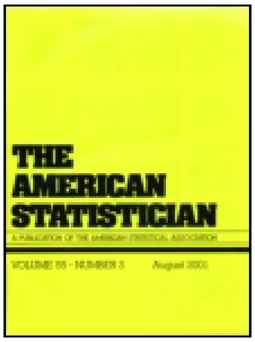


The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

# An Effective Confidence Interval for the Mean With Samples of Size One and Two

**Melanie M Wall, James Boen & Richard Tweedie**

**To cite this article:** Melanie M Wall, James Boen & Richard Tweedie (2001) An Effective Confidence Interval for the Mean With Samples of Size One and Two, The American Statistician, 55:2, 102-105, DOI: 10.1198/000313001750358400

**To link to this article:** https://doi.org/10.1198/000313001750358400

# An Effective Confidence Interval for the Mean With Samples of Size One and Two

Melanie M. WALL, James BOEN, and Richard TWEEDIE

It is counterintuitive that, with a sample of only one value from a normal distribution, one can construct a finite confidence interval of any size for the mean. It goes just as much against standard teaching that from a sample of size two such a CI might be shorter than that based on the $t$ statistic. We refine an earlier version of this first result, and use it to prove the second. For samples of three and larger, we show that the $t$-based interval cannot be improved using this approach.

KEY WORDS: Confidence intervals; Invariant estimators; Location estimation; Student's $t$.

## 1. INTRODUCTION

It has been known, but not well known, since the 1960s that from a single observation $x$ from a normal distribution with unknown mean $\mu$ and unknown standard deviation $\pi$, it is possible to create a confidence interval (CI) for $\mu$ with finite length (Abbot and Rosenblatt 1963; Machol and Rosenblatt 1966; Machol 1966, 1967). This remarkable result seems to completely contradict the standard statistical intuition that at least two observations are necessary in order to have some idea about variability.

Nonetheless, it is a special case of an even more surprising result (Blachman and Machol 1987; Edelman 1990; Rodriguez 1998) that for any $\alpha \in (0, 1)$ there exists a finite $100(1-\alpha)\%$ CI for the mean of any unimodal distribution, of the form

$$x \pm \zeta\, |x|. \tag{1}$$

Based on increasingly strong assumptions about the underlying distribution (unimodal, unimodal-symmetric, or normal), the value of $\zeta$ decreases; and for each of the three classes, the relevant value of $\zeta$ is solely a function of the size $1-\alpha$, and thus is appropriate for all distributions in that class.

This article discusses the case where the underlying distribution is normal. We first refine the results in Edelman (1990) using the methods of Blachman and Machol (1987), and give a formula for $\zeta$ that results in shorter confidence intervals than those in Edelman (1990).

We then show that it is possible to use (1) with samples of size $n > 1$, and that given a sample of size $n = 2$ from a normal distribution, this CI can yield better results in some cases than the usual Student's $t$ interval. For $n > 2$, however, the usual Student's $t$ interval is always better.

We conclude with a discussion of the intuitive reasons behind these results, and in particular we discuss the apparent contradiction with the known optimality of the CI based on Student's $t$.

## 2. A CONFIDENCE INTERVAL FOR $n$ / 1

Let $X$ be a normal random variable with mean $\mu$ and standard deviation $\pi$. To derive (1) we want to find a value $\zeta$ such that

$$\Pr(X - \zeta\,|X| \le \mu \le X + \zeta\,|X|) \ge 1 - \alpha\ ,$$

no matter what the values of $\mu$ and $\pi$. For such an interval the coverage probability for $\zeta > 1$ is

$$\begin{aligned}
\Pr(X - \zeta\,|X| &\le \mu \le X + \zeta\,|X|) \\
&= 1 - \Pr(|X-\mu| \ge \zeta |X|) \\
&= 1 - \Pr\left( \left| \frac{X-\mu}{\pi} \right| \ge \zeta \left| \frac{X-\mu}{\pi} + \frac{\mu}{\pi} \right| \right) \\
&= 1 - \left[ \Phi\left( \frac{|\mu|}{\pi} \frac{\zeta}{\zeta - 1} \right) - \Phi\left( \frac{|\mu|}{\pi} \frac{\zeta}{\zeta+1} \right) \right],
\end{aligned}$$

