An effective CI for the mean with samples of size 1 and 2

Melanie Wall, James Boen and Richard Tweedie[1]

Abstract

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. A "partly-Bayes" approach indicates that in some instances, with such small samples a quite reasonable confidence interval can be constructed.

keywords location estimation, Student's t, invariant estimators, confidence intervals

1 Introduction

It has been known, but not well known, since the 1960's that from a single observation x from a normal distribution with unknown mean μ and unknown standard deviation σ, it is possible to create a confidence interval (CI) for μ with finite length [1, 8, 9, 10]. 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 [2, 4, 11] 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 ζ decreases; and for each of the three classes, the relevant value of ζ is solely a function of the size $1-\alpha$, and thus is appropriate for all distributions in that class.

In this paper, we discuss the case where the underlying distribution is normal. We first refine the results in [4] using the methods of [2], and give a formula for ζ that results in shorter confidence intervals than those in [4].

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 then discuss intervals of the form

$$x \pm \zeta\ |x - A| \tag{2}$$

where A is some fixed predetermined constant. These have the same coverage properties as do those in (1), but if (using a prior assessment of the likely location of μ) we are able to pick A appropriately, we will get shorter intervals from this approach.

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455.

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 μ and standard deviation σ. To derive (1) we want to find a value ζ such that

$$P(X - \zeta\,|X| \leq \mu \leq X + \zeta\,|X|) \geq 1 - \alpha\,,$$

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

$$\begin{aligned} P(X - \zeta\,|X| \leq \mu \leq X + \zeta\,|X|) &= 1 - P(|X-\mu| \geq \zeta|X|) \\ &= 1 - P\left(\left|\frac{X-\mu}{\sigma}\right| \geq \zeta\,\left|\frac{X-\mu}{\sigma} + \frac{\mu}{\sigma}\right|\right) \\ &= 1 - \left[\,\Phi\left(\frac{|\mu|}{\sigma}\frac{\zeta}{\zeta-1}\right) - \Phi\left(\frac{|\mu|}{\sigma}\frac{\zeta}{\zeta+1}\right)\right] \end{aligned}$$

where Φ is the standard normal CDF. Note that this depends only on the ratio $\frac{|\mu|}{\sigma}$. In order to set the minimum coverage probability equal to $(1-\alpha)$, ζ needs to be chosen so that

$$\sup_{\frac{|\mu|}{\sigma}} \left[\Phi\left(\frac{|\mu|}{\sigma}\frac{\zeta}{\zeta-1}\right) - \Phi\left(\frac{|\mu|}{\sigma}\frac{\zeta}{\zeta+1}\right)\right] = \alpha\,. \tag{3}$$

We shall show that is possible to solve numerically for the value of ζ that satisfies (3). Table 1 gives the exact values of ζ associated with several values of α, assessed using (7) below.

Table 1: Values of ζ for constructing $100(1-\alpha)\%$ CIs for μ using (1)

α	.20	.15	.10	.05	.01
ζ	2.42	3.23	4.84	9.68	48.39

So, for example, a 90% confidence interval for μ is given by $x \pm 4.84|x|$. Note that this is narrower than the range $x \pm 5.84|x|$ given by the approximation (4) from [4].

Prior to deriving the results in Table 1, we present two simple closed-form approximations for ζ, using Figure 1 which displays the region where the confidence interval fails to cover μ.

Edelman [4] derives, as an approximation for ζ, the value necessary to ensure the rectangle formed with the dotted lines in Figure 1 has area less than or equal to α. This gives

$$\zeta \approx \frac{2\phi(1)}{\alpha} + 1 \tag{4}$$

where ϕ 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 ζ is typically rather conservative.

