United States
Department of
Agriculture

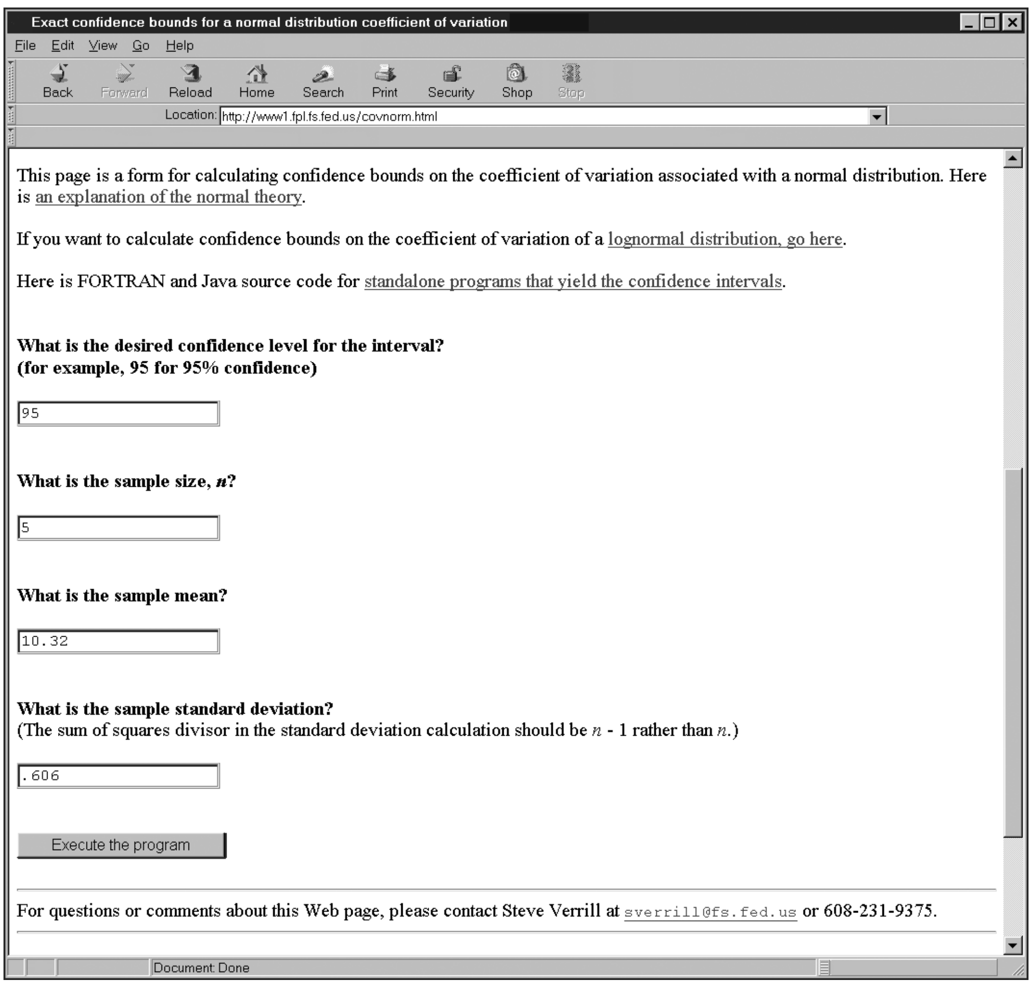
Forest Service

Forest
Products
Laboratory

Research
Paper
FPL–RP–609

Confidence Bounds for Normal and Lognormal Distribution Coefficients of Variation

Steve Verrill



Abstract

This paper compares the so-called exact approach for obtaining confidence intervals on normal distribution coefficients of variation to approximate methods. Approximate approaches were found to perform less well than the exact approach for large coefficients of variation and small sample sizes. Web-based computer programs are described for calculating confidence intervals on coefficients of variation for normally and lognormally distributed data.

Keywords: noncentral T, noise to signal ratio, variability, coefficient of variation, confidence bounds, normal, lognormal

May 2003

Verrill, Steve. 2003. Confidence bounds for normal and lognormal distribution coefficients of variation. Res. Pap. FPL-RP-609. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 13 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53726–2398. This publication is also available online at www.fpl.fs.fed.us. Laboratory publications are sent to hundreds of libraries in the United States and elsewhere.

The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

Contents

Confidence Bounds for Normal and Lognormal Distribution Coefficients of Variation

Steve Verrill, Mathematical Statistician
Forest Products Laboratory, Madison, Wisconsin

1 INTRODUCTION

The coefficent of variation (COV) of a distribution with mean μ and variance σ^2 is defined as the noise to signal ratio or σ/μ. (Sometimes this ratio is multiplied by 100 and reported as a percentage.) Building materials are often evaluated not only on the basis of mean strength but also on relative variability. Laboratory techniques are often compared on the basis of their COVs. Thus, scientists and engineers are interested in obtaining confidence intervals on population COVs.

We have developed easily accessible Web software that calculates confidence intervals on COVs for normal or lognormal distributions. A researcher who is interested in using this software, but who is not interested in the statistical theory, can skip to section 5.

Vangel (1996) discussed approximate methods for obtaining confidence intervals on COVs given normal data. He observed that the "exact" approach of Johnson and Welch (1940) is "computationally cumbersome." He also noted that the approximate methods are only appropriate for smaller COVs, and that McKay (1932) recommended that the approximations be used only for COV < 0.33.