where $\Phi$ is the standard normal cdf. Note that this depends only on the ratio $|\mu|/\pi$. To set the minimum coverage probability equal to $(1-\alpha)$, $\zeta$ needs to be chosen so that

$$\sup_{\frac{|\mu|}{\sigma}} \left[ \Phi\left( \frac{|\mu|}{\pi} \frac{\zeta}{\zeta - 1} \right) - \Phi\left( \frac{|\mu|}{\pi} \frac{\zeta}{\zeta+1} \right) \right] = \alpha\ . \tag{2}$$

We shall show that it is possible to solve numerically for the value of $\zeta$ that satisfies (2). Table 1 gives the exact values of $\zeta$ associated with several values of $\alpha$. So, for example, a $90\%$ confidence interval for $\mu$ is given by $x \pm 4.84|x|$. Prior to deriving the results in Table 1, we present two simple closed-form approximations for $\zeta$, using Figure 1 which displays the region where the confidence interval fails to cover $\mu$.

Edelman (1990) derived, as an approximation for $\zeta$, the value necessary to ensure the rectangle formed with the dotted lines in Figure 1 has area less than or equal to $\alpha$. This gives

$$\zeta \approx \frac{2\tau(1)}{\alpha} + 1, \tag{3}$$

where $\tau$ is the standard normal density. Since this area is always greater than the area under the normal density enclosed by the same upper and lower limit, this value for $\zeta$ is typically

Melanie M. Wall is Assistant Professor, James Boen is Professor, and Richard Tweedie is Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455 (E-mail: melanie@biostat.umn.edu).

Table 1. Values of $\zeta$ for constructing 100(1 − $\alpha$)% CIs for $\mu$ using (1)

| | $\alpha$ | | | | |
|---|---|---|---|---|---|
| | .20 | .15 | .10 | .05 | .01 |
| $\zeta$ | 2.42 | 3.23 | 4.84 | 9.68 | 48.39 |

rather conservative. For example, a 90% confidence interval for $\mu$, given by the approximation (3) from Edelman (1990) is $x \pm 5.84|x|$ which is wider than necessary, thus yielding higher than 90% coverage.

Blachman and Machol (1987) derived an approximation using the value of $\zeta$ needed so that the trapezoid formed by taking the first order approximation to the normal density at the point $|\mu|/\pi$ (bold line in Figure 1) has area less than or equal to $\alpha$. This gives

$$\zeta \approx \frac{2\tau(1)}{\alpha}. \tag{4}$$

The formula (4) is remarkably close to the exact numerical solution to (2) for $\zeta$. Numerical methods indicate that for all $\alpha < .3$, it is within .005 of the exact solution.

We now indicate how to derive the exact value of $\zeta$ which solves (2). For each $y$, denote by $(|\mu|/\pi)_y$ the "least favorable value" which maximizes the area $[\Phi((|\mu|/\pi)(y/(y-1))) - \Phi((|\mu|/\pi)(y/(y+1)))]$. In the Appendix, following the reasoning in Blachman and Machol (1987), we show that this least favorable value is

$$\left(\frac{|\mu|}{\pi}\right)_y = \left(1 - \frac{1}{y^2}\right)\sqrt{\frac{-y}{2}\log\left(\frac{y-1}{y+1}\right)}. \tag{5}$$

Thus, the $\zeta$ which solves (2) is the same as the $\zeta$ which solves

$$\Phi\left(\left(\frac{|\mu|}{\pi}\right)_\zeta \frac{\zeta}{\zeta-1}\right) - \Phi\left(\left(\frac{|\mu|}{\pi}\right)_\zeta \frac{\zeta}{\zeta+1}\right) = \alpha, \tag{6}$$

and (6) can easily be solved numerically using any software that can calculate quantiles of the normal distribution. The results are shown for common values of $\alpha$ in Table 1.