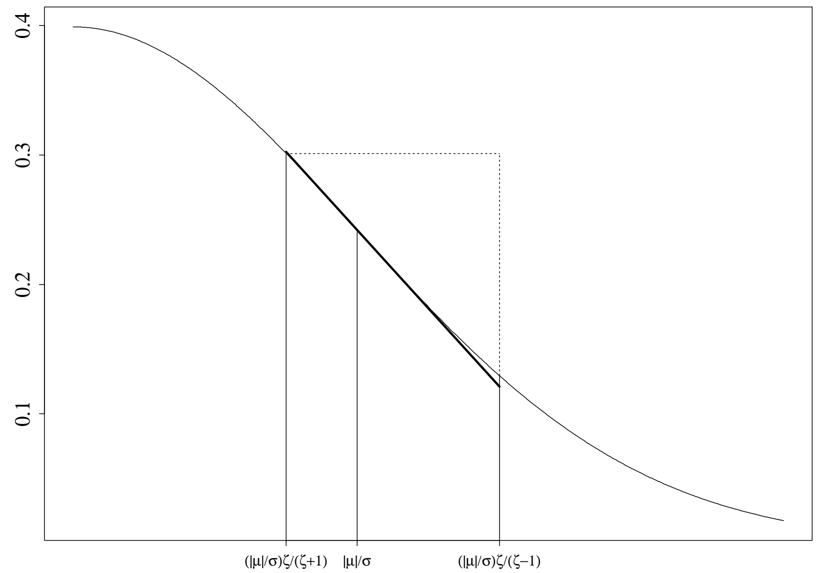

Figure 1: Region under the standard normal density where the confidence interval (1) fails to cover μ. Rectangle formed by the dotted line relates to (4) and trapezoid formed by the bold line relates to (5).

Blachman and Machol [2] derive an approximation using the value of ζ needed so that the trapezoid formed by taking the first order approximation to the normal density at the point $\frac{|\mu|}{\sigma}$ (bold line in Figure 1) has area less than or equal to α. This gives

$$\zeta \approx \frac{2\phi(1)}{\alpha}. \tag{5}$$

The formula (5) is remarkably close to the exact numerical solution to (3) for ζ. 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 ζ which solves (3). For each y, denote by $\left(\frac{|\mu|}{\sigma}\right)_y$ the "least favorable value" which maximizes the area $\left[\Phi(\frac{|\mu|}{\sigma}\frac{y}{y-1}) - \Phi(\frac{|\mu|}{\sigma}\frac{y}{y+1})\right]$. In the Appendix, following the reasoning in [2] we show that this least favorable value is

$$\left(\frac{|\mu|}{\sigma}\right)_y = \left(1 - \frac{1}{y^2}\right)\sqrt{\frac{-y}{2}\log\left(\frac{y-1}{y+1}\right)}. \tag{6}$$

Thus the ζ which solves (3) is the same as the ζ which solves

$$\Phi\left(\left(\frac{|\mu|}{\sigma}\right)_\zeta \frac{\zeta}{\zeta-1}\right) - \Phi\left(\left(\frac{|\mu|}{\sigma}\right)_\zeta \frac{\zeta}{\zeta+1}\right) = \alpha, \tag{7}$$

and (7) can easily be solved numerically using any software that can calculate quantiles of the normal distribution. The results are shown for common values of α 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 μ when we have more than one observation. If $x_1, x_2, \ldots x_n$ are iid observations from $N(\mu, \sigma^2)$, 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 μ by taking

$$\bar{x} \pm \zeta\,|\bar{x}| \tag{8}$$

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

The use of (8) when $n > 1$ may be of practical use if it can provide shorter confidence intervals than the usual confidence interval for μ formed with the sample standard deviation s and the Student's t-distribution, i.e.,

$$\bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}\ , \tag{9}$$

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 (8) versus (9), we compare $E(\zeta|\bar{X}|)$ to $E(t_{n-1}\frac{S}{\sqrt{n}})$. We derive an expression for $E(|\bar{X}|)$ in the Appendix for completeness, and an expression for $E(S)$ can be found in [3]. These lead to

$$E(\zeta|\bar{X}|) = \zeta\, E(|\bar{X}|) = \zeta\ \left[\mu\left(1 - 2\Phi\left(\frac{-\mu\sqrt{n}}{\sigma}\right)\right) + \sigma\sqrt{\frac{2}{n\pi}}\ e^{-\frac{n\mu^2}{2\sigma^2}}\right] \tag{10}$$

and

$$E(t_{n-1}\frac{S}{\sqrt{n}}) = t_{n-1}\ \frac{E(S)}{\sqrt{n}} = t_{n-1}\ \frac{\sigma}{\sqrt{n}}\ ,\ \left(\frac{n}{2}\right)\Big/\left[,\ \left(\frac{n-1}{2}\right)\sqrt{\frac{n-1}{2}}\right]\ . \tag{11}$$

To see that it is plausible that the margin of error for (8) might be smaller than that for (9) when $n = 2$, compare the quantiles of Student's t distribution given in Table 2 to the values for ζ given in Table 1 for some values of $|\bar{x}|$ and s. We see that $t_1 > \zeta$ for all α, although $t_2 < \zeta$ for all α.