In this paper, we review the exact approach that is appropriate for normally distributed data and describe Web resources that greatly ease any computational difficulties. We also demonstrate that there are cases in which COV < 0.33 and approximate approaches do not perform as well as the exact approach.

The appropriate technique for obtaining a confidence interval on the COV given lognormal data is comparatively straightforward. The statistical theory is described in Appendix I.

2 THE "EXACT" APPROACH

The "exact" approach to COV confidence intervals was outlined in Johnson and Welch (1940). Assume that we have a random sample of size n from a normal distribution with mean μ and variance σ^2. Then we have

$$\bar{X}/(S/\sqrt{n}\,) \sim F_{NCT}(n-1, \mu\sqrt{n}/\sigma)$$

where $F_{NCT}(n-1, \mu\sqrt{n}/\sigma)$ denotes a noncentral T distribution with $n-1$ degrees of freedom and noncentrality parameter $\mu\sqrt{n}/\sigma$, and $\bar{X}$ and S are the usual sample mean and standard deviation. Thus,

$$\mathrm{Prob}(A) = 1-\alpha \tag{1}$$

where A denotes the event

$$\{F_{NCT}(n-1, \sqrt{n}/\mathrm{COV})^{-1}(\alpha/2) \le \bar{X}/(S/\sqrt{n}\,) \le F_{NCT}(n-1, \sqrt{n}/\mathrm{COV})^{-1}(1-\alpha/2)\}$$

and

$$\mathrm{COV} \equiv \sigma/\mu$$

Suppose that COV is positive. If $\bar{X}$ is positive, define COV_L to be the solution to the nonlinear equation in β:

$$\alpha/2 = F_{NCT}(n-1, \sqrt{n}/\beta)(\bar{X}/(S/\sqrt{n})) \tag{2}$$

If $\bar{X}$ is positive, define COV_U to be the solution, if it exists, to the nonlinear equation in β:

$$1-\alpha/2 = F_{NCT}(n-1, \sqrt{n}/\beta)(\bar{X}/(S/\sqrt{n})) \tag{3}$$

If there is no solution to this equation, then set $\mathrm{COV}_U = \infty$. If $\bar{X}$ is negative, work with $-S/\bar{X}$ to obtain a confidence interval on $-\mathrm{COV}$ and then negate that confidence interval.

Now, for $\mathrm{COV} > 0$, we claim that

$$\{\bar{X} > 0\} \cap A = \{\mathrm{COV}_L \leq \mathrm{COV} \leq \mathrm{COV}_U\} \tag{4}$$

This is an immediate consequence of the fact that, for fixed x and fixed degrees of freedom k, $F_{NCT}(k, \gamma)(x)$ is a decreasing function of γ. For example, if the event on the left side of Equation (4) has occurred, then we have

$$\alpha/2 \leq F_{NCT}(n-1, \sqrt{n}/\mathrm{COV})(\bar{X}/(S/\sqrt{n})$$

and by the definition of COV_L we also have

$$\alpha/2 = F_{NCT}(n-1, \sqrt{n}/\mathrm{COV}_L)(\bar{X}/(S/\sqrt{n})$$

Thus, since $F_{NCT}(n-1, \gamma)(\bar{X}/(S/\sqrt{n})$ is a decreasing function of γ, we must have

$$\sqrt{n}/\mathrm{COV} \leq \sqrt{n}/\mathrm{COV}_L$$

or

$$\mathrm{COV}_L \leq \mathrm{COV}$$

From Equation (1) we have

$$1-\alpha = \mathrm{Prob}(A) \geq \mathrm{Prob}(\{\bar{X} > 0\} \cap A)$$
$$\geq \mathrm{Prob}(A) - \mathrm{Prob}(\bar{X} \leq 0) = 1 - \alpha - \mathrm{Prob}(\bar{X} \leq 0)$$

or (from Equation (4))

$$1-\alpha \geq \mathrm{Prob}(\{\mathrm{COV}_L \leq \mathrm{COV} \leq \mathrm{COV}_U\}) \geq 1 - \alpha - \mathrm{Prob}(\bar{X} \leq 0) \tag{5}$$

Thus, COV_L and COV_U almost yield an exact $1-\alpha$ confidence interval on COV. The discrepancy in the "exactness" is bounded by $\mathrm{Prob}(\bar{X} \leq 0)$. We discuss this discrepancy in the next section.

Note that by Equations (2) and (3), to obtain COV_L and COV_U we need a routine to calculate the noncentral T cumulative distribution function and a nonlinear equation solver. Presumably, this approach has been considered to be computationally cumbersome because of the need for these two routines. In section 5 we discuss an easy-to-use Web-based program that calculates COV_L and COV_U, and runs the cumbersome routines behind the scenes on the Web server.

3 PROBLEMS WITH THE "EXACT" APPROACH

Neglecting any computational difficulties, there are two relatively mild problems with the exact approach. First, from Equation (5) we see that because $\bar{X}$ can be negative even when μ is positive, the "exact" confidence interval is in fact very slightly nonconservative. However, for practical purposes we can ignore this problem because, in general, $\mathrm{Prob}(\bar{X} \leq 0)$ is negligible. We illustrate this in Table 1 with some relevant values for $\mathrm{Prob}(\bar{X} \leq 0) = \Phi(-\sqrt{n}/\mathrm{COV})$ where Φ is the N(0,1) cumulative distribution function.