## 3. CONFIDENCE INTERVALS FOR $n > 1$

The confidence interval (1) was originally derived with the intention of being used when there was only one observation from a distribution. Of course, (1) can also be used to form confidence intervals for $\mu$ when we have more than one observation. If $x_1, x_2, \ldots x_n$ are iid observations from $N(\mu, \pi^2)$, then $\bar{x} = \sum_{i=1}^n x_i/n$ is distributed $N(\mu, \frac{\sigma^2}{n})$, and therefore, based on the results of the last section, we can form a $100(1-\alpha)\%$ confidence interval for $\mu$ by taking

$$\bar{x} \pm \zeta\,|\bar{x}|, \tag{7}$$

where $\zeta$ is from Table 1. Note that the value of $\zeta$ does not depend on $n$, so that these intervals do not get shorter as more data are collected.

The use of (7) when $n > 1$ may be of practical use if it can provide shorter confidence intervals than the usual confidence interval for $\mu$ formed with the sample standard deviation $s$ and the Student's $t$ distribution; that is,

$$\bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}, \tag{8}$$

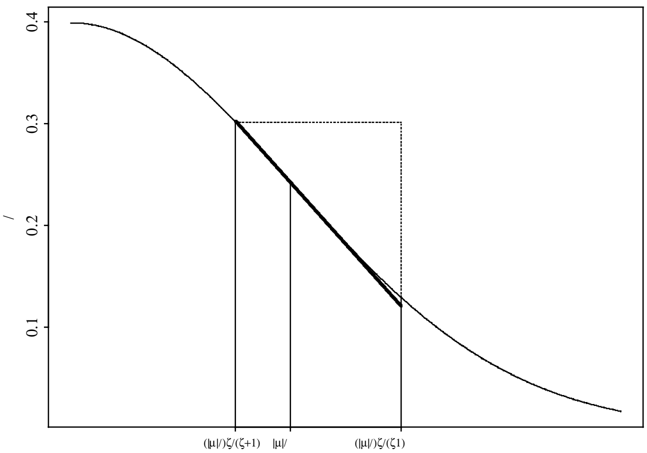


Figure 1. Region under the standard normal density where the confidence interval (1) fails to cover $\mu$. Rectangle formed by the dotted line relates to (3) and trapezoid formed by the bold line relates to (4).

where $t_{n-1}$ is the quantile associated with $\alpha/2$ from the Student's $t$ distribution with $n-1$ degrees of freedom. To compare the margin of error of (7) versus (8), we compare $E(\zeta|\bar{X}|)$ to $E(t_{n-1}\frac{S}{\sqrt{n}})$. The expression for $E(|\bar{X}|)$ is straightforward and an expression for $E(S)$ can be found in Cureton (1968). These lead to

$$\begin{aligned} E(\zeta|\bar{X}|) &= \zeta\, E(|\bar{X}|) \\ &= \zeta\left[\mu\left(1 - 2\Phi\left(\frac{-\mu\sqrt{n}}{\pi}\right)\right) + \pi\sqrt{\frac{2}{n\nu}}\; e^{-\frac{n\mu^2}{2\sigma^2}}\right], \end{aligned} \tag{9}$$

and

$$\begin{aligned} E\left(t_{n-1}\frac{S}{\sqrt{n}}\right) &= t_{n-1}\,\frac{E(S)}{\sqrt{n}} \\ &= t_{n-1}\,\frac{\pi}{\sqrt{n}}\,\Gamma\left(\frac{n}{2}\right)\Big/\left[\Gamma\left(\frac{n-1}{2}\right)\sqrt{\frac{n-1}{2}}\right]. \end{aligned} \tag{10}$$

To see that it is plausible that the margin of error for (7) might be smaller than that for (8) when $n = 2$, compare the quantiles of Student's $t$ distribution given in Table 2 to the values for $\zeta$ given in Table 1 for some values of $|\bar{x}|$ and $s$. We see that $t_1 > \zeta$ for all $\alpha$, although $t_2 < \zeta$ for all $\alpha$.