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

	α				
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

A careful comparison of $E(|\bar{X}|)$ and $\frac{E(S)}{\sqrt{n}}$ shows that when $n = 2$ there is indeed a region of the parameter space where (10) is smaller than (11). Figure 2 shows a perspective plot of $E(\zeta|\bar{X}|)$ and $E(t_{n-1}\frac{S}{\sqrt{n}})$ as functions of μ and σ for the case when $n = 2$ and $\alpha = .05$. In order 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 $E(\zeta|\bar{X}|) < E(t_{n-1}\frac{S}{\sqrt{n}})$. The plot on the right of Figure 3 represents this region, i.e. where the expected margin of error of (9) is larger than that of (8). We can see that

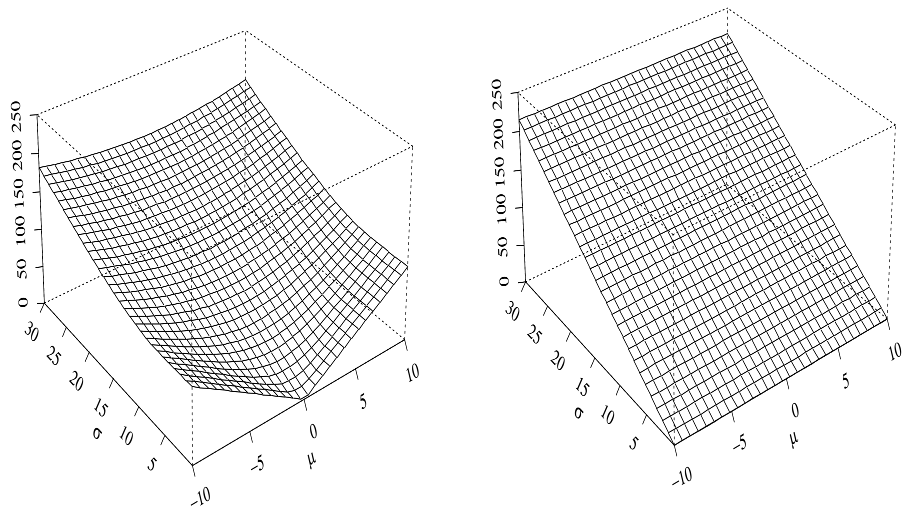

Figure 2: Given $n = 2$, the height in each of these plots represents the expected margin of error for a 95% CI at a given μ and σ. On the left is $E(\zeta|\bar{X}|)$ and on the right is $E(t_{n-1}\frac{S}{\sqrt{n}})$

when $n = 2$ and $\frac{|\mu|}{\sigma}$ is less than approximately $\frac{1}{2}$, (8) leads to a shorter (in expectation) confidence interval than (9). The amount of improvement in accuracy increases as sigma increases. On the other hand, if $\frac{|\mu|}{\sigma} >> \frac{1}{2}$ the average margin of error of (8) can become much worse than that of (9). To see this, consider what happens to the margin of error in Figure 3 (left plot) when σ is held constant and $|\mu|$ increases.

Outperforming the CI (9) using iid normal observations might seem like an impossible task given the well known and widely taught optimality properties of (9). The key fact to recall is that the CI (9) is the Uniformly Most Accurate *Equivariant* confidence interval for the parameter μ ([7], p. 329). Note that the random variable $\frac{\bar{X}-\mu}{|\bar{X}|}$ used in forming the confidence interval (8) is not a pivotal quantity since it depends on both μ and σ. 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 μ and σ, (8) can be more accurate than (9).

All of the above refers to the situation $n = 2$. When $n > 2$ the average margin of error for (8) can be shown to be uniformly larger than the average margin of error for (9), 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 α.

