The Canadian Journal of Statistics
Vol. 36, No. 4, 2008, Pages 623–637
La revue canadienne de statistique

Shrinkage confidence intervals for the normal mean: using a guess for greater efficiency

Robin WILLINK

Key words and phrases: Confidence coefficient; efficiency; information; interval estimation; point estimation; shrinkage estimator.

MSC 2000: Primary 62F25; secondary 62G05.

Abstract: If the unknown mean of a univariate population is sufficiently close to the value of an initial guess then an appropriate shrinkage estimator has smaller average squared error than the sample mean. This principle has been known for some time, but it does not appear to have found extension to problems of interval estimation. The author presents valid two-sided 95% and 99% "shrinkage" confidence intervals for the mean of a normal distribution. These intervals are narrower than the usual interval based on the Student distribution when the population mean lies in such an "effective interval." A reduction of 20% in the mean width of the interval is possible when the population mean is sufficiently close to the value of the guess. The author also describes a modification to existing shrinkage point estimators of the general univariate mean that enables the effective interval to be enlarged.

Intervalles de confiance à rétrécisseur pour l'espérance d'une loi normale : utilisation d'une valeur initiale au profit d'une meilleure efficacité

Résumé : Si l'espérance inconnue d'une population univariée est assez proche d'une valeur initiale, l'erreur quadratique moyenne d'un bon estimateur à rétrécisseur sera plus faible que celle de la moyenne échantillonnale. Ce principe est connu depuis longtemps, mais il ne semble pas avoir été appliqué à l'estimation par intervalle. L'auteur présente des intervalles de confiance "à rétrécisseur" bilatéraux à 95% et 99% pour l'espérance d'une loi normale. Ces intervalles sont plus courts que l'intervalle habituel fondé sur la loi de Student dès que l'espérance de la population se situe dans un "intervalle effectif." On peut observer une réduction de 20% de la longueur moyenne de l'intervalle quand l'espérance de la population est assez proche de la valeur initiale. L'auteur montre aussi comment certains estimateurs ponctuels à rétrécisseur d'une espérance univariée quelconque peuvent être modifiés de façon à élargir l'intervalle effectif.

1. INTRODUCTION

The problem of interest is the estimation of the mean μ of a univariate distribution from a random sample when a guess μ_0 of μ is available prior to sampling. In this context, the idea of *shrinkage estimation* relates to the technique of moving an unbiased point estimator partway towards μ_0. If the standard error of the unbiased estimator is comparable to or larger than $|\mu - \mu_0|$ then the reduction in variance can be sufficient to outweigh the ensuing squared bias, and the result can be a reduction in mean squared error. This idea that statistical efficiency can be increased when $\mu \approx \mu_0$ has been known for many years, but it does not appear to have been utilized in interval estimation. That is, there seems to be no analogous confidence-interval procedure for μ. Such a procedure would generate a narrower average interval when $|\mu - \mu_0|$ is small at the expense of a wider average interval in other situations.

This paper considers the two problems described: the point estimation of a univariate mean μ using shrinkage, and the interval estimation of μ using shrinkage. Section 2 shows how the quality of existing shrinkage point estimators can be improved by sacrificing accuracy at high values of $|\mu - \mu_0|$, where such an estimator is unlikely to be employed anyway. More notably, Section 3 presents two *shrinkage confidence intervals* for the normal case.

2. A MODIFICATION TO SHRINKAGE POINT ESTIMATORS

Thompson (1968a) developed shrinkage point estimators of the population mean and studied the performance of these estimators under normality. He also studied shrinkage estimators for the binomial proportion and the Poisson parameter (Thompson 1968a; Arnold 1969) and considered the idea of shrinking the unbiased estimator towards a nominal interval rather than a single value (Thompson 1968b). In this section we consider the first of these tasks, which is the point estimation of the mean of a continuous distribution from an independent sample.

Let $X_1, \ldots, X_n$ be a random sample from a distribution with unknown mean μ. The sample mean is $\overline{X} \equiv \sum_{i=1}^{n} X_i/n$ and the usual sample estimate of variance is $S^2 \equiv \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$. Sometimes there is a value μ_0 that is a reasonable guess of μ made before sampling. Thompson (1968a) considered estimating μ by

$$Q_1(h) \equiv \mu_0 + \frac{\overline{X} - \mu_0}{1 + hR}$$

with

$$R \equiv \frac{S^2/n}{(\overline{X} - \mu_0)^2}$$

and with h being a flexible parameter. Mehta & Srinivasan (1971) used a heuristic argument to derive the estimator

$$Q_2(a, b) \equiv \overline{X} - a(\overline{X} - \mu_0)\exp(-b/R),$$

which has greater flexibility through the use of two parameters a and b. Lemmer (1981a, 1981b, 1988) suggested a simpler estimator $Q_3(k) \equiv k\overline{X} + (1-k)\mu_0$, which does not involve the sample variance. The parameter k is such that $1 - k$ is "proportional to the experimenter's confidence in μ_0" (Lemmer 1988).

Figure 1 shows the results of simulations carried out to determine the performances of $Q_1(h)$, $Q_2(a, b)$ and $Q_3(k)$ with samples of size $n = 9$ drawn from a normal distribution. The ratio of the root mean squared error (RMSE) of each shrinkage estimator to the RMSE of $\overline{X}$ is plotted as a function of the absolute value of the normalized difference

$$\delta \equiv \frac{\mu - \mu_0}{\sigma/\sqrt{n}}.$$

The simulations involved setting μ_0 to zero, which we can do without loss of generality, and then generating 10^5 random samples for each value of $|\delta|$ indicated by the tick marks on the horizontal axis. The values of h used were studied by Thompson (1968a, Figure 2, lines (1.5) and (2.1)), the pairs of values of a and b used were presented by Mehta & Srinivasan (1971, Figure 2), and the values of k used were considered by Lemmer (1981b, Figure 1).

In each case there is a positive value of δ, say $\check{\delta}$, below which $y < 1$. That is, there is an effective interval $[-\check{\delta}, \check{\delta}]$ within which the RMSE of the estimator is smaller than that of $\overline{X}$. None of the estimators studied is uniformly better than any other, and none is uniformly better than $\overline{X}$, which is to be expected because $\overline{X}$ is known to be an admissible estimator of μ.

The usefulness of a shrinkage estimator in a practical situation depends on the analyst's confidence that $|\delta| \lessapprox \check{\delta}$. So the value of $\check{\delta}$ seems to be the first measure of quality of such an estimator. The gain in efficiency when $|\delta| \ll \check{\delta}$ is a second measure, while the loss of efficiency when $|\delta| \gg \check{\delta}$ is a third measure. As $|\delta| \to \infty$, the RMSE ratios for $Q_1(h)$ and $Q_2(a, b)$ approach 1 but the ratio for $Q_3(k)$ increases without bound. So, although $Q_3(k)$ exhibits good behaviour at small values of $|\delta|$, we do not consider it further. Of the remaining estimators studied in Figure 1, the one with the largest value of $\check{\delta}$ is $Q_2(0.302, 0.01)$, which has $\check{\delta} \approx 2.3$. When $\delta = 0$ this estimator has RMSE smaller than that of $\overline{X}$ by approximately 30%, i.e., mean squared error smaller by approximately 50%.

