidence that the true values of the four parametric functions simultaneously lie in each of these four intervals.

[Received January 1996. Revised February 1997.]

REFERENCES

Aitchison, J. (1964), "Confidence-Region Tests," *Journal of the Royal Statistical Society*, Series B, 26, 462–476.

——— (1965), "Likelihood-Ratio and Confidence Region Tests," *Journal of the Royal Statistical Society*, Series B, 27, 245–250.

Bartlett, M.S. (1955), "Approximate Confidence Intervals, III," *Biometrika*, 42, 201–204.

Brownlee, K.A. (1960), *Statistical Theory and Methodology in Science and Engineering*, New York: Wiley.

Cheng, R.C.H., and Iles, T.C. (1983),"Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables," *Technometrics*, 25, 77–86.

Cramer, H. (1951), *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

Douglas, J.B. (1993), "Confidence Regions for Parameter Pairs," *The American Statistician*, 47, 43–45.

Elderton, W. P., and Elderton, E.M. (1920), *Primer Of Statistics* (3rd ed.), London: A. & C. Black, Ltd.

Fieller, E.C. (1954), "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society*, Series B, 16, 175–185.

Hochberg, Y., and Tamhane, A.C. (1987), *Multiple Comparison Procedures*, New York: Wiley.

IMSL, Inc. (1987), *Statistics/Library User's Manual* (ver. 1.0), Houston: Author.

Kalbfleisch, J.G. (1985), *Probability and Statistical Inference: Volume 2, Statistical Inference* (2nd ed.), New York: Springer.

Kendall, M., and Stuart, A. (eds.) (1979), *The Advanced Theory Of Statistics: Volume 2, Inference and Relationship* (4th ed.), London: Charles Griffin & Company Ltd.

Meeker, W.Q., and Escobar, L.A. (1995), "Teaching About Approximate Confidence Regions Based on Maximum Likelihood Estimation," *The American Statistician*, 49, 48–53.

Mood, A.M. (1950), *Introduction to the Theory of Statistics*, New York: McGraw-Hill.

Nickerson, D.M. (1994), "Construction of a Conservative Confidence Region from Projections of an Exact Confidence Region in Multiple Linear Regression," *The American Statistician*, 48, 120–124.

Press, S.J. (1989), *Bayesian Statistics: Principles, Models, and Applications*, New York: Wiley.

Vangel, M.G. (1996), "Confidence Intervals for a Normal Coefficient of Variation," *The American Statistician*, 50, 21–26.

Wilks, S.S. (1962), *Mathematical Statistics*, New York: Wiley.

Wilks, S.S., and Daly, J.F. (1939), "An Optimum Property of Confidence Regions Associated with the Likelihood Function," *Annals of Mathematical Statistics*, 10, 225–235.