Uncertainty calculation for the ratio of dependent measurements

2003 Metrologia 40 177

(http://iopscience.iop.org/0026-1394/40/4/306)

INSTITUTE OF PHYSICS PUBLISHING METROLOGIA

Metrologia **40** (2003) 177–183 PII: S0026-1394(03)62330-3

Uncertainty calculation for the ratio of dependent measurements

Jan Hannig[1], **C M Wang**[2] **and Hari K Iyer**[1]

[1] Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA
[2] Statistical Engineering Division, National Institute of Standards and Technology, Boulder, CO 80305, USA

E-mail: hannig@stat.colostate.edu, hari@stat.colostate.edu and jwang@boulder.nist.gov

Received 2 April 2003
Published 13 June 2003
Online at stacks.iop.org/Met/40/177

Abstract
A measurand θ of interest is the ratio of two other quantities, μ_p and μ_q. A measurement experiment is conducted and results P and Q are obtained as estimates of μ_p and μ_q. The ratio $Y = P/Q$ is generally reported as the result for the measurand θ. In this paper we consider the problem of computing an uncertainty interval for θ having a prescribed confidence level of $1 - \alpha$. Although an exact procedure, based on an approach due to Fieller, is available for this problem, it is well known that this procedure can lead to unbounded confidence regions in certain situations. As a result, practitioners often use various non-exact methods. One such non-exact method is based on the propagation-of-errors approach described in the ISO *Guide to the Expression of Uncertainty in Measurement* to calculate a standard uncertainty u_y for Y. A confidence interval for θ with a presumed confidence level of 95% is obtained as $[Y - 2u_y, Y + 2u_y]$. In this paper we develop a highly accurate approximation for the coverage probability associated with the interval $[Y - ku_y, Y + ku_y]$. In particular, we demonstrate that, using $n - 1$ degrees of freedom for u_y, and the corresponding Student's t coverage factor $k = t_{1-\alpha/2:n-1}$ rather than $k = 2$, leads to uncertainty intervals $[Y - t_{1-\alpha/2:n-1}u_y,\ Y + t_{1-\alpha/2:n-1}u_y]$, that are nearly identical to Fieller's exact intervals whenever the measurement relative uncertainties are small, as is the case in most metrological applications. In addition, they are easy to compute and may be recommended for routine use in metrological applications. Improved coverage factors k can be derived based on the results of this paper for those exceptional situations where the t-interval may not have coverage probability sufficiently close to the desired value.

1. Introduction

There are many instances where a physical quantity of interest is defined as a ratio of two other quantities μ_p and μ_q. A measurement experiment provides values P and Q as the measured results for μ_p and μ_q, and the measurand of interest is estimated by $Y = P/Q$. Furthermore, it is quite common to use a method of propagation of errors based on a first-order Taylor-series expansion for the calculation of the standard uncertainty associated with Y.

Calculating confidence intervals for ratios, particularly when the numerator and denominator estimates are distributed as Gaussian random variables, is a problem that has attracted the attention of numerous researchers [1–7]. The method of propagation of errors proposed in the ISO *Guide to the Expression of Uncertainty in Measurement* (ISO GUM) [8], provides formulae for computing an approximate standard uncertainty for a ratio, but specifies neither how an associated number of degrees of freedom is to be computed nor how an uncertainty interval is to be obtained for ratios. Metrologists routinely use a coverage factor of 2 to multiply the standard uncertainty and obtain an expanded uncertainty, for calculating a confidence interval for the ratio quantity. This confidence interval is presumed to have a level of confidence equal to 95%.

In the context of estimation of a ratio, the following two distinct scenarios must be recognized.

(a) Independent measurements: independent measurement experiments are conducted for measuring μ_p and μ_q, respectively. There are no common error components in the measured results P and Q. Consequently, P and Q may be assumed to be statistically independent.
(b) Dependent measurements: μ_p and μ_q are measured in the course of a single measurement experiment. It is expected that common sources of error exist that contribute to errors in P and Q. Consequently, there is statistical dependence between the results P and Q.

It turns out that an exact confidence interval procedure has been discussed in the statistical literature for the dependent case, based on what is now referred to as Fieller's method [1]. In this paper we compare, using a statistical simulation study, the performance of the exact Fieller uncertainty interval for θ with the uncertainty interval calculated using the standard uncertainty u_y based on propagation of errors and a coverage factor $k = 2$. The latter approach is referred to as POE_2. In the process, we make the observation that use of a coverage factor $k = t_{1-\alpha/2:n-1}$ rather than $k = 2$, with u_y, improves the performance of the POE_2 interval. This modified interval will be referred to as a POE_t interval with $n - 1$ degrees of freedom.