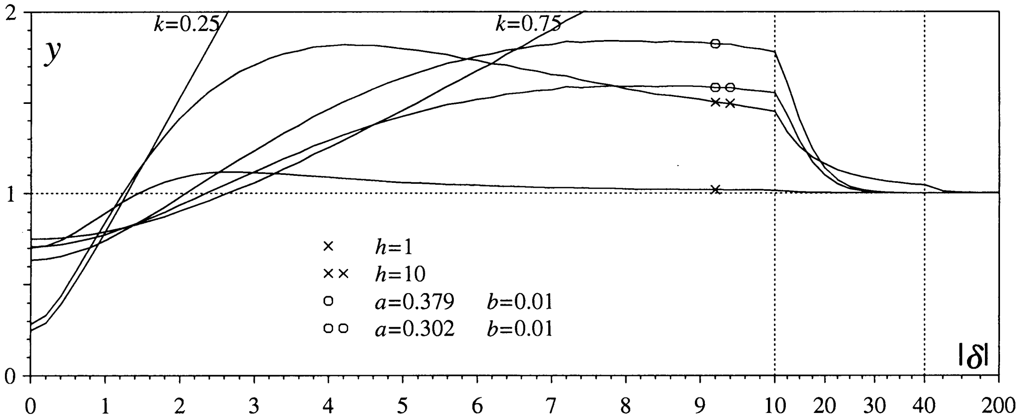

FIGURE 1: Root mean squared error of shrinkage estimators $Q_1(h)$, $Q_2(a, b)$ and $Q_3(k)$ relative to root mean squared error of the sample mean. $y = \mathrm{RMSE}(\text{shrinkage estimator})/\mathrm{RMSE}(\overline{X})$. Each estimator is identified by its parameter label(s) and value(s).

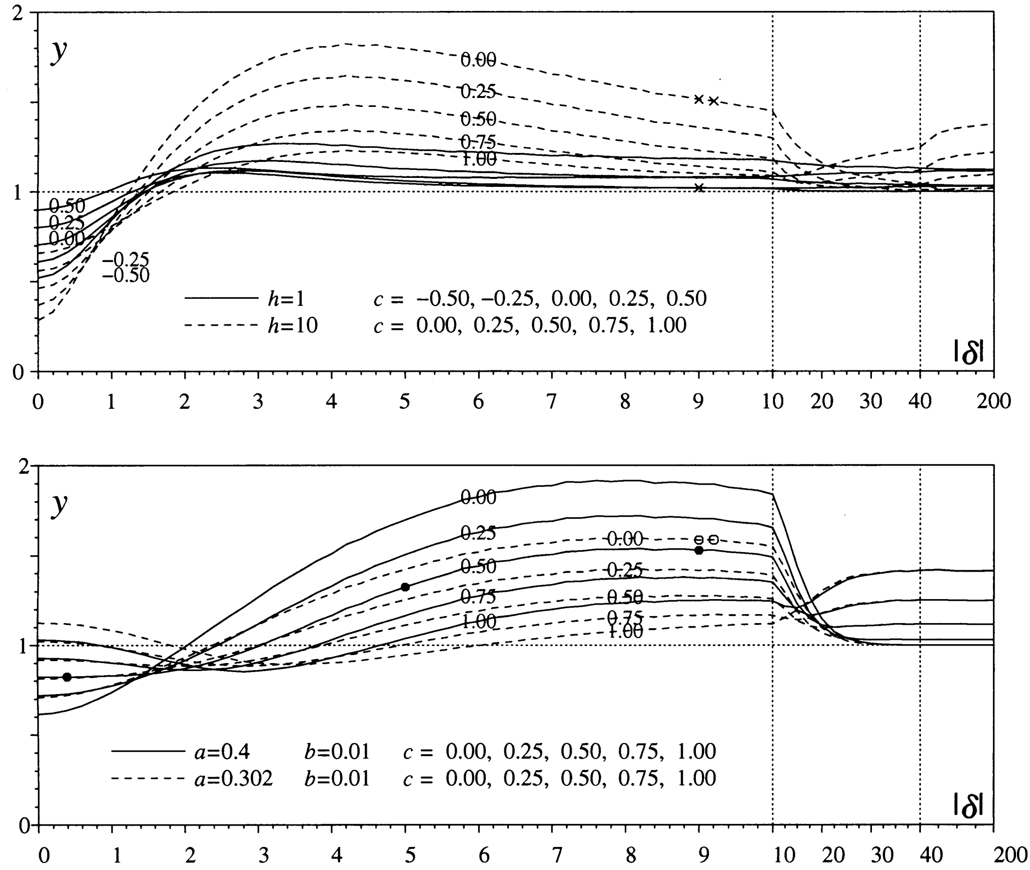

FIGURE 2: Performance of modified estimators $Q_1'(h, c)$ (upper plot) and $Q_2'(a, b, c)$ (lower plot) for the values of c indicated. The results marked • are for the favoured estimator $Q_2'(0.4, 0.01, 0.5)$. The results marked ××, × and ∘∘ correspond to estimators with $c = 0$ studied in Figure 1.

If we could devise an estimator with an RMSE ratio that increased with $|\delta|$ more slowly and without the large peak that is evident in Figure 1, then it seems likely that $\breve{\delta}$ could be made larger. However, the admissibility of $\overline{X}$ implies that if our modified estimator is to have an RMSE ratio that increases monotonically with $|\delta|$, then it must have a ratio greater than unity at $|\delta| \to \infty$. These concepts motivate us to consider the estimators

$$Q_1'(h,c) \equiv Q_1(h) + cV$$

and

$$Q_2'(a,b,c) \equiv Q_2(a,b) + cV$$

with

$$V \equiv \frac{S/\sqrt{n}}{1+R} \times \operatorname{sgn}(\overline{X} - \mu_0).$$

The variables S^2 and $\overline{X}$ are independent under normality, and $\Pr(|R| > \varepsilon) \to 0$ for all $\varepsilon > 0$ as $|\delta| \to \infty$. Therefore V becomes independent of both $Q_1(h)$ and $Q_2(a,b)$ as $|\delta| \to \infty$. It follows that, under normality, the RMSE ratios for both $Q_1'(h,c)$ and $Q_2'(a,b,c)$ approach $\surd(1+c^2)$.

Figure 2 demonstrates the performances of $Q_1'(h,c)$ and $Q_2'(a,b,c)$ for different values of c. The modification enables the value of $\breve{\delta}$ to be increased. The estimator that we choose to highlight is $Q_2'(0.4, 0.01, 0.5)$, which has $\breve{\delta} \approx 3.5$ and has an RMSE lower than that of $\overline{X}$ by 10% or more for all $|\delta| \lessapprox 2.5$. If 3.5 is regarded as excessive as an upper bound for $|\delta|$, then more efficient estimators such as $Q_2'(0.4, 0.01, 0.25)$ and $Q_2'(0.4, 0.01, 0)$ are available. (Again, the simulations involved 10^5 trials at each value of $|\delta|$ given by the tick marks. This is also the case with Figures 3 and 4 below.)

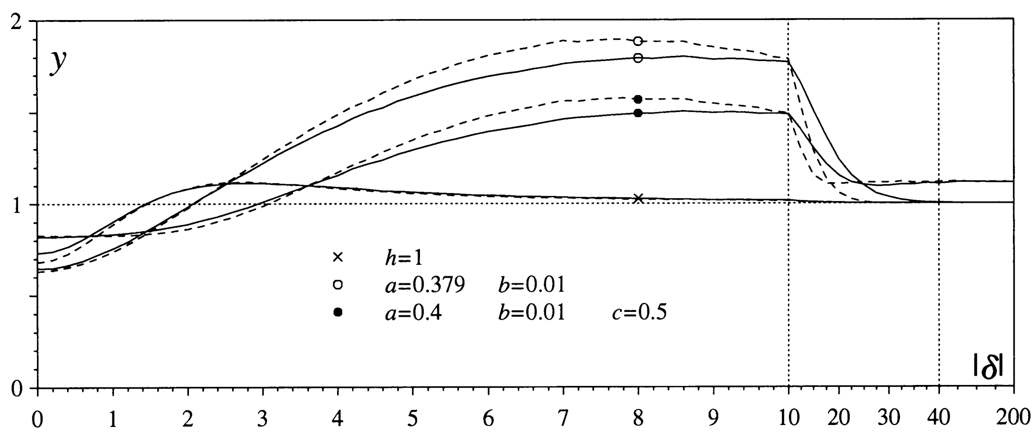

FIGURE 3: Performance of various shrinkage estimators with different sample sizes. $y = \text{RMSE(shrinkage estimator)}/\text{RMSE}(\overline{X})$. Solid lines: $n = 4$. Dashed lines: $n = \infty$.