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

$$E(\zeta|\bar{X}|) > E(t_{n-1}\frac{S}{\sqrt{n}}) \tag{12}$$

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

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

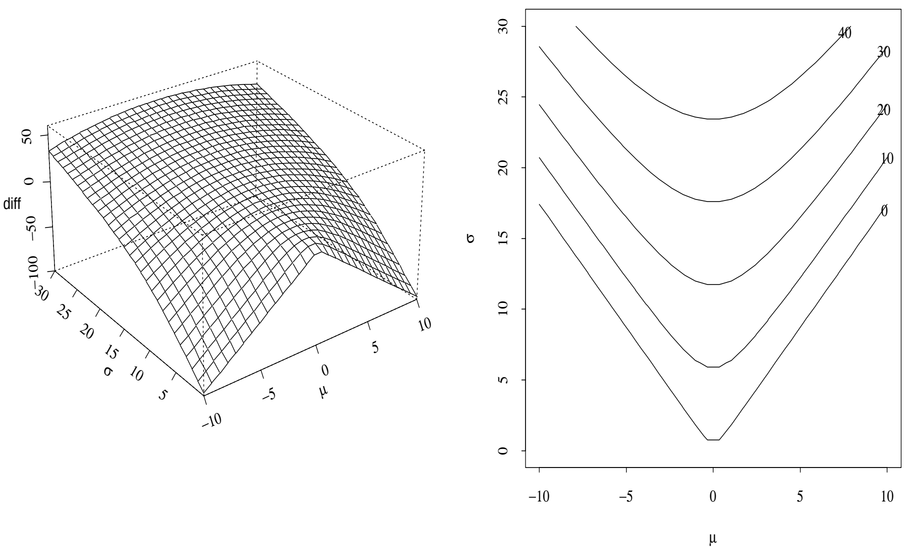

Figure 3: The height of the plot on the left is the difference $E(t_{n-1}\frac{S}{\sqrt{n}}) - E(\zeta|\bar{X}|)$ based on a 95% CI with $n = 2$ and the plot on the right is a contour plot of the difference shown on the left but only for the region where (8) is more accurate than (9)

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

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

Hence, we can conclude (12), if we can show that $t_2 < .797\,\zeta$ for all $\alpha \in (0, .20]$. Now, $t_2 = \sqrt{\frac{2(1-\alpha)^2}{2\alpha-\alpha^2}}$ ([6], p. 112). If we use the numerically demonstrated fact that (5) is within .005 from the true ζ for $\alpha \in (0, .20]$, it suffices to show $\sqrt{\frac{2(1-\alpha)^2}{2\alpha-\alpha^2}} \leq .797\,(\frac{2\phi(1)}{\alpha} - .005)$. The quantity $\alpha\sqrt{\frac{2(1-\alpha)^2}{2\alpha-\alpha^2}}$ is monotonically increasing since the the first derivative w.r.t α is always positive when $\alpha \in (0, .20]$. Thus, from Table 1 and 2 we conclude that $\max_{\alpha\in(0,.20]} \frac{t_2}{\zeta} = \frac{1.89}{2.42} < .797$ as required. ∎

4 A partially Bayesian approach

Consider now the more general confidence interval of the form

$$x \pm \zeta\,|x - A| \tag{13}$$

where A is some fixed predetermined constant. In our development so far we have assumed $A = 0$. Since A is fixed, the coverage probability of (13) for μ is the same as that developed in Section 2 and thus the values for ζ given in Table 1 will still apply. On the other hand, the choice of A will clearly affect the margin of error of the confidence interval. For the example we posed above, when $n = 2$ and $\alpha = .05$, the region in the parameter space where (13) would provide on average a smaller margin of error than (9) is where $\frac{|\mu-A|}{\sigma}$ is less than approximately $\frac{1}{2}$.

Thus, when we can use some form of prior information concerning the mean and standard deviation only (but not necessarily any other prior distributional properties) to choose A to be within a half a standard deviation of μ, we can provide a better 95% confidence interval with $n = 2$ using $\bar{x} \pm \zeta \, |\bar{x} - A|$ than using (9).

There have been a number of approaches to eliciting prior information on moments [5]. Given the full prior distributions on μ and σ, we also can calculate the prior probability of events such as $\frac{|\mu - A|}{\sigma} < \frac{1}{2}$ for some A, and thus of the prior probability that (8) will be more accurate than (9).