This paper is organized as follows. In section 2, we define the statistical model used to describe the results from the measurement experiment. We give detailed instructions for calculating uncertainty intervals using each of the three methods—POE_2, POE_t, and Fieller's interval. We also develop approximate bounds for the coverage probability associated with POE-type intervals with any specified coverage factor k. In section 3 we explain how we conducted our statistical simulation study and the different parameters that were varied in the study. Furthermore, we discuss the results and make recommendations. The final section summarizes our results.

2. The statistical model

Let θ be a measurand of interest. Suppose

$$\theta = \frac{\mu_p}{\mu_q},$$

where μ_p and μ_q are quantities that are measured directly with P and Q denoting the corresponding results. We write

$$P - \mu_p = \epsilon \qquad \text{and} \qquad Q - \mu_q = \delta,$$

where ϵ and δ are measurement errors in P and Q, respectively. Suppose the ensemble of possible errors ϵ and δ are described by a bivariate probability density function. The measurand θ is generally estimated by Y, where

$$Y = \frac{P}{Q}.$$

For this result to be useful one also needs to quantify the associated uncertainty. This is typically accomplished by providing a quantity called the standard uncertainty for Y, denoted by u_y. According to the ISO GUM, u_y may be calculated using the method of propagation of errors.

The standard uncertainty is often used to compute a confidence interval $[L, U]$ for the measurand, in this case θ, by using the formula

$$L = Y - k\,u_y \qquad \text{and} \qquad U = Y + k\,u_y,$$

where L is the lower limit and U is the upper limit of the confidence interval. The quantity k is called a coverage factor. The value of k depends on the confidence level to be associated with the interval. Sometimes a coverage factor of $k = 2$ or $k = 3$ is used with implied confidence levels of 95% and 99%, respectively. However, the actual confidence level will generally differ from the nominal value, but the difference is often, but not always, small enough to be of no consequence in most applications.

2.1. Propagation of errors

Suppose the errors ϵ and δ have zero means and standard deviations equal to $\sigma_p/\sqrt{n}$ and $\sigma_q/\sqrt{n}$, respectively, where n is the number of independent repeat observations (P_i, Q_i) whose means are the reported results P and Q, respectively. Suppose also that the correlation between P and Q is ρ. Here σ_p and σ_q are the standard deviations associated with a single measurement of μ_p and μ_q. In certain circumstances, it may be known that ρ is zero, particularly when there are no common sources of error during the measurements of μ_p and μ_q. However, in this paper, we do not impose any restrictions on ρ.

Let us denote the usual sample estimate of σ_p and σ_q by S_p and S_q, respectively. The estimate of ρ is given by $\hat{\rho}$, where

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(P_i - P)(Q_i - Q)/(n-1)}{S_p S_q}.$$

There are $\nu = n - 1$ degrees of freedom associated with S_p, S_q, and $\hat{\rho}$. The propagation-of-errors method leads to the following formula for the standard uncertainty of Y:

$$u_y = \frac{P}{\sqrt{n}Q}\left[\frac{S_p^2}{P^2} + \frac{S_q^2}{Q^2} - \frac{2\hat{\rho}S_pS_q}{PQ}\right]^{1/2}. \qquad (1)$$

In the appendix we show that, when $(P, Q)^t$ has a bivariate Gaussian distribution, it is reasonable to associate $\nu = n - 1$ degrees of freedom with u_y. The ISO GUM does not give explicit procedures for calculating a number of degrees of freedom for u_y. This is because the commonly used Welch–Satterthwaite formula applies only when the expression for u_y is a linear function of a number of sample variance terms (with no correlation term being present).

The coverage probability associated with the interval $[Y - ku_y, Y + ku_y]$ depends on the sample size n, the correlation coefficient ρ, and the coefficients of variation of P and Q, which are denoted by $\kappa_p = \sigma_p/\mu_p$ and $\kappa_q = \sigma_q/\mu_q$. We use $p(n, \kappa_p, \kappa_q, \rho)$, or simply p, to denote this coverage probability. It can be evaluated by the use of either a numerical double integration or a Monte Carlo approach. Alternatively, a large sample approximation for p, derived using asymptotic methods, can be used. See theorem 1 in the appendix. The following simple bounds are a direct consequence of this theorem.