Table 1. Prob($\bar{\mathbf{X}} \le \mathbf{0}$)

	n		
COV	3	5	7
0.50	0.27E-03	0.39E-05	0.61E-07
0.40	0.75E-05	0.11E-07	0.19E-10
0.30	0.39E-08	0.45E-13	0.58E-18

Our second concern derives from the fact that if

$$1 - \alpha/2 \ge F_T(n-1)(\bar{X}/(S/\sqrt{n})) \tag{6}$$

where $F_T(n-1)$ denotes the cumulative distribution function of a central T distribution with $n-1$ degrees of freedom, then Equation (3) would have no positive solution. Regardless of how large COV became, we would still have

$$1 - \alpha/2 > F_{NCT}(n-1, \sqrt{n}/COV)(\bar{X}/(S/\sqrt{n}))$$

In this case, the upper bound of the COV confidence interval would be ∞.

We consider this to be a relatively mild problem for three reasons. First, as noted by Lehmann (1986, page 352), the upper bound given by solving Equation (3) for β is uniformly most accurate invariant under scale transformations. (For the non-statistician, this simply means that the approach is theoretically optimal.) Second, since a standard confidence interval for μ in this case has lower bound

$$\bar{X} - (S/\sqrt{n})F_T(n-1)^{-1}(1-\alpha/2)$$

inequality (6) implies that the confidence interval for μ includes 0. Thus, it is reasonable for a confidence interval for σ/μ to include ∞. Third, as we shall see in the next section, this problem is even more pronounced for the approximate intervals.

4 PROBLEMS WITH APPROXIMATE CONFIDENCE INTERVALS

Vangel (1996) discusses four approximate confidence intervals for a COV. He focuses on two of them given by (for a $1-\alpha$ confidence interval)

$$\Lambda_1 = \left\{ K\left[\left(\frac{u_1}{n} - 1\right)K^2 + \frac{u_1}{n-1}\right]^{-1/2},\ K\left[\left(\frac{u_2}{n} - 1\right)K^2 + \frac{u_2}{n-1}\right]^{-1/2} \right\}$$

and

$$\Lambda_4 = \left\{ K\left[\left(\frac{u_1+2}{n} - 1\right)K^2 + \frac{u_1}{n-1}\right]^{-1/2},\ K\left[\left(\frac{u_2+2}{n} - 1\right)K^2 + \frac{u_2}{n-1}\right]^{-1/2} \right\}$$

where $K = S/\bar{X}$, $u_1 = F_{\chi^2}(n-1)^{-1}(1-\alpha/2)$, $u_2 = F_{\chi^2}(n-1)^{-1}(\alpha/2)$, and $F_{\chi^2}(n-1)$ denotes the cumulative distribution function of a central chi-squared distribution with $n-1$ degrees of freedom. Λ_1 is based on McKay's (1932) approximation to the distribution of the COV, and Λ_4 is based on Vangel's modification to this approximation.

Now we saw in section 2 that the exact confidence interval will have an infinite upper bound when

$$S/\bar{X} \ge \sqrt{n}/F_T(n-1)^{-1}(1-\alpha/2) \equiv B_e$$

Some straightforward algebra demonstrates that the upper Λ_1 bound will have to be set to ∞ if

$$S/\bar{X} \geq \sqrt{n}\sqrt{\frac{u_2}{(n-1)(n-u_2)}} \equiv B_1$$

and the upper Λ_4 bound will have to be set to ∞ if

$$S/\bar{X} \geq \sqrt{n}\sqrt{\frac{u_2}{(n-1)(n-u_2-2)}} \equiv B_4$$

Clearly, $B_1 < B_4$. Also, using the asymptotic normality of a Chi-squared statistic, one can show that for larger n, $B_4 < B_e$. For smaller n we simply calculate the Bs and find that $B_1 < B_4 < B_e$. Thus whenever the upper confidence bound for the exact interval is infinite, the upper bounds for Λ_1 and Λ_4 are also infinite. Further there will be occasions in which the exact upper bound is finite but the approximate upper bounds are infinite. The probability of this occurring for the Λ_4 interval is

$$P_4 \equiv \text{Prob}(B_4 \leq S/\bar{X} < B_e) = \text{Prob}(\sqrt{n}/B_4 \geq \bar{X}/(S/\sqrt{n}\,) > \sqrt{n}/B_e)$$

For the Λ_1 interval this probability is

$$P_1 \equiv \text{Prob}(B_1 \leq S/\bar{X} < B_e) = \text{Prob}(\sqrt{n}/B_1 \geq \bar{X}/(S/\sqrt{n}\,) > \sqrt{n}/B_e)$$

Since $\bar{X}/(S/\sqrt{n}\,)$ has a noncentral T distribution with $n-1$ degrees of freedom and noncentrality parameter $\sqrt{n}/\text{COV}$, we can calculate these probabilities with the aid of a noncentral T distribution routine. We present values of these probabilities in Table 2 for $n = 3, 5, 7, 9, 11$, $\text{COV} = 0.5, 0.4, 0.3, 0.2$, and $\alpha = 0.10, 0.05, 0.01$. Clearly the problem of infinite upper bounds is more severe for the approximate intervals than for the exact interval.