5 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" suggest that it can hardly come from $\mu = 1000$ and $\sigma = 20$. This reasoning indicates that some μ, σ pairs are highly unlikely. Thus any confidence region in (μ, σ)-space should exclude some μ, σ pairs, and should not be infinite.

The rather surprising result, even given this, is that there is marginally a finite CI for μ which is correct no matter what the unknown value of σ. Again, however, one might argue that if $x = 10$ then no matter what the value of μ there is little probability that σ will be 100,000 or more. Thus we are intuitively working in a "rectangle" in (μ, σ)-space and the marginal result of that calculation leads to the finite CI for μ alone. The proof above 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 non-artificial 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 [2, 8, 9, 10] 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 (8) may well lead to greater accuracy of estimation.

Appendix

The result of our first lemma is essentially given in [2] without proof, but we give it here with proof because [2] is not trivial to read.

Lemma: Let $a = \left[\Phi(\frac{|\mu|}{\sigma}\frac{y}{y-1}) - \Phi(\frac{|\mu|}{\sigma}\frac{y}{y+1})\right]$. Then the value of $\frac{|\mu|}{\sigma}$ 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 $v \equiv \frac{|\mu|}{\sigma}$. Setting $\frac{\partial a}{\partial v}$ equal to zero and solving for v gives successively:

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

Our second lemma is standard, but we give it for completeness.

Lemma: Given $X \sim N(\mu, \sigma^2)$

$$E(|X|) = \mu\left(1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right) + \sigma\sqrt{\frac{2}{\pi}}e^{-\frac{\mu^2}{2\sigma^2}}$$

Proof:

$$\begin{aligned}
E(|X|) &= \int_{-\infty}^{-\infty} |X|\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\mu)^2}{2\sigma^2}}dX \\
&= \int_{0}^{\infty} X\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\mu)^2}{2\sigma^2}}dX - \int_{-\infty}^{0} X\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\mu)^2}{2\sigma^2}}dX \\
&= \int_{-\frac{\mu}{\sigma}}^{\infty} (\mu+U\sigma)\frac{1}{\sqrt{2\pi}}e^{-\frac{U^2}{2}}dU - \int_{-\infty}^{-\frac{\mu}{\sigma}} (\mu+U\sigma)\frac{1}{\sqrt{2\pi}}e^{-\frac{U^2}{2}}dU \\
&= \mu\left(1-2\Phi\left(-\frac{\mu}{\sigma}\right)\right) + \frac{\sigma}{2\sqrt{2\pi}}\int_{\frac{\mu^2}{\sigma^2}}^{\infty} e^{-\frac{V}{2}}dV - \frac{\sigma}{2\sqrt{2\pi}}\int_{-\infty}^{\frac{\mu^2}{\sigma^2}} e^{-\frac{V}{2}}dV \\
&= \mu\left(1-2\Phi\left(-\frac{\mu}{\sigma}\right)\right) + \sigma\sqrt{\frac{2}{\pi}}e^{-\frac{\mu^2}{2\sigma^2}} \qquad \blacksquare
\end{aligned}$$

References

[1] 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.

[2] 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.

[3] Cureton, E.E. (1968). Unbiased estimation of the standard deviation. *The American Statistician*, **22**, No. 1, 22.

[4] Edelman, D. (1990). A confidence interval for the center of an unknown unimodal distribution based on a sample size 1. *The American Statistician*, **44**, No. 4, 285-287.

[5] Goldstein, M. (1981). Revising previsions: A geometric interpretation. *J.R. Statistical Society B*, **43**, 105-130.

[6] Johnson, N., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions - 2*. New York: John Wiley and Sons.

[7] Lehman, E.L. (1986). *Testing statistical hypotheses*, 2nd ed. New York: John Wiley and Sons.

[8] Machol, R.E., and Rosenblatt, J. (1966). Confidence interval based on single observation. *Proceedings of IEEE*, **54**, 1087-1088.

[9] Machol, R.E. (1966). Confidence interval based on single observation. *Proceedings of IEEE*, **54**, 1976.

[10] Machol, R.E. (1967). Confidence interval based on single observation. *Proceedings of IEEE*, **55**, 122.

[11] Rodriguez, C.C. (1998). Confidence intervals from one observation. Public webpage: *http://omega.albany.edu:8008/confint.html.*