A careful comparison of $E(|\bar{X}|)$ and $E(S)/\sqrt{n}$ shows that when $n = 2$ there is indeed a region of the parameter space where (9) is smaller than (10). Figure 2 shows a perspective plot of $E(\zeta|\bar{X}|)$ and $E(t_{n-1}(S/\sqrt{n}))$ as functions of $\mu$ and $\pi$ for the case when $n = 2$ and $\alpha = .05$. To compare these surfaces we show their difference over the parameter space in Figure 3 (left plot). The region of interest is that where the $E(\zeta|\bar{X}|) < E(t_{n-1}(S/\sqrt{n}))$. The plot on the right of Figure 3 represents this region—that is, where the expected margin of error of (8) is larger than that of (7). We can see that when $n = 2$

Table 2. Quantile values of Student's t-distribution with one degree of freedom

| | $\alpha$ | | | | |
|---|---|---|---|---|---|
| $t_{n-1}$ | .20 | .15 | .10 | .05 | .01 |
| $t_1$ | 3.08 | 4.17 | 6.31 | 12.71 | 63.66 |
| $t_2$ | 1.89 | 2.28 | 2.92 | 4.30 | 9.92 |

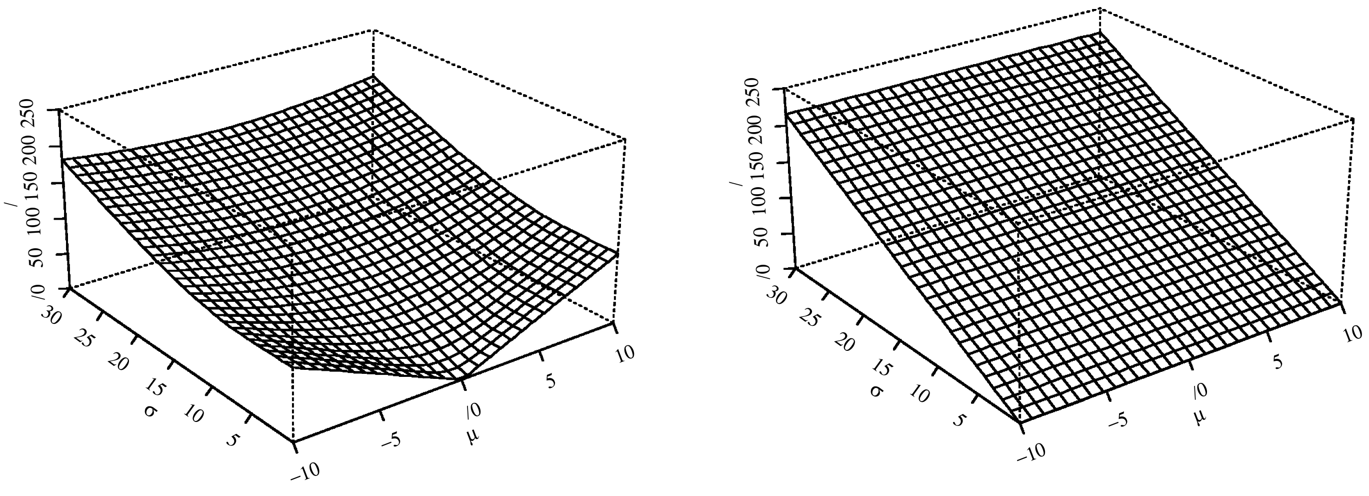


*Figure 2. Given n = 2, the height in each of these plots represents the expected margin of error for a 95% CI at a given $\mu$ and $\sigma$. On the left is $E(\zeta|X|)$ and on the right is $E(t_{n-1}(S/\sqrt{n}))$.*

and $|\mu|/\sigma$ is less than approximately $1/2$, (7) leads to a shorter (in expectation) confidence interval than (8). The amount of improvement in accuracy increases as sigma increases. On the other hand, if $|\mu|/\sigma >> \frac{1}{2}$ the average margin of error of (7) can become much worse than that of (8). To see this, consider what happens to the margin of error in Figure 3 (left plot) when $\sigma$ is held constant and $|\mu|$ increases.