Table 2. Probability that the exact confidence interval upper bound is finite and the approximate confidence interval upper bound is infinite

COV	n	90% CI P_4	90% CI P_1	95% CI P_4	95% CI P_1	99% CI P_4	99% CI P_1
0.50	3	0.224	0.513	0.190	0.367	0.058	0.100
0.50	5	0.214	0.487	0.352	0.603	0.359	0.465
0.50	7	0.087	0.253	0.212	0.431	0.497	0.669
0.50	9	0.028	0.103	0.092	0.231	0.387	0.581
0.50	11	0.008	0.036	0.034	0.104	0.233	0.402
0.40	3	0.211	0.562	0.225	0.466	0.084	0.146
0.40	5	0.083	0.297	0.217	0.487	0.426	0.598
0.40	7	0.013	0.074	0.058	0.193	0.339	0.561
0.40	9	0.001	0.013	0.010	0.051	0.141	0.303
0.40	11	0.000	0.002	0.001	0.010	0.043	0.121
0.30	3	0.133	0.518	0.232	0.561	0.131	0.234
0.30	5	0.007	0.080	0.050	0.224	0.347	0.605
0.30	7	0.000	0.003	0.002	0.023	0.082	0.238
0.30	9	0.000	0.000	0.000	0.001	0.008	0.042
0.30	11	0.000	0.000	0.000	0.000	0.000	0.004
0.20	3	0.019	0.272	0.121	0.502	0.214	0.411
0.20	5	0.000	0.001	0.000	0.015	0.065	0.257
0.20	7	0.000	0.000	0.000	0.000	0.000	0.008
0.20	9	0.000	0.000	0.000	0.000	0.000	0.000
0.20	11	0.000	0.000	0.000	0.000	0.000	0.000

A second problem with the approximate confidence intervals relative to the exact intervals concerns their lengths and configurations. Simulations show that even for COVs as large as 0.5 and n as small as 3, the coverages of the approximate intervals are satisfactory. (See Tables 7 to 9 in Appendix II.) However, for larger COVs and smaller n's, the lengths and configurations of the approximate intervals are inferior to those of the exact intervals. In particular, we performed simulations (4,000 trials per COV, sample size combination) that indicate that even if we restrict our attention to situations in which both the approximate and exact intervals have finite upper bounds, then four reasonable measures of confidence interval (CI) performance are superior for the exact intervals.

The simulation estimates of the expected value of

$$\text{(exact CI interval length)/(approximate CI interval length)}$$

are presented in Table 3 for $n = 3, 5, 7, 9, 11$, $\text{COV} = 0.5, 0.4, 0.3, 0.2$, and $\alpha = 0.10, 0.05, 0.01$.

The simulation estimates of the expected value of

$$\text{(max. distance of an exact CI point from the true COV)/}$$

$$\text{(max. distance of an approximate CI point from the true COV)}$$

are presented in Table 4.

The simulation estimates of the probability that the exact confidence interval is shorter than an approximate interval are presented in Table 5.

The simulation estimates of the probability that the maximum distance between the true COV and a point in the confidence interval is shorter for the exact interval than for an approximate interval are presented in Table 6.

The program that was used to perform the simulations is available on the Web at

`http://www1.fpl.fs.fed.us/cov.sim.html`

Table 3. Ratios of the exact confidence interval length to the approximate confidence interval lengths[1]

Confidence Level	COV	n = 3 Λ_4	3 Λ_1	5 Λ_4	5 Λ_1	7 Λ_4	7 Λ_1	9 Λ_4	9 Λ_1	11 Λ_4	11 Λ_1
0.90	0.500	0.910	0.720	0.856	0.685	0.875	0.717	0.902	0.761	0.920	0.805
	0.400	0.923	0.719	0.910	0.735	0.940	0.799	0.960	0.856	0.969	0.894
	0.300	0.943	0.730	0.963	0.801	0.982	0.893	0.988	0.934	0.991	0.952
	0.200	0.980	0.780	0.995	0.921	0.997	0.963	0.998	0.976	0.998	0.982
0.95	0.500	0.901	0.713	0.820	0.666	0.828	0.680	0.852	0.724	0.880	0.755
	0.400	0.898	0.725	0.864	0.699	0.899	0.750	0.929	0.809	0.950	0.858
	0.300	0.917	0.713	0.921	0.757	0.963	0.849	0.980	0.912	0.986	0.940
	0.200	0.944	0.744	0.985	0.880	0.994	0.950	0.996	0.970	0.997	0.979
0.99	0.500	0.876	0.730	0.768	0.645	0.751	0.635	0.758	0.657	0.771	0.666
	0.400	0.883	0.733	0.807	0.679	0.799	0.671	0.832	0.712	0.870	0.763
	0.300	0.896	0.722	0.826	0.683	0.878	0.740	0.934	0.827	0.961	0.890
	0.200	0.884	0.714	0.916	0.759	0.976	0.888	0.989	0.947	0.993	0.967

Table 4. Ratios of the maximum distance between the true COV and a point in the exact confidence interval and the maximum distances for the approximate confidence intervals[1]