2.1. Insensitivity to sample size and distribution.

Thompson (1968a) showed that the performance of his estimator under normality with $n = 9$ was improved only slightly when σ^2 was known. This suggests that the value of n will have little effect. In agreement with this, Figure 3 shows results for some of the estimators studied in Figures 1 and 2 with $n = 4$ and $n = \infty$ under normality. (The simulations for $n = \infty$ involved drawing values of $\overline{X}$ from the normal distribution with mean $|\delta|$ and variance 1, and setting $S^2/n = 1$ in the equations for R and V.) The results imply that we may consider the value of $\breve{\delta}$ and the pattern of performance to be broadly independent of n.

It is also appropriate to examine the robustness of the estimators to departures from normality. Figure 4 shows results for samples of size $n = 9$ drawn from (i) the normal distribution, (ii) Student's t-distribution with 4 degrees of freedom, (iii) the distribution of chi-squared with 4 degrees of freedom, which is visually quite asymmetric, and (iv) the reflection of that chi-squared distribution. The abscissa is now δ to accommodate sampling from an asymmetric distribution. The results indicate that the value of $\breve{\delta}$ is acceptably insensitive to departures from normality.

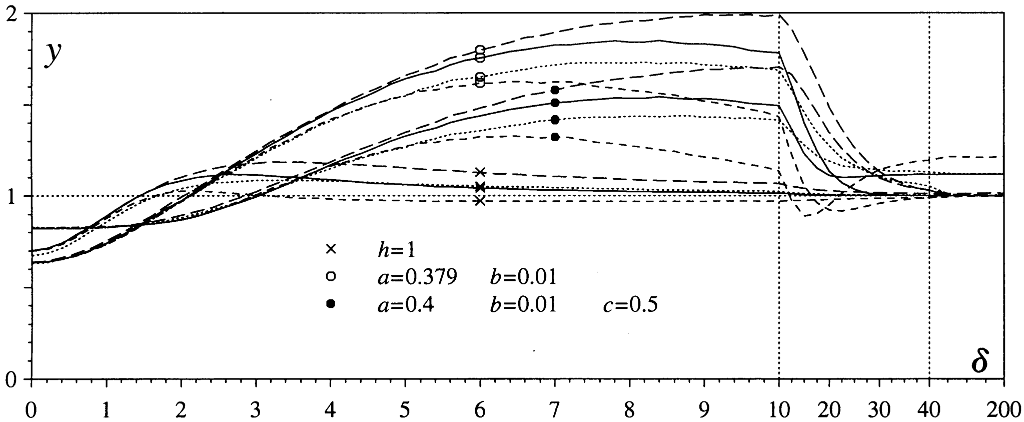

FIGURE 4: Performance of various shrinkage estimators at $n = 9$ with different distributions. $y = \mathrm{RMSE}(\text{shrinkage estimator})/\mathrm{RMSE}(\overline{X})$. Solid lines: normal distribution. Dotted lines: t distribution with $\nu = 4$. Short-dashed lines: chi-square distribution with $\nu = 4$. Long-dashed lines: reflection of chi-square distribution with $\nu = 4$.

Let us summarize our analysis for this problem of point estimation of the mean μ from a random sample. The results shown in Figures 2–4 suggest that if the underlying distribution is unimodal, then the shrinkage estimator $Q_2'(0.4, 0.01, 0.5)$ is suitable for estimating μ when it is thought that $|\delta| \lessapprox 3.5$.

3. SHRINKAGE CONFIDENCE INTERVALS

We now describe the construction of shrinkage confidence intervals for μ under the premise of normality. The intervals have confidence coefficients equalling or exceeding the nominal coefficient for all values of μ and σ, and the intervals are typically narrower than the traditional interval based on the t-distribution when $\mu \approx \mu_0$.

The majority of results relating to confidence intervals in shrinkage estimation are for the contrasting problem of estimating the mean vector of a p-dimensional multivariate normal distribution (George & Casella 1994; Tseng & Brown 1997; Takada 1998; Kazimi & Brownstone 1999; Tseng 2004). There are at least two important differences between shrinkage estimation in that context and shrinkage estimation in our univariate problem. First, in the multivariate context the original estimate is shrunk towards the zero vector even though that vector does not have the status of a meaningful guess. Second, the resulting estimate can dominate the maximum likelihood estimate in terms of mean squared error, which is not the case in the univariate problem (Muirhead 1982, p. 123). The multivariate normal problem therefore differs from the problem addressed in this paper both conceptually and practically, and most of the results presented in that context do not seem relevant here.

We regard the most relevant question raised in the multivariate normal context to be that of finding confidence sets with the stated coverage probability but with reduced volume (Tseng 2004, § 1, Question (2)(b)). Analogously, the task here is to obtain a procedure with at least the stated probability of generating an interval enclosing the univariate mean μ and with reduced typical width when $\mu \approx \mu_0$.

Let the generation of an interval that encloses μ be called *success*. A procedure with probability of success greater than or equal to p whatever the values of μ and σ is a valid $100p\%$ confidence-interval procedure for μ, and the interval itself can be called *valid*. Set

$$\dot{p} \equiv 0.5 + p/2.$$

The usual valid two-sided confidence interval is the exact interval

$$[X_{\rm L},\ X_{\rm H}] \equiv [\overline{X} - t_{\dot{p}} S/\sqrt{n}, \overline{X} + t_{\dot{p}} S/\sqrt{n}], \tag{1}$$

with $t_{\dot{p}}$ being the $\dot{p}$ quantile of Student's t-distribution with $n-1$ degrees of freedom. We now describe two different valid procedures that generate intervals narrower than $[X_{\rm L}, X_{\rm H}]$ when μ is near μ_0 at the expense of wider intervals in other situations.

3.1. First interval.

Define $U_\varepsilon^+ \equiv t_{\dot{p}+\varepsilon} S/\sqrt{n}$ and $U_\varepsilon^- \equiv t_{\dot{p}-\varepsilon} S/\sqrt{n}$ with ε being a predetermined positive value less than $1-\dot{p}$. The first interval proposed is

$$[X'_{\rm L}, X'_{\rm H}] \equiv \begin{cases} \left[X_{\rm L}, \min\{\overline{X} + U_\varepsilon^+, \mu_0\}\right], & X_{\rm H} < \mu_0 \quad \text{(case 1)}, \\ \left[\min\{\overline{X} - U_\varepsilon^-, \mu_0\}, \max\{\overline{X} + U_\varepsilon^-, \mu_0\}\right], & X_{\rm L} \le \mu_0 \le X_{\rm H} \quad \text{(case 2)}, \\ \left[\max\{\overline{X} - U_\varepsilon^+, \mu_0\}, X_{\rm H}\right], & X_{\rm L} > \mu_0 \quad \text{(case 3)}. \end{cases} \tag{2}$$

This interval is wider than $[X_{\rm L}, X_{\rm H}]$ in cases 1 and 3 but is narrower in case 2. Figure 5 shows the three cases schematically and indicates the limits that would change if the shrinkage procedure were used.

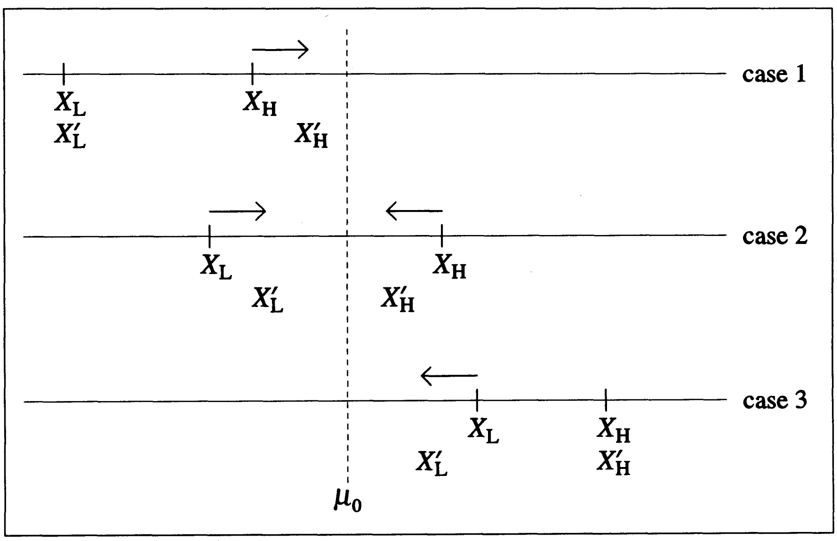

FIGURE 5: Schematic diagram of shrinkage of interval limits for the three different cases in (2).