The coverage probability p of the POE confidence interval using k as the coverage factor, lies approximately between p_L and p_U, where

$$p_L = \Pr(|T_{n-1}| \leqslant k) + 4\,l(n,k)c_1(n,k)\,k^3\kappa_q^2, \qquad (2)$$
$$p_U = \Pr(|T_{n-1}| \leqslant k) + 4\,l(n,k)c_2(n,k)\,k^3\kappa_q^2$$

and

$$l(n,k) = \frac{1}{8\sqrt{\pi}}\frac{\Gamma(n/2)}{\Gamma((n+1)/2)}\frac{(n-1)^{(n+1)/2}}{(n-1+k^2)^{(n+5)/2}}. \qquad (3)$$

T_{n-1} is a student's t random variable with $n-1$ degrees of freedom, $c_1(n,k) = \min\{4n-4-k^2(n-2), n-1+k^2\}$, and $c_2(n,k) = \max\{4n-4-k^2(n-2), n-1+k^2\}$. In particular, when $k^2 > 3$ (as is often the case in practical applications), we have $c_1(n,k) = 4n-4-k^2(n-2)$ and $c_2(n,k) = n-1+k^2$. The bounds (2) work well for all values κ_p, κ_q, and ρ. The values of $4l(n,k)c_1(n,k)\kappa_q^2$ and $4l(n,k)c_2(n,k)\kappa_q^2$ are generally very small. Thus, a coverage factor k obtained from a t-distribution with $n-1$ degrees of freedom will provide an interval with coverage probability very nearly equal to the nominal value in most practical applications. Tables 1 and 2 display p_L and p_U for the 95% POE$_2$ and POE$_t$ intervals for some values of n and for $\kappa_q = 10\%$ and $\kappa_q = 50\%$, respectively.

The bounds in (2) allow us to obtain coverage factors k such that the resulting interval will have actual coverage equal to or greater than the nominal value. These coverage factors are obtained by solving for values of k for which the lower bound equals the nominal confidence level. However, tables 1 and 2 indicate that the bounds for the coverage probability of the POE$_t$ intervals are very tight and are close to the nominal value, hence there is no need to use these bounds to improve the coverage factor $k = t_{1-\alpha/2:n-1}$ for this problem.

Example. We use an example from the ISO GUM annex H.2 to illustrate the calculation of POE$_2$ and POE$_t$ and compare the results. In this example, the amplitude V (in V) of a sinusoidally alternating potential difference across the terminals of a circuit element, the amplitude I (in mA) of the alternating current passing through it, and the phase-shift angle ϕ (in rad) of the alternating potential difference relative to the alternating current are the primary quantities that are measured. The measurands of interest are the three impedance components, which are functions of the true values of V, I, and ϕ. For illustrative purposes, we consider only one of the measurands, which is given by $\mu_z = \mu_V/\mu_I$.

Five sets of simultaneous observations (V_i, I_i, ϕ_i), $i = 1, \ldots, 5$, are available. The summary statistics associated with (V, I) are shown below:

$$\bar{V} = 4.999, \qquad S_V = 0.007\,176\,4,$$
$$\bar{I} = 19.661, \qquad S_I = 0.021\,177\,8,$$
$$\hat{\rho}(V, I) = -0.3553.$$

The measurand μ_z is estimated by $Z = \bar{V}/\bar{I} = 254.2597\,\Omega$. Based on (1), we obtain the standard uncertainty of Z, $u_z = 0.2363\,\Omega$. The 95% POE$_t$ confidence interval on μ_z is found to be $(253.6035, 254.9159)\,\Omega$, while the 95% POE$_2$ interval is $(253.787, 254.7324)\,\Omega$. The actual coverage probability of the POE$_2$ interval, based on simulation results reported in section 3, is about 0.88. Therefore, it is not surprising that the POE$_2$ interval is shorter. Also, the bounds for p based on the asymptotic theory, shown in tables 1 and 2, are in excellent agreement with the simulation results.

Since an exact interval procedure is available for the ratio quantity being discussed here, it is of interest to compare the exact intervals with the intervals obtained using POE$_2$ and POE$_t$. We begin with a description of Fieller's exact method.

Table 1. Lower and upper bounds for the coverage probability for the POE$_2$ and POE$_t$ intervals with $\kappa_q = 10\%$.

	POE$_2$		POE$_t$	
n	p_L	p_U	p_L	p_U
3	0.816 70	0.816 87	0.949 95	0.950 09
5	0.884 00	0.884 12	0.949 92	0.950 14
10	0.923 48	0.923 55	0.949 97	0.950 09
30	0.945 06	0.945 08	0.950 00	0.950 02

Table 2. Lower and upper bounds for the coverage probability for the POE$_2$ and POE$_t$ intervals with $\kappa_q = 50\%$.