Confidence Level	COV	n = 3 Λ_4	3 Λ_1	5 Λ_4	5 Λ_1	7 Λ_4	7 Λ_1	9 Λ_4	9 Λ_1	11 Λ_4	11 Λ_1
0.90	0.500	0.902	0.725	0.841	0.655	0.864	0.687	0.894	0.740	0.914	0.787
	0.400	0.918	0.695	0.903	0.705	0.936	0.778	0.957	0.842	0.967	0.883
	0.300	0.941	0.697	0.961	0.782	0.981	0.881	0.987	0.927	0.990	0.946
	0.200	0.980	0.757	0.994	0.913	0.997	0.958	0.998	0.973	0.998	0.980
0.95	0.500	0.890	0.722	0.797	0.629	0.807	0.639	0.836	0.690	0.867	0.726
	0.400	0.890	0.701	0.851	0.661	0.889	0.719	0.922	0.786	0.945	0.839
	0.300	0.913	0.676	0.916	0.729	0.960	0.832	0.977	0.900	0.984	0.930
	0.200	0.942	0.715	0.985	0.867	0.994	0.942	0.996	0.965	0.997	0.975
0.99	0.500	0.861	0.743	0.734	0.600	0.716	0.581	0.725	0.603	0.741	0.616
	0.400	0.873	0.709	0.783	0.625	0.776	0.623	0.813	0.672	0.855	0.729
	0.300	0.889	0.679	0.813	0.642	0.867	0.709	0.927	0.804	0.956	0.872
	0.200	0.880	0.678	0.912	0.736	0.974	0.875	0.988	0.937	0.992	0.960

[1]4,000 trials per COV, n combination

Table 5. Empirical probability that the exact confidence interval is shorter than the approximate interval given that both intervals are finite[2]

		n									
Confidence		3		5		7		9		11	
Level	COV	Λ_4	Λ_1	Λ_4	Λ_1	Λ_4	Λ_1	Λ_4	Λ_1	Λ_4	Λ_1
	0.50	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.90	0.40	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.30	0.58	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.20	0.37	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.50	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.95	0.40	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.30	0.74	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.20	0.64	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.50	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	0.40	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.30	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.20	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 6. Empirical probability that the maximum distance between the true COV and a point in the confidence interval is shorter for the exact than for the approximate interval given that both intervals are finite[2]

		n									
Confidence		3		5		7		9		11	
Level	COV	Λ_4	Λ_1	Λ_4	Λ_1	Λ_4	Λ_1	Λ_4	Λ_1	Λ_4	Λ_1
	0.50	0.86	0.52	1.00	0.72	1.00	0.77	1.00	0.76	1.00	0.77
0.90	0.40	0.77	0.62	1.00	0.77	1.00	0.78	1.00	0.76	1.00	0.75
	0.30	0.62	0.73	0.99	0.81	1.00	0.78	1.00	0.74	1.00	0.72
	0.20	0.37	0.81	0.95	0.80	0.99	0.75	1.00	0.72	1.00	0.69
	0.50	0.79	0.52	1.00	0.74	1.00	0.81	1.00	0.81	1.00	0.83
0.95	0.40	0.77	0.60	0.99	0.79	1.00	0.83	1.00	0.83	1.00	0.82
	0.30	0.74	0.73	0.98	0.86	1.00	0.84	1.00	0.81	1.00	0.79
	0.20	0.64	0.84	0.92	0.87	0.97	0.82	0.99	0.79	1.00	0.77
	0.50	0.76	0.50	0.95	0.74	1.00	0.84	1.00	0.88	1.00	0.91
0.99	0.40	0.81	0.60	0.92	0.82	1.00	0.90	1.00	0.91	1.00	0.91
	0.30	0.86	0.74	0.94	0.91	0.98	0.93	1.00	0.93	1.00	0.91
	0.20	0.87	0.87	0.96	0.95	0.93	0.93	0.94	0.91	0.98	0.89

[2]4,000 trials per COV, n combination

It is clear from these tables that for larger COVs and smaller n's, the exact intervals outperform the approximate intervals on all four of these measures. This holds true even for COVs below McKay's 0.33 cutoff. The disparity in performance is accentuated for the 0.99 level confidence intervals.

High COVs are seldom seen in scientific measurements or in the properties of man-made materials, but they are not uncommon in the properties of naturally occurring materials. For example, the COVs of various wood strength properties can exceed 0.30 (see Table 4-6 of *The Wood Handbook* (1999)).

5 COMPUTATIONAL RESOURCES

A Web-based program that makes use of Equations (2) and (3) to obtain exact confidence intervals on COVs for normal distributions can be run at

`http://www1.fpl.fs.fed.us/covnorm.html`

The Web form front end for for this program is displayed in Figure 1. To use the program, type in the desired confidence level, the sample size, and the sample mean and standard deviation of the data. For example, suppose the data set consists of the five data points 9.68, 9.94, 10.82, 11.09, and 10.05 (generated from a $N(10, 1^2)$ distribution), and you want a 95% confidence interval on the corresponding COV. Enter the confidence level (95) in the first (top) box, the sample size (5) in the second box, the sample mean (10.32) in the third box, and the sample standard deviation (0.606) in the last box. (See Figure 1.) Then click the **Execute the program** button, and the program will return a Web page that presents the calculated COV (.059) and the corresponding confidence interval ([.035,.170]).