Outperforming the CI (8) using iid normal observations might seem like an impossible task given the well known and widely taught optimality properties of (8). The key fact to recall is that the CI (8) is the uniformly most accurate *equivariant* confidence interval for the parameter $\mu$ (Lehman 1986, p. 329). Note that the random variable $(X-\mu)/|X|$ used in forming the confidence interval (7) is not a pivotal quantity since it depends on both $\mu$ and $\sigma$. Thus, the CI $\bar{x} \pm \zeta|\bar{x}|$ is not in the equivariant class of confidence intervals, and so it is indeed possible that for some points in the parameter space of $\mu$ and $\sigma$, (7) can be more accurate than (8).

All of the above refers to the situation $n = 2$. When $n > 2$ the average margin of error for (7) can be shown to be uniformly larger than the average margin of error for (8), and so this non-equivariant method is not practical for larger samples where the standard methods are better. The following proposition demonstrates this result for the most useful range of $\alpha$.

*Proposition:* For $n > 2$ and $\alpha \in (0, .20]$

$$E(\zeta|\bar{X}|) > E\left(t_{n-1}\frac{S}{\sqrt{n}}\right). \tag{11}$$

*Proof:* Note that $\min_\mu E(\zeta|\bar{X}|)$ occurs when $\mu = 0$. This can be seen by examining the derivative of $E(\zeta|\bar{X}|)$ with respect