	POE$_2$		POE$_t$	
n	p_L	p_U	p_L	p_U
3	0.822 67	0.825 76	0.948 82	0.952 31
5	0.886 81	0.889 74	0.947 92	0.953 42
10	0.924 26	0.926 08	0.949 22	0.952 23
30	0.945 13	0.945 66	0.959 98	0.950 59

2.2. Fieller's method

We outline the derivation of an exact confidence set for θ, originally derived by Fieller [3]. We begin by noting that the distribution of the quantity $W = P - \theta Q$ is normal with mean zero and variance given by

$$\sigma_w^2 = \frac{\sigma_p^2 - 2\theta\,\rho\sigma_p\sigma_q + \theta^2\sigma_q^2}{n}.$$

The ratio

$$(n-1)\frac{S_p^2 - 2\theta\hat{\rho}S_pS_q + \theta^2S_q^2}{\sigma_p^2 - 2\theta\,\rho\sigma_p\sigma_q + \theta^2\sigma_q^2}$$

has a χ^2_{n-1} distribution and is independent of W. Hence

$$\frac{n\,(P-\theta Q)^2}{S_p^2 - 2\theta\hat{\rho}S_pS_q + \theta^2S_q^2}$$

has an F-distribution with 1 and $n-1$ degrees of freedom. Therefore,

$$\Pr\left[\frac{n(P-\theta Q)^2}{S_p^2 - 2\theta\hat{\rho}S_pS_q + \theta^2S_q^2} \leqslant F_0\right] = 1-\alpha,$$

where $F_0 = F_{1-\alpha:1,n-1}$ is an F-table value such that the area under the F-density function with 1 and $n-1$ degrees of freedom between 0 and F_0 is $1-\alpha$. From this we deduce that

$$\Pr[(nQ^2 - F_0S_q^2)\theta^2 - 2(nPQ - F_0\hat{\rho}S_pS_q)\theta + (nP^2 - F_0S_p^2) \leqslant 0] = 1-\alpha.$$

The two roots of the quadratic equation (in θ)

$$(nQ^2 - F_0 S_q^2)\theta^2 - 2(nPQ - F_0\hat{\rho}S_pS_q)\theta + (nP^2 - F_0S_p^2) = 0$$

are, respectively,

$$L = \frac{(nPQ - F_0\hat{\rho}S_pS_q) - R_{pq}}{nQ^2 - F_0S_q^2} \quad (4)$$

and

$$U = \frac{(nPQ - F_0\hat{\rho}S_pS_q) + R_{pq}}{nQ^2 - F_0S_q^2}, \quad (5)$$

where

$$R_{pq}^2 = (nPQ - F_0\hat{\rho}S_pS_q)^2 - (nQ^2 - F_0S_q^2)(nP^2 - F_0S_p^2).$$

Provided that $nQ^2 - F_0S_q^2 > 0$, it can be shown that the roots L and U above are real and distinct and that $[L, U]$ is a $1-\alpha$ confidence interval for θ. The condition $nQ^2 - F_0S_q^2 > 0$ is equivalent to the statement that an α level test of the hypothesis H_0: $\mu_q = 0$ against H_a: $\mu_q \neq 0$ rejects the hypothesis H_0 in favour of H_a. This is generally the case in metrological applications, due to the fact that the variability in the data is likely to be very small. We will assume the condition holds. Observe that the interval is asymmetric relative to the reported value Y for θ.

If the coefficient of variation for Q is large, then the Fieller method may not result in a proper confidence interval. This seldom happens in metrological applications and hence we do not discuss this further. The interested reader may consult [3].

Example. We first obtain $F_0 = F_{0.95:1,4} = 7.708\,647$. Since $n\bar{I} - F_0S_I^2 = 1932.77 > 0$, we can use (4) and (5) to obtain the Fieller interval for μ_z. The 95% Fieller interval is calculated as (253.6042, 254.9165) Ω, which is almost identical to the POE_t interval.

3. Simulation study

The coverage probabilities of the POE_2 and POE_t intervals were examined using statistical simulation. This probability depends only on σ_p/μ_p, σ_q/μ_q, ρ, and n. Hence, without loss of generality, we can fix the values of μ_p and μ_q, and we choose $\mu_p = 4.999$ and $\mu_q = 19.661$ as in the ISO GUM annex H.2. Thus, the value of θ in the simulation study is 4.999/19.661.