The procedure treats the random limits of the interval symmetrically, so the interval realized is the negative of the interval that would result from negating all the data and then applying the method. However, the interval realized is not symmetrical about the observed sample mean.

When $\varepsilon = 0$ the interval (2) reduces to the usual interval $[X_{\mathrm{L}}, X_{\mathrm{H}}]$, which has mean width $2t_{\dot{p}}E(S)/\sqrt{n}$. When $|\delta| \to \infty$ the interval is either that of case 1 or case 3, and the mean width of the interval is $(t_{\dot{p}} + t_{\dot{p}+\varepsilon})E(S)/\sqrt{n}$.

The usual interval $[X_{\mathrm{L}}, X_{\mathrm{H}}]$ is exact, so to show that $[X'_{\mathrm{L}}, X'_{\mathrm{H}}]$ has probability of success greater than or equal to p, it is sufficient to show that the probability of success gained when adopting (2) equals or exceeds the probability of success lost, i.e., that

$$\Pr\big([X'_{\mathrm{L}}, X'_{\mathrm{H}}] \ni \mu \cap [X_{\mathrm{L}}, X_{\mathrm{H}}] \not\ni \mu\big) \geq \Pr\big([X'_{\mathrm{L}}, X'_{\mathrm{H}}] \not\ni \mu \cap [X_{\mathrm{L}}, X_{\mathrm{H}}] \ni \mu\big). \tag{3}$$

We now show that this inequality holds.

A gain of success occurs if the upper limit is increased from below μ to above μ (which can only occur in case 1) or if the lower limit is reduced from above μ to below μ (which can only occur in case 3). The probability of the first of these events is

$$\Pr\big(X_{\mathrm{H}} < \mu_0 \cap X_{\mathrm{H}} < \mu \cap \min\{\overline{X} + U_\varepsilon^+, \mu_0\} > \mu\big) \tag{4}$$

and the probability of the second of these events is

$$\Pr\big(X_{\mathrm{L}} > \mu_0 \cap X_{\mathrm{L}} > \mu \cap \max\{\overline{X} - U_\varepsilon^+, \mu_0\} < \mu\big). \tag{5}$$

These events are the only means by which a gain of success occurs. They are disjoint, so the probability on the left-hand side of (3) is the sum of these two probabilities.

Similarly, a loss of success occurs if the upper limit is reduced beyond μ (which can only occur in case 2) or if the lower limit is increased over μ (which also can only occur in case 2). The first event has probability

$$\Pr\big(X_{\mathrm{L}} < \mu_0 < X_{\mathrm{H}} \cap X_{\mathrm{H}} > \mu \cap \max\{\overline{X} + U_\varepsilon^-, \mu_0\} < \mu\big) \tag{6}$$

and the second event has probability

$$\Pr\big(X_{\mathrm{L}} < \mu_0 < X_{\mathrm{H}} \cap X_{\mathrm{L}} < \mu \cap \min\{\overline{X} - U_\varepsilon^-, \mu_0\} > \mu\big). \tag{7}$$

Likewise, these events are the only means by which a loss occurs. They too are disjoint, so the probability on the right-hand side of (3) is the sum of these two probabilities.

Suppose $\mu > \mu_0$. Then the probabilities (4) and (7) are zero. Also, in this case $X_{\mathrm{L}} > \mu$ implies that $X_{\mathrm{L}} > \mu_0$, so the probability (5) is

$$\Pr\big(X_{\mathrm{L}} > \mu \cap \overline{X} - U_\varepsilon^+ < \mu\big) = \Pr\big(T > t_{\dot{p}} \cap T < t_{\dot{p}+\varepsilon}\big),$$

where

$$T \equiv \frac{\overline{X} - \mu}{S/\sqrt{n}}. \tag{8}$$

Under normality, T has Student's t-distribution with $n-1$ degrees of freedom, in which case this probability is equal to ε. Also when $\mu > \mu_0$, the probability (6) is

$$\Pr\big(X_{\mathrm{L}} < \mu_0 < X_{\mathrm{H}} \cap X_{\mathrm{H}} > \mu \cap \overline{X} + U_\varepsilon^- < \mu\big),$$

which cannot exceed

$$\begin{aligned}\Pr\big(X_{\mathrm{H}} > \mu \cap \overline{X} + U_\varepsilon^- < \mu\big) &= \Pr\big(T > -t_{\dot{p}} \cap T < -t_{\dot{p}-\varepsilon}\big) \\ &= \varepsilon.\end{aligned}$$

It follows that the required inequality (3) holds when $\mu > \mu_0$. By the argument of symmetry, we can also conclude that the inequality holds when $\mu < \mu_0$. In addition, there can be neither gain nor loss when $\mu = \mu_0$ because the use of the maximum and minimum functions prevents any limit from moving through μ_0. Thus the confidence-interval procedure is valid under the premise of normality.

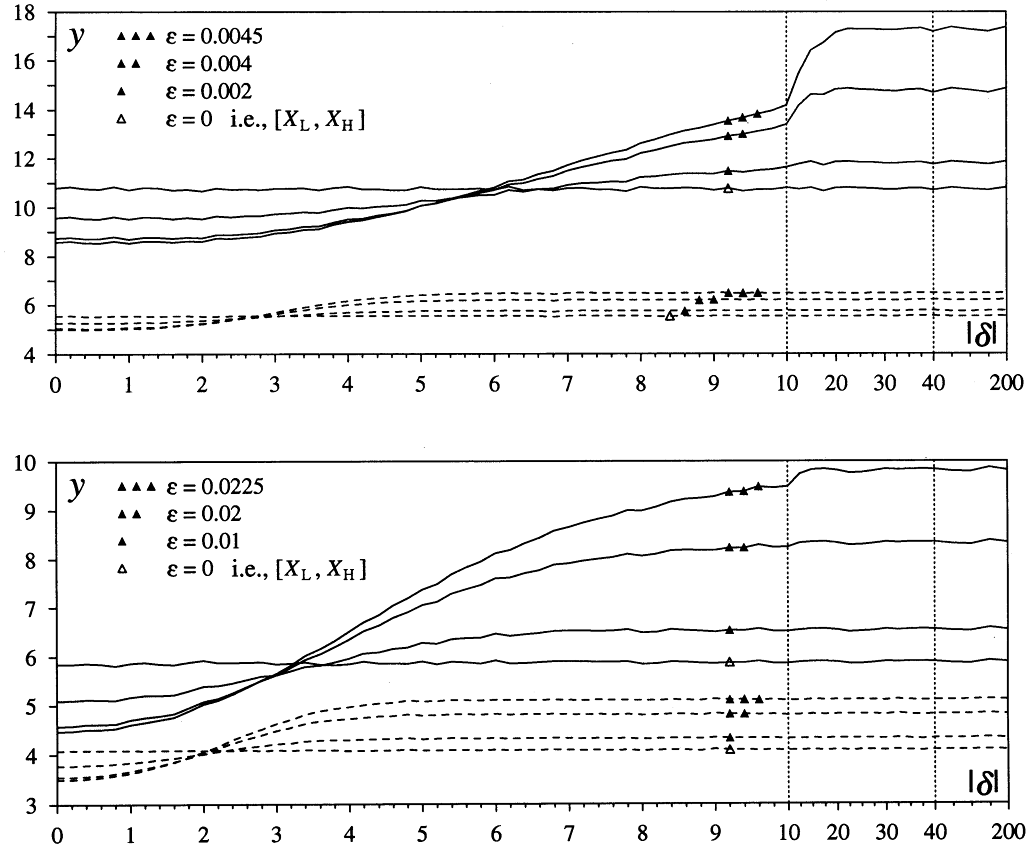

FIGURE 6: Mean width of $100p\%$ shrinkage confidence interval (2). $y = E(X'_H - X'_L)/(\sigma/\sqrt{n})$. Upper plot: $p = 0.99$. Lower plot: $p = 0.95$. Solid lines: $n = 4$. Dashed lines: $n = 25$.