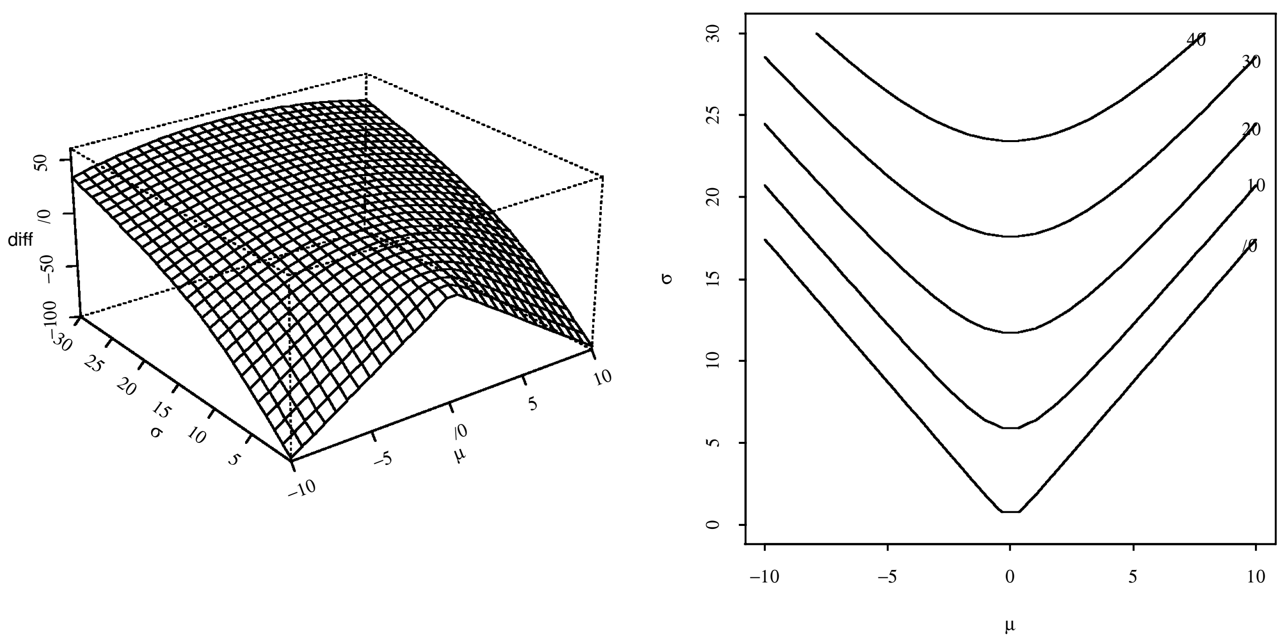


*Figure 3. The height of the plot on the left is the difference $E(t_{n-1}(S/\sqrt{n})) - E(\zeta|X|)$ based on a 95% CI with n = 2 and the plot on the right is a contour plot of the difference shown on the left but only for the region where (7) is more accurate than (8).*

to $\mu$ and observing that it only can equal zero when $\mu = 0$. Thus,

$$E(\zeta|\bar{X}|) > \sqrt{\frac{2}{\pi}}\,\zeta\,\frac{\sigma}{\sqrt{n}} > .797\,\zeta\,\frac{\sigma}{\sqrt{n}}\,.$$

Now $\Gamma\left(n/2\right) \Big/ \left[\Gamma\left((n-1)/2\right)\sqrt{(n-1)/2}\right] < 1$ for all $n$, and so for $n \geq 3$

$$E\left(t_{n-1}\frac{S}{\sqrt{n}}\right) < t_{n-1}\,\frac{\sigma}{\sqrt{n}} \leq t_2\frac{\sigma}{\sqrt{n}}\,.$$

Hence, we can conclude (11), if we can show that $t_2 < .797\,\zeta$ for all $\alpha \in (0, .20]$. Now, $t_2 = \sqrt{2(1-\alpha)^2/(2\alpha-\alpha^2)}$ (Johnson and Kotz 1970, p. 112). If we use the numerically demonstrated fact that (4) is within .005 from the true $\zeta$ for $\alpha \in (0, .20]$, it suffices to show $\sqrt{2(1-\alpha)^2/(2\alpha-\alpha^2)} \leq .797\left((2\phi(1))/\alpha - .005\right)$. The quantity $\alpha\sqrt{2(1-\alpha)^2/(2\alpha-\alpha^2)}$ is monotonically increasing since the the first derivative with respect to $\alpha$ is always positive when $\alpha \in (0, .20]$. Thus, from Table 1 and 2 we conclude that $\max_{\alpha \in (0,.20]}(t_2/\zeta) = 1.89/2.42 < .797$ as required.

## 3. CONCLUSION

Why does this totally implausible approach work?

To try and give an intuition for this, consider a single observation of $x = 10$. This could come from a normal distribution with $\mu = 0$ and $\sigma = 10$, or from a normal distribution with $\mu = 10$ and $\sigma = 1$. However, "common sense" suggests that it can hardly come from $\mu = 1,000$ and $\sigma = 20$. This reasoning indicates that some $\mu$, $\sigma$ pairs are highly unlikely. Thus, any confidence region in $(\mu, \sigma)$-space should exclude some $\mu$, $\sigma$ pairs, and should not be infinite.

The rather surprising result, even given this, is that there is marginally a finite CI for $\mu$ which is correct no matter what the unknown value of $\sigma$. Again, however, one might argue that if $x = 10$, then no matter what the value of $\mu$ there is little probability that $\sigma$ will be 100,000 or more. Thus we are intuitively working in a "rectangle" in $(\mu, \sigma)$-space and the marginal result of that calculation leads to the finite CI for $\mu$ alone. The foregoing proof formalizes this argument, showing that without any appeal to prior distributions, we can calculate the actual size of the finite CI.

When is this sort of work useful?

Clearly there is a good use for this example in the classroom. We teach that one needs an idea of variability in order to do estimation: it is useful to hone this intuition with examples such as this, to make us realize that $x$ itself tells us something about the variability as well as the mean. We also teach that Student's $t$ leads to a uniformly most accurate CI. We do not always mention, and certainly do not always stress, that this only applies if we restrict ourselves to *equivariant* CIs. It is valuable to have a nonartificial example that shows the need for the equivariance in this statement.