We used the following grid of values for the unknown parameters in the simulation study to estimate coverage probabilities:

(a) $(\mu_p, \mu_q) = (4.999, 19.661)$,
(b) $n = \{3, 5, 10, 30\}$,
(c) $\kappa_p = \kappa_q = \{0.1\%, 0.5\%, 1\%, 2\%, 5\%, 10\%\}$,
(d) $\rho = \{\pm 0.01, \pm 0.1, \pm 0.5, \pm 0.7, \pm 0.99\}$.

In total, there were 1440 combinations of simulation parameters. For each combination of parameters, we simulated 100 000 independent realizations of (P, Q) according to their joint normal distribution (see appendix). Both $k = t_{0.975:n-1}$ and $k = 2$ were used to calculate $[L, U]$ and estimate p. Although the Fieller interval is exact, we included it in the study to verify the simulation and to compare with other intervals. The percentage of times that the intervals contained $\theta = 4.999/19.661$ and the widths of the intervals were both recorded. Due to the large quantity of data, for each given value of the parameters (μ_p, μ_q), we show only the case where the coverage probability of the POE_2 interval has the largest absolute deviation from the nominal value of 0.95 among the cases with different values of ρ. We also show the results only for $n = 5$ and 10. Tables 3 and 4 display the results. In these tables the columns labelled 'simulated' report the Monte Carlo estimates of coverage probabilities, those labelled 'theorem 1' give the large-sample approximation for the coverage probability based on theorem 1 (see appendix), and columns labelled 'Relative error' give the magnitude of the relative error of the approximation based on theorem 1 when compared to the Monte Carlo estimates.

For the parameters considered, the coverage probability of the POE_t interval is very close to the nominal value of 0.95. For the POE_2 interval, the coverage probability depends mainly on the sample size n. When $n = 3$, the coverage probability ranges from 0.8149 to 0.8176 for the 360 combinations of κ_p, κ_q, and ρ considered. For $n = 30$ the ranges of the coverage probabilities are 0.9436 and 0.9467. In all cases, the relative errors of the approximations based on theorem 1 are about one-tenth of 1% or smaller, indicating that the large-sample approximation performs extremely well even for sample sizes as low as 5. Additional investigations, not reported here, confirm that, for values of κ_q in the practical range, the approximation based on theorem 1 is highly satisfactory even for $n = 2$.

The results suggest that if one were to take the propagation-of-errors approach, then $k = t_{0.975:n-1}$ should be used to construct the 95% confidence interval for θ; $k = 2$ works only when n is large. This is not surprising since the value of $t_{0.975:n-1}$ is approximately equal to 2 when n is large. It is interesting to note that the POE_t intervals were nearly identical to the Fieller intervals in almost all cases.

Since both the POE_t and the Fieller intervals have comparable coverage probability, and the POE_t interval is symmetric about Y but the Fieller interval is not, we also compared their average widths. Although it is known that the expected width of the Fieller interval is infinite, we computed the empirical average widths for the Fieller intervals obtained in the simulation study. For the parameters considered, the POE_t interval is slightly shorter, on average, than the Fieller interval for the simulated data sets; the ratio of the average width of the POE_t interval to the average width of the Fieller interval ranged from 0.8882 to 1.0000.

The above study indicates that the POE_t interval is very similar to the exact Fieller interval. The POE_t interval is also consistent with the ISO uncertainty guidelines and thus can be recommended in metrological applications. The POE_2 approach is not suitable when n is small.

4. Conclusions

We have considered three methods for constructing confidence intervals for ratios of dependent measurements. They are (a) propagation of errors with coverage factor $k = 2$ (POE_2); (b) propagation of errors with coverage factor $k = t_{1-\alpha/2:n-1}$ (POE_t), and (c) Fieller's method. The Fieller interval is known

Table 3. Coverage probabilities of discussed intervals for the ratio θ ($n = 5$).