Figure 6 shows the normalized mean width of (2) obtained from 10^4 trials carried out at each of the values of $|\delta|$ indicated by the tick marks. The confidence levels used were $p = 0.99$ and $p = 0.95$, and the sample sizes studied were $n = 4$ and $n = 25$. When $p = 0.99$, $n = 4$ and $\varepsilon = 0.0045$, interval (2) has $\breve{\delta} \approx 5.8$ and has mean width approximately 20% less than $[X_{\mathrm{L}}, X_{\mathrm{H}}]$ when $\delta = 0$. Likewise, when $p = 0.95$, $n = 4$ and $\varepsilon = 0.0225$, interval (2) has $\breve{\delta} \approx 3.2$ and has mean width more than 20% less than $[X_{\mathrm{L}}, X_{\mathrm{H}}]$ when $\delta = 0$. So the potential gain in efficiency when using (2) is considerable.

The value of $\varepsilon = 0.0045$ for $p = 0.99$ and the value of $\varepsilon = 0.0225$ for $p = 0.95$ are both given by $\varepsilon = 0.9(1 - \dot{p})$, in which case $\dot{p} - \varepsilon = 0.95p + 0.05$ and $\dot{p} + \varepsilon = 0.05p + 0.95$. This choice could also be made with other values of p. Accordingly, the $100p\%$ two-sided confidence interval proposed for μ is

$$[X'_{\mathrm{L}}, X'_{\mathrm{H}}] \equiv \begin{cases} \left[X_{\mathrm{L}}, \min\{\overline{X} + t_{p_2} S/\sqrt{n}, \mu_0\}\right], & X_{\mathrm{H}} < \mu_0, \\ \left[\min\{\overline{X} - t_{p_1} S/\sqrt{n}, \mu_0\}, \max\{\overline{X} + t_{p_1} S/\sqrt{n}, \mu_0\}\right], & X_{\mathrm{L}} \le \mu_0 \le X_{\mathrm{H}}, \\ \left[\max\{\overline{X} - t_{p_2} S/\sqrt{n}, \mu_0\}, X_{\mathrm{H}}\right], & X_{\mathrm{L}} > \mu_0 \end{cases} \tag{9}$$

with $p_1 = 0.95p + 0.05$ and $p_2 = 0.05p + 0.95$.

Consider the idea of a one-sided shrinkage interval. A left-infinite $100\dot{p}\%$ shrinkage interval for μ, say $(-\infty, X]$, would tend to have X less than X_{H} when $|\delta|$ is small, but greater than X_{H}

when $|\delta|$ is large. Similarly, a right-infinite shrinkage interval would involve a limit greater than X_{L} when $|\delta|$ is small, but less than X_{L} when $|\delta|$ is large.

The one-sided intervals formed in the natural way from the two-sided intervals just described would be $(-\infty, X'_{\mathrm{H}}]$ and $[X'_{\mathrm{L}}, \infty)$. However, these are not valid $100\dot{p}\%$ confidence intervals for μ. To show this, we examine the change in the probability of success when preferring $(-\infty, X'_{\mathrm{H}}]$ to the exact left-infinite $100\dot{p}\%$ confidence interval $(-\infty, X_{\mathrm{H}}]$. The probability of a gain is given by the single probability (4), which is zero when $\mu > \mu_0$. Yet the probability of a loss is positive for all $\mu \neq \mu_0$. So the interval $(-\infty, X'_{\mathrm{H}}]$ does not have success probability reaching $\dot{p}$ when $\mu > \mu_0$. By symmetry, the interval $[X'_{\mathrm{L}}, \infty)$ does not have success probability reaching $\dot{p}$ when $\mu < \mu_0$.

3.2. Second interval.

The second $100p\%$ shrinkage confidence interval we present is of the form

$$[\overline{X} - Z, \overline{X} + Z] \qquad Z = \begin{cases} t_1 S/\sqrt{n} & \left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| \leq t_2, \\ t_3 S/\sqrt{n} & \left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| > t_2, \end{cases} \tag{10}$$

where $t_1 < t_{\dot{p}} < t_3$. (Note that the meaning of the integer subscript in t_1, for example, is different from the meaning of the fractional subscript in $t_{\dot{p}}$, for example.) For best performance, this interval would involve a small value of t_1 and a large value of t_2, because there would then be a high probability of generating an interval considerably narrower than $[X_{\mathrm{L}}, X_{\mathrm{H}}]$ when $\mu \approx \mu_0$. Unlike the interval realized in (2), the interval realized in (10) is symmetrical about the observed sample mean.

It remains to identify values of t_1, t_2 and t_3 that make (10) a valid $100p\%$ confidence interval for μ. The probability of success is seen to be

$$\begin{aligned} p^* &\equiv p^*(\delta; n, t_1, t_2, t_3) \\ &= \Pr\big(|T| \leq t_1 \cap |T'| \leq t_2\big) + \Pr\big(|T| \leq t_3 \cap |T'| > t_2\big), \end{aligned} \tag{11}$$

where T is defined in (8) and

$$T' \equiv \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$

We require $p^* \geq p$ for all δ. Jointly, T and T' have Owen's doubly non-central bivariate t-distribution with $n - 1$ degrees of freedom and with non-centrality parameters 0 and δ (Owen 1965). So p^* can be calculated for any value of the unknown δ using equations given by Owen (1965) and Chou (1992). Our method of calculation is described in the Appendix. Note that the probability p^* is maximized with respect to δ when $|\delta| \rightarrow \infty$, in which case $p^* = \Pr(|T| < t_3)$. When $\delta = 0$ success occurs if $|T| < t_1$ or if $t_2 < |T| < t_3$, so we must have $t_2 \leq t_3$.

Let us make the reparametrization $t_q = t_2$, $at_q = t_1$ and $bt_q = t_3$, where q is a probability. It was found by experimentation that the intervals

$$[\overline{X} - Z, \overline{X} + Z] \qquad Z = \begin{cases} at_q S/\sqrt{n}, & \left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| \leq t_q, \\ bt_q S/\sqrt{n}, & \left|\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}\right| > t_q, \end{cases} \qquad n = 2, 3, \ldots, 50 \tag{12}$$

with $(a, b, q) = (0.91, 1.29, 0.995)$ and $(a, b, q) = (0.95, 1.08, 0.995)$ are valid 99% confidence intervals for μ and that the intervals $(a, b, q) = (0.85, 1.46, 0.975)$ and $(a, b, q) =$

$(0.9, 1.17, 0.975)$ are valid 95% confidence intervals. Note that $q = \dot{p}$ for each of these intervals. (These intervals have parameters satisfying $p^*(0; \infty, t_1, t_2, t_3) \overset{\rightharpoonup}{=} p$, where $\overset{\rightharpoonup}{=}$ means "is equal to or just greater than." That is, $2\{\Phi(a\Phi^{-1}(q)) - 0.5\} + 2\{\Phi(b\Phi^{-1}(q)) - q\} \overset{\rightharpoonup}{=} p$, with $\Phi(\,\cdot\,)$ being the standard normal distribution function.)

The minimum probability of success with respect to the unknown parameter δ is an indication of the potential efficiency of the procedure. For the interval with $(a, b, q) = (0.91, 1.29, 0.995)$ this minimum probability is maximized with respect to n at approximately 99.11%, which is observed with $n = 4$. For $(a, b, q) = (0.95, 1.08, 0.995)$ the corresponding figures are 99.02% and $n = 6$. For $(a, b, q) = (0.85, 1.46, 0.975)$ the figures are 95.47% and $n = 3$, and for $(a, b, q) = (0.9, 1.17, 0.975)$ the figures are 95.11% and $n = 3$.