Whether the approach is useful in practice is difficult to judge. The authors who largely developed this approach (Blachman and Machol 1987; Machol and Rosenblatt 1966; Machol 1966, 1967) are from NASA, and it appears plausible that in their work, small samples of 1 or 2 really do exist. There are many other experiments where such sample sizes also apply. In these cases, using (7) may well lead to greater accuracy of estimation.

## APPENDIX

The result of this lemma is essentially given in Blachman and Machol (1987) without proof, but we give it here with proof because Blachman and Machol (1987) is not trivial to read.

*Lemma:* Let $a = [\Phi((|\mu|/\sigma)(y/(y-1))) - \Phi((|\mu|/\sigma)(y/(y+1)))]$. Then the value of $|\mu|/\sigma$ which maximizes $a$ is

$$\left(\frac{|\mu|}{\sigma}\right)_y = \left(1 - \frac{1}{y^2}\right)\sqrt{\frac{-y}{2}\log\left(\frac{y-1}{y+1}\right)}.$$

*Proof:* Define $v \equiv |\mu|/\sigma$. Setting $(\partial a/\partial v)$ equal to zero and solving for $v$ gives successively:

$$\begin{aligned}
\phi\left(\frac{yv}{y-1}\right)\frac{y}{y-1} - \phi\left(\frac{yv}{y+1}\right)\frac{y}{y+1} &\equiv 0 \\
e^{-\frac{v^2}{2}\left[\left(\frac{y}{y-1}\right)^2 - \left(\frac{y}{y+1}\right)^2\right]} &\equiv \frac{y-1}{y+1} \\
-\frac{v^2}{2}\left[\left(\frac{y(y+1)}{y^2-1}\right)^2 - \left(\frac{y(y-1)}{y^2-1}\right)^2\right] &\equiv \log\left(\frac{y-1}{y+1}\right) \\
-\frac{v^2}{2}\,\frac{4y^3}{(y^2-1)^2} &\equiv \log\left(\frac{y-1}{y+1}\right) \\
v &\equiv \left(1 - \frac{1}{y^2}\right)\sqrt{\frac{-y}{2}\log\left(\frac{y-1}{y+1}\right)} \\
v &\equiv \left(1 - \frac{1}{y^2}\right)\sqrt{y\,\operatorname{archcot} y}
\end{aligned}$$

*[Received April 2000. Revised October 2000.]*

## REFERENCES

Abbot, J. H., and Rosenblatt, J. (1963), "Two Stage Estimation With One Observation on the First Stage," *Annals of the Institute of Statistical Mathematics*, 14, 229–235.

Blachman, N. M., and Machol, R. E. (1987), "Confidence Intervals Based on One or More Observations," *IEEE Transactions on Information Theory*, IT-33, No. 3, 373–382.

Cureton, E. E. (1968), "Unbiased Estimation of the Standard Deviation," *The American Statistician*, 22, 22.

Edelman, D. (1990), "A Confidence Interval for the Center of an Unknown Unimodal Distribution Based on a Sample Size 1," *The American Statistician*, 44, 285–287.

Goldstein, M. (1981), "Revising Previsions: A Geometric Interpretation," *Journal of the Royal Statistical Society*, Ser. B, 43, 105–130.

Johnson, N., and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions–2*, New York: Wiley.

Lehman, E. L. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: Wiley.

Machol, R. E. (1966), "Confidence Interval Based on Single Observation," *Proceedings of the IEEE*, 54, 1976.

——— (1967), "Confidence Interval Based on Single Observation," *Proceedings of the IEEE*, 55, 122.

Machol, R. E., and Rosenblatt, J. (1966), "Confidence Interval Based on Single Observation," *Proceedings of IEEE*, 54, 1087–1088.

Rodriguez, C. C. (1998), "Confidence Intervals From One Observation," http://omega.albany.edu:8008/confint.html.