			POE$_2$			POE$_t$			
κ_p	κ_q	ρ	Simulated	Theorem 1	Relative error	Simulated	Theorem 1	Relative Error	Fieller
0.001	0.001	−0.99	0.8832	0.8839	0.000 77	0.9495	0.9500	0.000 53	0.9495
	0.005	−0.99	0.8834	0.8839	0.000 55	0.9492	0.9500	0.000 84	0.9491
	0.01	−0.99	0.8832	0.8839	0.000 78	0.9493	0.9500	0.000 74	0.9491
	0.02	−0.99	0.8833	0.8839	0.000 67	0.9492	0.9500	0.000 84	0.9492
	0.05	−0.99	0.8833	0.8839	0.000 69	0.9493	0.9500	0.000 72	0.9492
	0.1	−0.99	0.8836	0.8840	0.000 45	0.9490	0.9499	0.000 97	0.9492
0.005	0.001	0.99	0.8831	0.8839	0.000 89	0.9495	0.9500	0.000 53	0.9495
	0.005	−0.99	0.8833	0.8839	0.000 66	0.9496	0.9500	0.000 42	0.9495
	0.01	−0.99	0.8832	0.8839	0.000 78	0.9496	0.9500	0.000 42	0.9495
	0.02	−0.01	0.8837	0.8839	0.000 21	0.9499	0.9500	0.000 10	0.9499
	0.05	−0.99	0.8833	0.8839	0.000 69	0.9493	0.9500	0.000 72	0.9491
	0.1	−0.99	0.8836	0.8840	0.000 45	0.9491	0.9499	0.000 86	0.9492
0.01	0.001	0.50	0.8832	0.8839	0.000 77	0.9495	0.9500	0.000 53	0.9495
	0.005	0.99	0.8831	0.8839	0.000 89	0.9491	0.9500	0.000 95	0.9492
	0.01	−0.70	0.8835	0.8839	0.000 44	0.9497	0.9500	0.000 32	0.9498
	0.02	−0.99	0.8835	0.8839	0.000 44	0.9495	0.9500	0.000 52	0.9495
	0.05	−0.99	0.8832	0.8839	0.000 81	0.9493	0.9500	0.000 72	0.9491
	0.1	−0.99	0.8835	0.8840	0.000 57	0.9492	0.9499	0.000 76	0.9491
0.02	0.001	0.01	0.8831	0.8839	0.000 89	0.9494	0.9500	0.000 63	0.9494
	0.005	0.99	0.8832	0.8839	0.000 77	0.9493	0.9500	0.000 74	0.9493
	0.01	0.99	0.8832	0.8839	0.000 78	0.9490	0.9500	0.001 05	0.9492
	0.02	−0.70	0.8834	0.8839	0.000 55	0.9498	0.9500	0.000 21	0.9498
	0.05	−0.99	0.8836	0.8839	0.000 35	0.9496	0.9500	0.000 40	0.9494
	0.1	−0.99	0.8836	0.8840	0.000 45	0.9493	0.9499	0.000 65	0.9491
0.05	0.001	0.99	0.8834	0.8839	0.000 55	0.9495	0.9500	0.000 53	0.9496
	0.005	−0.70	0.8832	0.8839	0.000 77	0.9496	0.9500	0.000 42	0.9495
	0.01	0.99	0.8830	0.8839	0.001 00	0.9495	0.9500	0.000 53	0.9495
	0.02	−0.01	0.8832	0.8839	0.000 78	0.9497	0.9500	0.000 32	0.9497
	0.05	−0.99	0.8834	0.8839	0.000 58	0.9494	0.9500	0.000 61	0.9495
	0.1	−0.99	0.8835	0.8840	0.000 57	0.9492	0.9499	0.000 76	0.9495
0.1	0.001	−0.50	0.8833	0.8839	0.000 66	0.9497	0.9500	0.000 32	0.9497
	0.005	0.70	0.8831	0.8839	0.000 89	0.9496	0.9500	0.000 42	0.9495
	0.01	0.10	0.8832	0.8839	0.000 78	0.9494	0.9500	0.000 63	0.9494
	0.02	0.99	0.8831	0.8839	0.000 89	0.9495	0.9500	0.000 52	0.9495
	0.05	−0.99	0.8834	0.8839	0.000 58	0.9494	0.9500	0.000 61	0.9494
	0.1	−0.70	0.8838	0.8840	0.000 25	0.9498	0.9499	0.000 16	0.9498

to be exact, whereas the other two intervals are approximate. We have found that the POE$_t$ and Fieller intervals are very similar, and both perform better than the POE$_2$ method for the parameters considered in the study. The POE$_t$ interval is also consistent with the ISO uncertainty guidelines, and hence can be recommended in metrological applications. For those exceptional situations, where the coverage probability of the POE$_t$ interval is not sufficiently close to the desired value, improved coverage factors derived based on theorem 1 in the appendix may be used. A table of such values is provided in table 5.