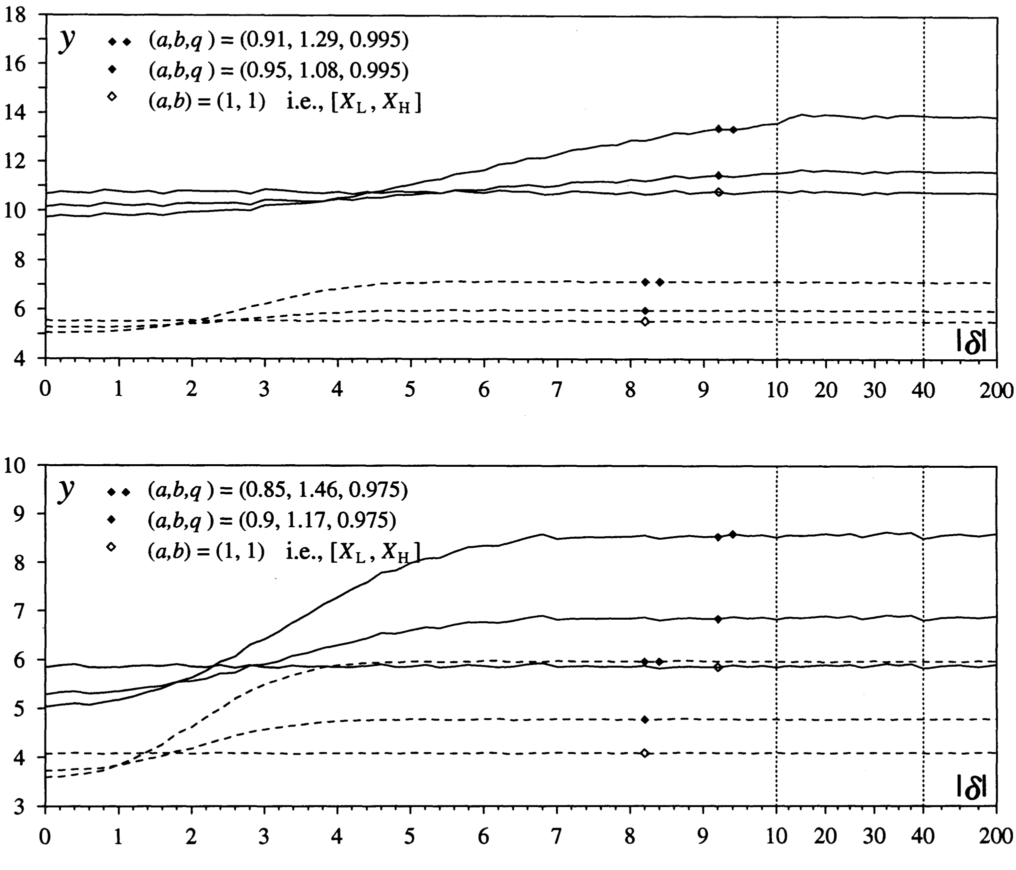

FIGURE 7: Mean width of $100p\%$ shrinkage confidence interval (12). $y = 2E(Z)/(\sigma/\sqrt{n})$.
Upper plot: $p = 0.99$. Lower plot: $p = 0.95$. Solid lines: $n = 4$. Dashed lines: $n = 25$.

Figure 7 shows the normalized mean width of (12) for the same values of p and n studied in Figure 6. Again we used 10^4 trials. The setting of $(a, b) = (1, 1)$ corresponds to the usual interval $[X_L, X_H]$. The levels of performance observed with (12) are not as high as those observed with (9).

Like the one-sided intervals examined in Section 3.1, the intervals $(-\infty, \overline{X} + Z]$ and $[\overline{X} - Z, \infty)$ are not valid 99.5% and 97.5% confidence intervals for μ. This can be shown by evaluating the probability of success of the interval $(-\infty, \overline{X} + Z]$, which is

$$\Pr(\overline{X} + Z \geq \mu) = \Pr\big(T \geq -t_1 \cap |T'| \leq t_2\big) + \Pr\big(T \geq -t_3 \cap |T'| > t_2\big).$$

For example, when $n = 20$ and $\delta = 1$ this probability is 0.991, 0.993, 0.954 and 0.962 for the four intervals given. These figures are below the nominal confidence coefficients of 0.995, 0.995, 0.975 and 0.975, so the intervals are not valid. The analysis with $[\overline{X} - Z, \infty)$ will obey symmetry, and will have success probability falling below the nominal value for sufficiently negative δ.

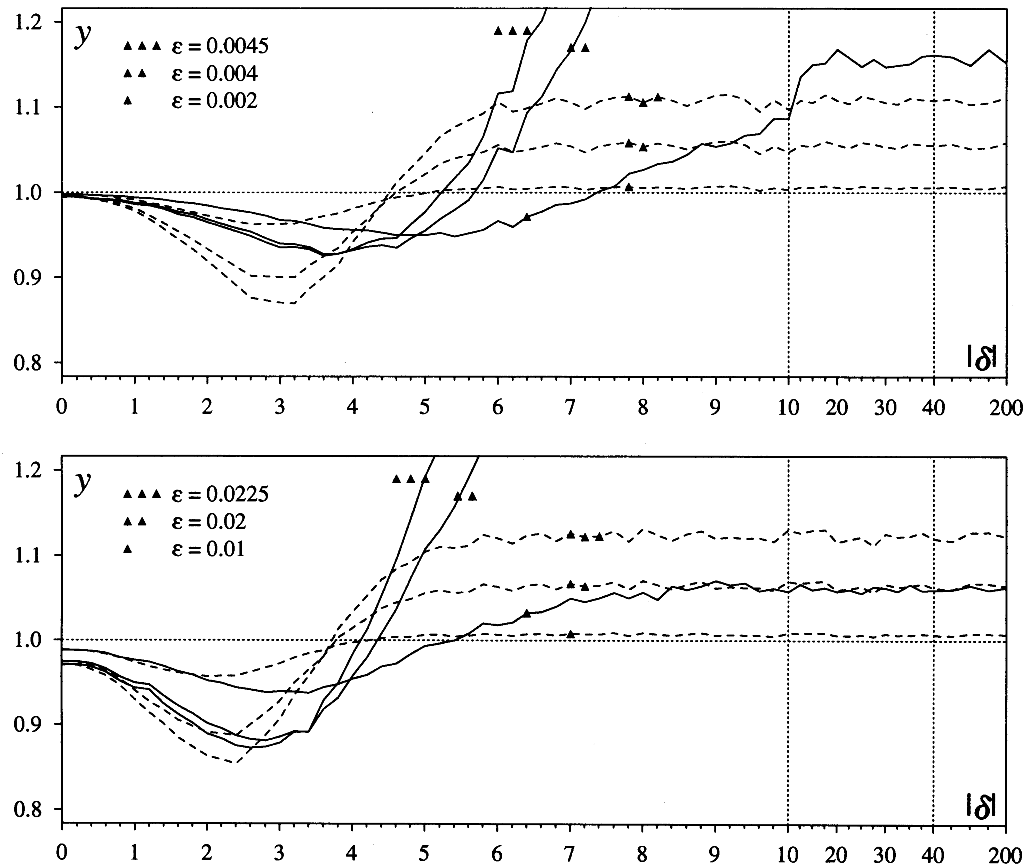

FIGURE 8: Performance of X'_{mid} as a point estimator. $y = \text{RMSE}(X'_{\text{mid}})/\text{RMSE}(\overline{X})$.
Upper plot: $p = 0.99$. Lower plot: $p = 0.95$. Solid lines: $n = 4$. Dashed lines: $n = 25$.
The asymptotic values of y for $\varepsilon = 0.0045$ and $\varepsilon = 0.004$ in the upper plot are approximately 3.7 and 2.4.
The asymptotic values of y for $\varepsilon = 0.0225$ and $\varepsilon = 0.02$ in the lower plot are approximately 2.4 and 1.7.

3.3. Point estimator.

The purpose of an interval estimation procedure is to obtain intervals that contain the unknown value of the parameter with at least the frequency implied by the quoted confidence level. There is no need to associate the interval with any point estimator to achieve this, so there is no scientific reason why the interval obtained should be symmetric about any particular point estimator. Therefore, in view of its better performance, we suggest that (9) is more useful than (12).