Note that the sample standard deviation should be calculated with a $n-1$ denominator rather than a n denominator where n is the sample size. Thus the appropriate formula for the sample standard deviation is

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1)}$$

where $\bar{x}$ is the sample mean. Most calculators and spreadsheets will have an option to calculate s.

A Web-based program that obtains exact confidence intervals on COVs for lognormal distributions can be run at `http://www1.fpl.fs.fed.us/covln.html`. To use this program, type in the desired confidence level, the sample size, and the standard deviation of the natural logs of the data values. For example, suppose the data set consists of the five data points 8.52, 8.43, 8.24, 9.00, and 10.72 (generated from a lognormal$(2.3, .1^2)$ distribution), and you want a 95% confidence interval on the corresponding COV. Type 95 in the first box, 5 in the second, and .106, the sample standard deviation of the natural logs of the data, in the third. (See Figure 2.) Then click the **Execute the program** button, and the program will return a Web page that presents the corresponding confidence interval ([.064,.312]).

As in the normal case, the sample standard deviation (of the natural logs of the data) should be calculated with a $n-1$ denominator rather than a n denominator where n is the sample size. Most calculators and spreadsheets will have an option to calculate this value (after you have calculated the natural logs of the data values).

The public domain source code for complete FORTRAN and Java implementations of a program to calculate exact COV confidence intervals for normal distributions can be found at

`http://www1.fpl.fs.fed.us/covnorm.code.html`

If users want to do their own programming, public domain noncentral T distribution functions and nonlinear equation solvers are readily available.

Public domain FORTRAN or C code to calculate the noncentral T distribution can be found in the DCDFLIB library. DCDFLIB is a public domain library of "routines for cumulative distribution functions, their inverses, and their parameters." It was produced by Barry Brown, James Lovato, and Kathy Russell of the Department of Biomathematics, M.D. Anderson Cancer Center, The University of Texas. DCDFLIB can be found at

`http://odin.mdacc.tmc.edu/biomath/anonftp/page_2.html`

Public domain Java code to calculate the noncentral T distribution can be found at

`http://www1.fpl.fs.fed.us/distributions.html`

Public domain FORTRAN and C code to solve a nonlinear equation can be found at

`http://gams.nist.gov/serv.cgi/Class/F1b/`

Public domain Java code to solve a nonlinear equation can be found at

`http://www1.fpl.fs.fed.us/optimization.html`

6 CONCLUDING REMARKS

For normally distributed data, large coefficients of variation (COVs), and small sample sizes, "exact" confidence intervals yield, on average, smaller lengths and smaller maximum distances from the true COV than do Vangel's (1996) approximate intervals. In addition, the upper bounds of the approximate intervals are infinite more frequently than are the upper bounds of the exact intervals.

We have developed Web resources that make it simple to obtain exact confidence intervals for COVs for both normally and lognormally distributed data.

REFERENCES

Johnson, N.L. and Welch, B.L. (1940), "Applications of the Non-central t-Distribution," *Biometrika*, 31, 362–389.

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

McKay, A.T. (1932), "Distribution of the Coefficient of Variation and the Extended 't' Distribution," *Journal of the Royal Statistical Society*, 95, 695–698.

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

The Wood Handbook (1999), General Technical Report FPL-GTR-113, USDA Forest Service Forest Products Laboratory, Madison, WI.