Appendix

Suppose $(P_1, Q_1)^t, \ldots, (P_n, Q_n)^t$ is an independently and identically distributed sample of size n from a bivariate normal distribution with mean vector $(\mu_p, \mu_q)^t$ and covariance matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_p^2 & \rho\sigma_p\sigma_q \\ \rho\sigma_p\sigma_q & \sigma_q^2 \end{pmatrix}.$$

Let $\boldsymbol{S}$ be the sample variance covariance matrix (with $n - 1$ degrees of freedom), i.e.

$$\boldsymbol{S} = \begin{pmatrix} S_p^2 & \hat{\rho}S_pS_q \\ \hat{\rho}S_pS_q & S_q^2 \end{pmatrix}$$

and $\boldsymbol{A} = (n-1)\boldsymbol{S}$. The matrix $\boldsymbol{A}$ is called the sample sum of squares and cross-products matrix. For any fixed vector $\boldsymbol{L}$, $\boldsymbol{L}^t\boldsymbol{A}\boldsymbol{L}/\boldsymbol{L}^t\boldsymbol{\Sigma}\boldsymbol{L}$ is distributed as χ^2 with $n-1$ degrees of freedom ([9], p 535). In particular, with $\boldsymbol{L}^t = (1/Q, -P/Q^2)$, and conditional on P and Q, the quantity $G = n(n-1)u_y^2/\sigma_y^2$ has a χ^2 distribution with $n-1$ degrees of freedom, where

$$\sigma_y^2 = \frac{\sigma_p^2}{Q^2} - \frac{2P\sigma_p\sigma_q\rho}{Q^3} + \frac{P^2\sigma_q^2}{Q^4}.$$

Thus, it is reasonable to associate $n-1$ degrees of freedom with u_y.

Based on the above result, the coverage probability associated with interval $[Y - ku_y, Y + ku_y]$ is given by

$$p = \Pr\left[Y - ku_y \leqslant \theta \leqslant Y + ku_y\right],$$

$$p = \Pr\left[G \geqslant \frac{n(n-1)(Y-\theta)^2}{k^2\sigma_y^2}\right].$$

Table 4. Coverage probabilities of discussed intervals for the ratio θ ($n = 10$).

			POE$_2$			POE$_t$			
κ_p	κ_q	ρ	Simulated	Theorem 1	Relative error	Simulated	Theorem 1	Relative error	Fieller
0.001	0.001	0.99	0.9229	0.9234	0.000 59	0.9500	0.9500	0.000 00	0.9501
	0.005	0.10	0.9226	0.9234	0.000 92	0.9500	0.9500	0.000 00	0.9501
	0.01	0.01	0.9228	0.9234	0.000 70	0.9501	0.9500	0.000 11	0.9502
	0.02	0.10	0.9224	0.9234	0.001 14	0.9495	0.9500	0.000 53	0.9496
	0.05	0.10	0.9226	0.9234	0.000 93	0.9495	0.9500	0.000 52	0.9494
	0.1	0.01	0.9227	0.9235	0.000 84	0.9494	0.9500	0.000 60	0.9497
0.005	0.001	0.99	0.9235	0.9234	0.000 06	0.9507	0.9500	0.000 74	0.9507
	0.005	0.99	0.9230	0.9234	0.000 48	0.9501	0.9500	0.000 11	0.9501
	0.01	0.50	0.9229	0.9234	0.000 59	0.9499	0.9500	0.000 11	0.9497
	0.02	0.10	0.9227	0.9234	0.000 81	0.9500	0.9500	0.000 00	0.9501
	0.05	−0.01	0.9228	0.9234	0.000 71	0.9501	0.9500	0.000 11	0.9500
	0.1	0.10	0.9229	0.9235	0.000 63	0.9496	0.9500	0.000 39	0.9496
0.01	0.001	−0.70	0.9236	0.9234	0.000 17	0.9506	0.9500	0.000 63	0.9506
	0.005	−0.99	0.9235	0.9234	0.000 06	0.9506	0.9500	0.000 63	0.9506
	0.01	0.99	0.9230	0.9234	0.000 49	0.9501	0.9500	0.000 10	0.9501
	0.02	0.50	0.9229	0.9234	0.000 60	0.9498	0.9500	0.000 21	0.9497
	0.05	0.10	0.9229	0.9234	0.000 60	0.9501	0.9500	0.000 11	0.9501
	0.1	0.10	0.9227	0.9235	0.000 85	0.9494	0.9500	0.000 60	0.9497
0.02	0.001	0.50	0.9234	0.9234	0.000 05	0.9506	0.9500	0.000 63	0.9506
	0.005	0.99	0.9234	0.9234	0.000 05	0.9505	0.9500	0.000 53	0.9506
	0.01	−0.99	0.9235	0.9234	0.000 06	0.9506	0.9500	0.000 63	0.9506
	0.02	0.99	0.9231	0.9234	0.000 38	0.9502	0.9500	0.000 21	0.9501
	0.05	0.50	0.9228	0.9234	0.000 71	0.9495	0.9500	0.000 52	0.9497
	0.1	0.10	0.9232	0.9235	0.000 31	0.9500	0.9500	0.000 03	0.9501
0.05	0.001	0.99	0.9236	0.9234	0.000 17	0.9509	0.9500	0.000 95	0.9510
	0.005	0.50	0.9236	0.9234	0.000 17	0.9506	0.9500	0.000 63	0.9506
	0.01	0.99	0.9234	0.9234	0.000 05	0.9507	0.9500	0.000 74	0.9507
	0.02	−0.99	0.9238	0.9234	0.000 38	0.9508	0.9500	0.000 84	0.9507
	0.05	0.99	0.9231	0.9234	0.000 40	0.9503	0.9500	0.000 29	0.9501
	0.1	0.50	0.9226	0.9235	0.000 97	0.9496	0.9500	0.000 42	0.9497
0.1	0.001	0.99	0.9235	0.9234	0.000 06	0.9510	0.9500	0.001 05	0.9510
	0.005	0.50	0.9233	0.9234	0.000 16	0.9506	0.9500	0.000 63	0.9506
	0.01	0.50	0.9236	0.9234	0.000 16	0.9506	0.9500	0.000 63	0.9506
	0.02	0.99	0.9237	0.9234	0.000 27	0.9507	0.9500	0.000 74	0.9507
	0.05	−0.99	0.9237	0.9234	0.000 26	0.9510	0.9500	0.001 06	0.9506
	0.1	0.99	0.9234	0.9235	0.000 16	0.9504	0.9500	0.000 33	0.9501