However, perhaps a point estimator is to be quoted either with (9) or with its more general form (2). A natural choice would be the mid-point $X'_{\text{mid}} \equiv (X'_{\text{L}} + X'_{\text{H}})/2$. Figure 8 shows the root mean squared error of X'_{mid} relative to that of $\overline{X}$ during the simulations used in obtaining Figure 6. (The curves are rougher than those in Figures 1–4 because the number of trials, 10^4, is smaller.) The value of $\breve{\delta}$ for X'_{mid} depends on p, n and ε but, for the sets of parameters values

examined, the value of $\breve{\delta}$ exceeds the value of 3.5 found for the point estimator $Q_2'(0.4, 0.01, 0.5)$ advocated in Section 2. The reduction in RMSE with X'_{mid} can reach 10%, which is the summary value obtained with $Q_2'(0.4, 0.01, 0.5)$. However, when $|\delta|$ is large, X'_{mid} with the higher values of ε is considerably less efficient than $Q_2'(0.4, 0.01, 0.5)$.

So, despite its inefficiency at high values of $|\delta|$, X'_{mid} is suggested as a suitable point estimator for accompanying (9). Conversely, the performance of X'_{mid} might be thought sufficiently good for it to be employed when the primary requirement is a point estimate and when an accompanying confidence interval is expected.

4. EXAMPLE

We now present a fictitious example involving (9). By choosing such an example, we are able to describe a proper usage of the method and to contrast its applicability with that of a Bayesian method.

The managers of a forestry company are to decide whether the company will plant trees of species A or trees of species B. All the managers agree that the measure of quality (of wood) in individual trees of the same species closely follows a normal distribution, and all agree on the mean quality of species A. However, their opinions are sharply divided over the mean quality of species B. A company scientist is therefore asked to obtain a 95% interval estimate of the mean quality μ of trees in species B, and to do so in a way that all the managers can accept.

From experience, the scientist is confident that the value of μ lies in the interval $[70, 90]$, and so considers the use of the shrinkage interval (9) with $\mu_0 = 80$. The variability observed in natural processes is high, and the destruction and testing of trees is expensive and time consuming. This situation of large σ and small n encourages the scientist to think that $|\delta|$ may well be small enough for the shrinkage procedure that generates (9) to be suitable.

Only $n = 8$ trees of species B are available for examination, and these can be regarded as a random sample from the corresponding population. The scientist proceeds to have the quality of wood assessed in each of these trees. The sample variance S^2 is independent of the sample mean $\overline{X}$, so no information about μ is obtained by observing the sample variance S^2. Therefore, provided that $\overline{X}$ remains unobserved, the decision to apply (9) instead of (1) can legitimately be made after observing S^2. The data are handled and processed accordingly (perhaps by subtracting a single randomly chosen large constant from every measurement) and the sample variance is calculated to be $s^2 = 48.97$. The scientist is confident that $70 \leq \mu \leq 90$ and so, after calculating $s/\sqrt{n} \approx 2.47$, is reasonably confident that $|\delta| < 3.5$. This leads to the decision to apply (9) with $p = 0.95$.

The (ordered) data set is then observed to be $\{71.9, 72.3, 75.5, 80.1, 84.8, 85.6, 88.1, 89.2\}$ and the sample mean is calculated to be $\bar{x} = 80.94$. The random interval (9) with $p = 0.95$ is of case 2 in Figure 5 and takes the value $[76.2, 85.7]$. This interval is correctly quoted to the managers as an interval in which they can have 95% confidence, and the midpoint of this interval, 80.9, is quoted as a point estimate. (Note that (1) gives the wider interval $[74.0, 87.9]$.)

The approach taken by the scientist can be contrasted with a Bayesian approach, which is unsuitable in this situation because the prior beliefs of the managers about the value of μ show a wide spread. Each manager has their own opinion and, in accordance with Bayesian reasoning, would not find meaning in any posterior distribution other than that derived from their own prior distribution. So, whether or not the practice of encoding belief as a probability distribution is trusted, a Bayesian analysis is inappropriate because there is no agreed prior information about μ.

The validity of the shrinkage procedure, however, does not depend on whether the managers agree with, or even know about, the scientist's choice of μ_0. If they trust the scientist's integrity and competence as a statistician, then they will agree that the procedure had probability at least 95% of generating an interval containing μ and will accept the result as a valid 95% confidence interval for μ.

This example shows that the use of a shrinkage procedure can involve a different kind of prior

information than is found in a Bayesian analysis. For the shrinkage procedure to achieve the 95% probability of success claimed, the prior information only need be available to the analyst; no agreement about μ_0 is needed. In contrast, the prior information in a Bayesian analysis must be agreed upon by all if a 95% credibility interval derived from the posterior distribution is to be accepted by all.

5. ON COMPARISON WITH BAYESIAN METHODS

Some readers will look for a comparison of the performances of the shrinkage intervals and Bayesian intervals. As stressed in this paper, the random intervals (9) and (12) are constructed to have at least the nominal probability of success whatever the values of μ and σ. This frequentist requirement of validity is our principal performance criterion. If we are to properly compare the widths of Bayesian intervals with those of (9) and (12), then we can only do so using intervals that also satisfy this criterion.

It is well known that the highest posterior density $100p\%$ credibility interval obtained using the Bayesian procedure with the prior density function $f(\mu, \sigma) \propto 1/\sigma$ is identical to the interval realized by (1). Therefore, that procedure, which uses a uniform marginal density function for μ, has the required property of validity. However, the intervals that it generates are wider than the shrinkage intervals when $|\delta| < \breve{\delta}$, i.e., when $\mu \approx \mu_0$. It seems intuitive that to obtain narrower intervals when $\mu \approx \mu_0$ the marginal prior density for μ must favour values near μ_0, i.e., must have a mode near μ_0. It is also intuitive that using such a prior distribution will result in a procedure with less than 95% probability of success when $\mu \not\approx \mu_0$ unless the statistician deliberately calculates posterior intervals with probability exceeding p to compensate for this effect. This in turn would widen the intervals. So, unless this curious step is taken, a Bayesian procedure cannot have the required property of validity while competing favourably with the shrinkage procedures in terms of width.

Presumably, many Bayesian statisticians, especially those who have subjectivist views, will not find value or meaning in the property of validity that we have taken as our starting point. They might consider instead that the frequency of success implied by routine use of a standard figure of credibility, e.g. 95%, should relate instead to a hypothetical large set of problems with differing values of the parameters μ and σ. Evidently, the Bayesian approach with specific prior distributions for μ and σ will only result in a success frequency of 95% in such a set of problems if these prior distributions match the frequency distributions of the parameters in this set. Therefore, such a comparison makes scientific sense only when these frequency distributions are known and when the Bayesian statistician has no reason to vary the prior distributions used throughout this set of problems. These conditions appear unrealistic and seem to go against the Bayesian ethos. Consequently, a comparison of success rate or interval width is left to those for whom it would have greater meaning.

Some might regard the shrinkage methods and Bayesian methods as automatically comparable because they both involve prior "information." This view fails to acknowledge that there are different types of information. Recall the example of Section 4, where the only prior information required for the application of the shrinkage interval corresponded to confidence that μ lay in the interval $[70, 90]$, and where it was sufficient that this information be held by the statistician alone. In contrast, the prior information required in a Bayesian analysis is information fully quantifed in the form of a complete probability density function, and this has to be agreed upon by all the interested parties. If it is acknowledged that no such quantification is perfect or if there is no strict agreement over prior belief, then any trustworthy Bayesian analysis also requires a sensitivity analysis (O'Hagan 1994, pp. 109, 169).

6. CONCLUSION

The concept of shrinkage discussed here involves the idea that the efficiency of procedures for estimating the population mean μ can be improved when μ lies in an effective interval centred

on a predetermined guess μ_0. The usefulness of the shrinkage point estimators and shrinkage confidence intervals presented depends on the relationship of the unknown $\delta \equiv (\mu-\mu_0)/(\sigma/\sqrt{n})$ to the standardized length of the effective interval $\breve{\delta}$. The techniques are valuable when $|\delta|$ is a priori thought to be smaller than $\breve{\delta}$, so we would look for applications where σ is expected to be large relative to $|\mu - \mu_0|$ and where the sample size is small.