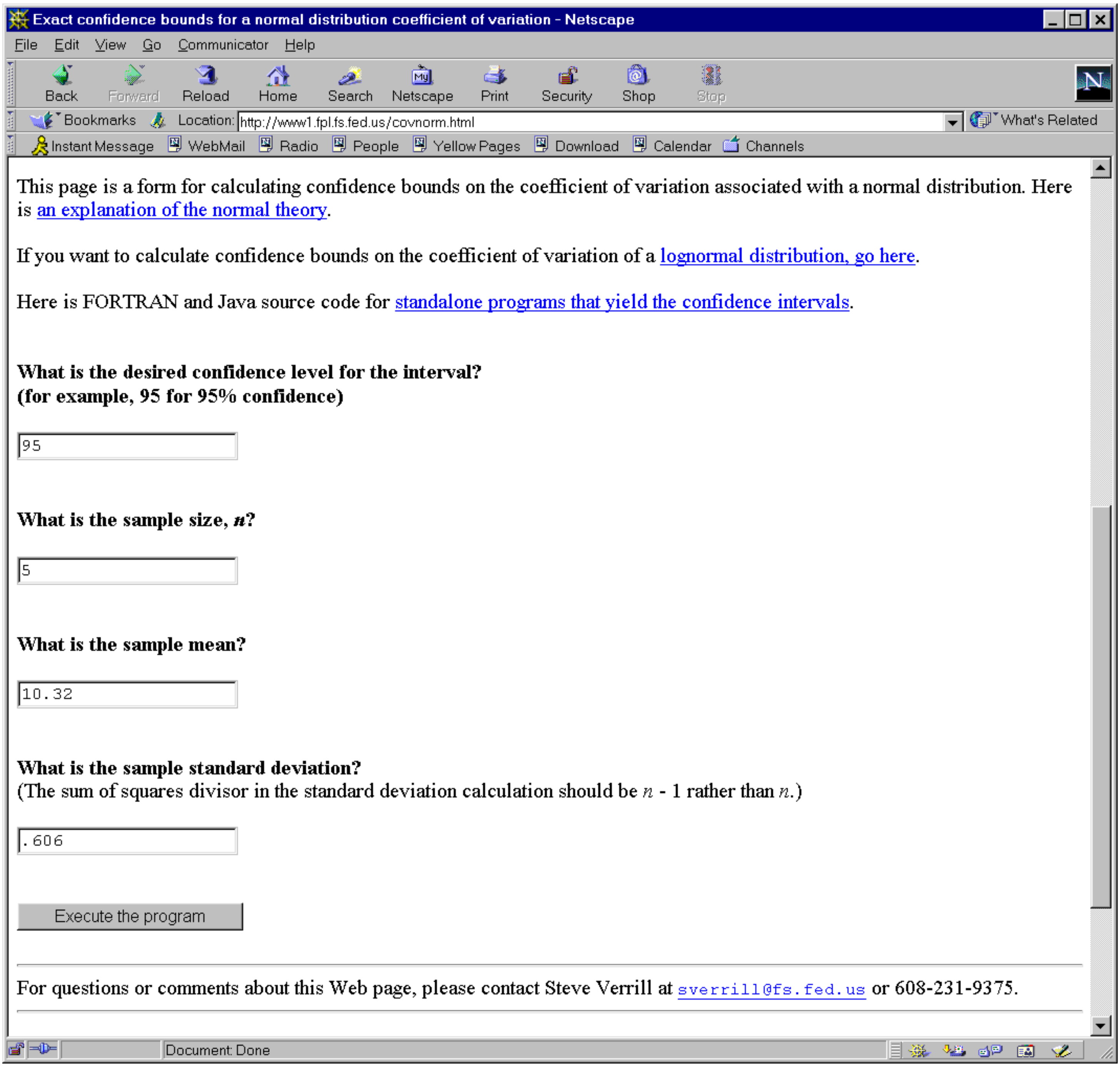


Figure 1: The Web form for a confidence interval on the COV for normally distributed data

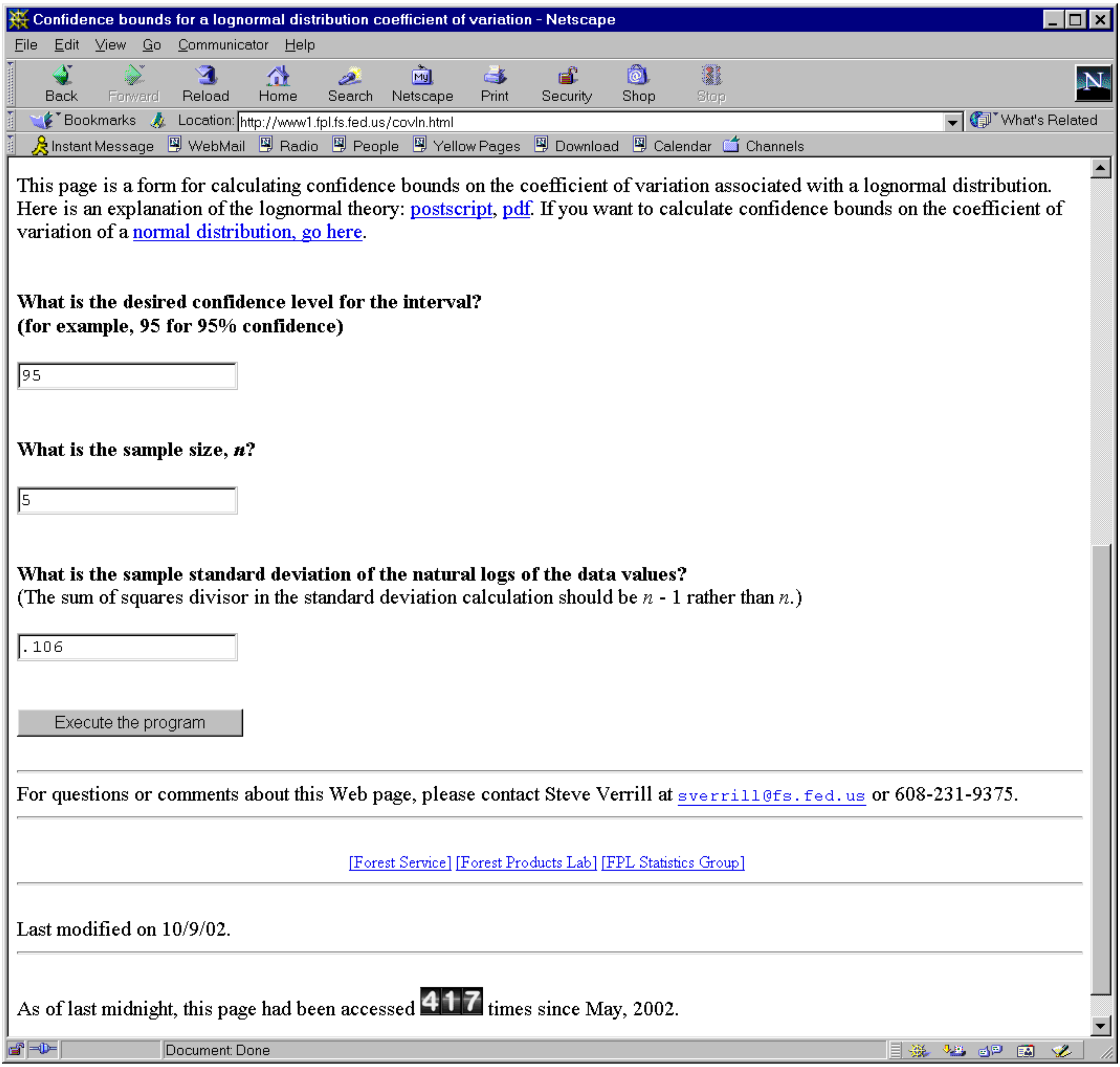


Figure 2: The Web form for a confidence interval on the COV for lognormally distributed data