The methods in this paper have been discussed under the view that the chief measure of the quality of the shrinkage procedure is the size of $\breve{\delta}$. Section 2 showed how the value of $\breve{\delta}$ could be increased in the point estimation problem by sacrificing efficiency in situations where a shrinkage estimator would not usually be applied anyway. The estimator favoured by the author is $Q_2'(0.4, 0.01, 0.5)$, which has $\breve{\delta} \approx 3.5$ and has root mean squared error smaller than that of $\overline{X}$ by approximately 10% for all $\delta \lessapprox 2.5$, i.e., mean squared error smaller by approximately 20%.

Section 3 considered the problem of estimating the mean of a normal distribution using a confidence interval, and presented two double-sided shrinkage confidence intervals. Interval (9) appears to have better performance than interval (12). When $n = 4$, the 99% and 95% confidence intervals for μ obtained using (9) have $\breve{\delta} \approx 5.8$ and $\breve{\delta} \approx 3.2$ respectively, and both can be more efficient than the corresponding standard interval by up to 20% in mean interval width. The midpoint of interval (9) performs well as an accompanying point estimator. The construction of valid one-sided shrinkage confidence intervals would be a further useful development.

APPENDIX

The probability of success of interval (10) is given in equation (11), which can be written as

$$\begin{aligned}
p^* &= \Pr(|T| \le t_1) + \Pr(t_1 < |T| \le t_3 \cap |T'| > t_2) \\
&= \Pr(|T| \le t_1) \\
&\quad + \Pr(T \le t_3 \cap T' > t_2) - \Pr(T < t_1 \cap T' > t_2) \\
&\quad + \Pr(T \le t_3 \cap T' < -t_2) - \Pr(T < t_1 \cap T' < -t_2) \\
&\quad + \Pr(T \ge -t_3 \cap T' > t_2) - \Pr(T > -t_1 \cap T' > t_2) \\
&\quad + \Pr(T \ge -t_3 \cap T' < -t_2) - \Pr(T > -t_1 \cap T' < -t_2)
\end{aligned} \tag{13}$$

for evaluation using equations for tail probabilities of Owen's doubly non-central t-distribution given by Chou (1992, Eq. 1.1). Equation (13) in turn can be written as $p^* = \Pr(|T| \le t_1) + A - B + C - D + E - F + G - H$ with the terms being labelled in the order of their appearance. Then

$$\begin{aligned}
A &= \Pr(T \le t_3 \cap T' \ge t_2) = ch(t_3, t_2, n-1, 0, \delta), \\
B &= \Pr(T \le t_1 \cap T' \ge t_2) = ch(t_1, t_2, n-1, 0, \delta), \\
C &= \Pr(T \le t_3) - \Pr(T \le t_3 \cap T' \ge -t_2) = pt(t_3, n-1) - ch(t_3, -t_2, n-1, 0, \delta), \\
D &= \Pr(T \le t_1) - \Pr(T \le t_1 \cap T' \ge -t_2) = pt(t_1, n-1) - ch(t_1, -t_2, n-1, 0, \delta), \\
E &= \Pr(T \ge -t_3) - \Pr(T' \le t_2 \cap T \ge -t_3) = pt(t_3, n-1) - ch(t_2, -t_3, n-1, \delta, 0), \\
F &= \Pr(T \ge -t_1) - \Pr(T' \le t_2 \cap T \ge -t_1) = pt(t_1, n-1) - ch(t_2, -t_1, n-1, \delta, 0), \\
G &= \Pr(T' \le -t_2 \cap T \ge -t_3) = ch(-t_2, -t_3, n-1, \delta, 0), \\
H &= \Pr(T' \le -t_2 \cap T \ge -t_1) = ch(-t_2, -t_1, n-1, \delta, 0).
\end{aligned}$$

Here $pt(\cdot)$ indicates the probability integral of the t-distribution and $ch(\cdot)$ indicates the probability given in Chou's equation 1.1, which would be written as the function $ch(t_1, t_2, f, \delta_1, \delta_2)$ using Chou's notation for the arguments. This function involves the function

$$Q_f(t_i, \delta_i; 0, R) \equiv \frac{\sqrt{(2\pi)}}{\Gamma(f/2)2^{f/2-1}} \int_0^R \Phi\left(\frac{t_i x}{\sqrt{f}} - \delta_i\right) x^{f-1} \phi(x)\, dx \tag{14}$$

of Owen (1965), in which $\Phi(\,\cdot\,)$ and $\phi(\,\cdot\,)$ are the standard normal distribution and density functions and $0 \leq R < \infty$. To evaluate the integral in (14) we make the transformation $y = x/(1+x)$, which sets the upper limit of integration to be the finite quantity $S = R/(1+R)$. The integral becomes $\int_0^S L(y)\,dy$, where

$$L(y) \equiv \Phi\left(\frac{t_i y}{(1-y)\sqrt{f}} - \delta_i\right)\phi\left(\frac{y}{1-y}\right)\frac{y^{f-1}}{(1-y)^{f+1}}.$$

The integral is then accurately approximated by $S/10^4 \times \sum_{k=1}^{10^4} L(kS/10^4 - 0.5S/10^4)$. Replacing 10^4 by 10^3 makes no difference to the minimum probabilities 99.11%, 99.02%, 95.47% and 95.11% quoted in Section 3.2. This implies that 10^4 is sufficiently large for adequate results to be obtained with this form of numerical integration.

REFERENCES

J. C. Arnold (1969). A modified technique for improving an estimate of the mean. *Biometrics*, 25, 588–591.

Y.-M. Chou (1992). A bivariate noncentral t-distribution with applications. *Communications in Statistics: Theory and Methods*, 21, 3427–3462.

E. I. George & G. Casella (1994). An empirical Bayes confidence report. *Statistica Sinica*, 4, 617–638.

C. Kazimi & D. Brownstone (1999). Bootstrap confidence bands for shrinkage estimators. *Journal of Econometrics*, 90, 99–127.

H. H. Lemmer (1981a). Note on shrinkage estimators for the binomial distribution. *Communications in Statistics A: Theory and Methods*, A10, 1017–1027.

H. H. Lemmer (1981b). From ordinary to Bayesian shrinkage estimators. *South African Statistical Journal*, 15, 57–72.

H. H. Lemmer (1988). Shrinkage estimators. In *Encyclopedia of Statistical Sciences, Volume 8* (S. Kotz, N. L. Johnson & C. B. Read, eds.), Wiley, New York, pp. 452–456.

J. S. Mehta & R. Srinivasan (1971). Estimation of the mean by shrinkage to a point. *Journal of the American Statistical Association*, 66, 86–90.

R. J. Muirhead (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

A. O'Hagan (1994). *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference.* Edward Arnold, London.

D. B. Owen (1965). A special case of a bivariate noncentral t-distribution. *Biometrika*, 52, 437–446.

Y. Takada (1998). Asymptotic improvement of the usual confidence set in a multivariate normal distribution with unknown variance. *Journal of Multivariate Analysis*, 64, 118–130.

J. R. Thompson (1968a). Some shrinkage techniques for estimating the mean. *Journal of the American Statistical Association*, 63, 113–122.

J. R. Thompson (1968b). Accuracy borrowing in the estimation of the mean by shrinkage to an interval. *Journal of the American Statistical Association*, 63, 952–962.

Y.-L. Tseng (2004). Set-induced minimax estimators for a multivariate normal mean. *Journal of Statistical Planning and Inference*, 121, 53–64.

Y.-L. Tseng & L. D. Brown (1997). Good exact confidence sets for a multivariate normal mean. *The Annals of Statistics*, 25, 2228–2258.

Received 7 November 2007
Accepted 18 June 2008

Robin WILLINK: r.willink@irl.cri.nz
Industrial Research Limited
P.O. Box 31-310
Lower Hutt 5040, New Zealand