APPENDIX I — CONFIDENCE INTERVAL FOR THE COEFFICIENT OF VARIATION OF A LOGNORMAL DISTRIBUTION

Assume that X has a lognormal(μ, σ^2) distribution. That is

$$\ln(X) \sim \mathrm{N}(\mu, \sigma^2)$$

We want a confidence interval on

$$\mathrm{COV} = \sqrt{\mathrm{Var}(X)}/\mathrm{E}(X)$$

where $\mathrm{Var}(X)$ denotes the population variance of X and $\mathrm{E}(X)$ denotes the population expectation of X. Straightfoward calculations yield

$$\mathrm{E}(X) = \exp(\mu + \sigma^2/2)$$

and

$$\mathrm{Var}(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

Thus

$$\begin{aligned}\mathrm{COV} &= \left(\exp(\mu + \sigma^2/2)\sqrt{\exp(\sigma^2) - 1}\right) / \exp(\mu + \sigma^2/2)\\ &= \sqrt{\exp(\sigma^2) - 1}\end{aligned}$$

Now let $X_1, \ldots, X_n$ be a sample from the lognormal distribution. Then $Y_1 = \ln(X_1), \ldots, Y_n = \ln(X_n)$ is a sample from a $\mathrm{N}(\mu, \sigma^2)$ distribution and we know that a $1 - \alpha$ level confidence interval for σ^2 is $[a_L, a_U]$, where

$$\begin{aligned}a_L &\equiv (n-1)S_n^2/(F_{\chi^2}(n-1)^{-1}(1-\alpha/2))\\ a_U &\equiv (n-1)S_n^2/(F_{\chi^2}(n-1)^{-1}(\alpha/2))\\ S_n^2 &\equiv \sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)\end{aligned}$$

and $F_{\chi^2}(n-1)(x)$ denotes the cumulative distribution function of a central chi-squared distribution with $n-1$ degrees of freedom. Thus

$$\left[\sqrt{\exp(a_L) - 1}, \sqrt{\exp(a_U) - 1}\right]$$

is a $1 - \alpha$ level confidence interval for COV.

APPENDIX II — SIMULATION COVERAGES OF CONFIDENCE INTERVALS

Table 7. Simulation coverages of nominal 90% confidence intervals[3]

	n											
	3			5			7			9		
COV	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1
0.50	0.903	0.906	0.890	0.901	0.907	0.896	0.900	0.906	0.900	0.902	0.909	0.900
0.40	0.899	0.901	0.893	0.901	0.904	0.897	0.897	0.901	0.897	0.904	0.907	0.903
0.30	0.899	0.899	0.896	0.907	0.907	0.903	0.899	0.900	0.897	0.904	0.905	0.901
0.20	0.903	0.903	0.901	0.904	0.904	0.903	0.894	0.895	0.895	0.897	0.897	0.896

Table 8. Simulation coverages of nominal 95% confidence intervals[4]

	n											
	3			5			7			9		
COV	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1
0.50	0.950	0.952	0.942	0.949	0.956	0.944	0.950	0.955	0.949	0.948	0.952	0.949
0.40	0.947	0.950	0.943	0.951	0.951	0.950	0.950	0.952	0.950	0.950	0.951	0.950
0.30	0.947	0.948	0.946	0.949	0.949	0.949	0.950	0.950	0.948	0.952	0.952	0.950
0.20	0.956	0.956	0.955	0.953	0.954	0.953	0.950	0.950	0.948	0.950	0.950	0.948

Table 9. Simulation coverages of nominal 99% confidence intervals[5]

	n											
	3			5			7			9		
COV	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1	exact	Λ_4	Λ_1
0.50	0.990	0.992	0.989	0.992	0.994	0.991	0.990	0.992	0.992	0.988	0.989	0.988
0.40	0.990	0.991	0.988	0.991	0.992	0.990	0.990	0.993	0.990	0.992	0.992	0.991
0.30	0.990	0.990	0.988	0.990	0.990	0.989	0.990	0.990	0.989	0.991	0.991	0.990
0.20	0.993	0.993	0.992	0.989	0.989	0.989	0.990	0.990	0.991	0.988	0.988	0.988

[3] 4000 trials per COV, n combination. The half-width of an approximate 90% confidence interval on the coverage is $1.96 \times \sqrt{.9 \times .1/4000} = .0093$.

[4] 4000 trials per COV, n combination. The half-width of an approximate 95% confidence interval on the coverage is $1.96 \times \sqrt{.95 \times .05/4000} = .0068$.

[5] 4000 trials per COV, n combination. The half-width of an approximate 99% confidence interval on the coverage is $1.96 \times \sqrt{.99 \times .01/4000} = .0031$.