Using the standard conditioning argument of probability, we have

$$p = \mathrm{E}\left[\Pr\left(G \geqslant \frac{n(n-1)(Y-\theta)^2}{k^2\sigma_y^2}\,\middle|\, P, Q\right)\right],$$

$$p = \mathrm{E}\left[1 - H_{n-1}\left(\frac{n(n-1)(Y-\theta)^2}{k^2\sigma_y^2}\right)\right],$$

where $H_{n-1}(\cdot)$ represents the cumulative distribution function of a χ^2 random variable with $n-1$ degrees of freedom.

The expression for p may be further simplified as follows. Let $P_s = \sqrt{n}(P/\mu_p - 1)$ and $Q_s = \sqrt{n}(Q/\mu_q - 1)$. Then, the distribution of $(P_s, Q_s)^t$ is given by

$$\begin{pmatrix} P_s \\ Q_s \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \kappa_p^2 & \rho\kappa_p\kappa_q \\ \rho\kappa_p\kappa_q & \kappa_q^2 \end{bmatrix}\right).$$

Expressing p in terms of P_s and Q_s, we get

$$p = \mathrm{E}_{P_s,Q_s}\left[1 - H_{n-1}\left(\frac{(n-1)Y_s}{k^2}\right)\right],$$

where

$$Y_s = \frac{(P_s - Q_s)^2}{\kappa_p^2 - 2\rho\kappa_p\kappa_q Z_s + \kappa_q^2 Z_s^2}$$

and

$$Z_s = 1 + \frac{P_s - Q_s}{Q_s + \sqrt{n}}.$$

Then, for each given value of the parameter vector $(\kappa_p, \kappa_q, \rho)$, and sample size n, p can either be evaluated using numerical double integration or estimated by a Monte Carlo approach. However, an excellent approximation for p is available, as indicated in theorem 1.

Theorem 1. *Let T_{n-1} be a random variable having a t distribution with $n-1$ degrees of freedom. Then,*

$$p = \Pr(|T_{n-1}| < k) + C_n + \mathrm{o}\left(\frac{1}{n}\right) \qquad (6)$$

with

$$C_n = l(n,k)[(n-1)k^3(4\kappa_q^2 a_1 + 5a_2^2) - k^5\{(n-3)a_2^2 - 4\kappa_q^2 a_1\}],$$

where

$$a_1 = \frac{2\kappa_p^2(1-\rho^2)}{\kappa_p^2 - 2\rho\kappa_p\kappa_q + \kappa_q^2} - 1,$$

$$a_2 = \frac{2\kappa_q(\kappa_q - \rho\kappa_p)}{\sqrt{\kappa_p^2 - 2\rho\kappa_p\kappa_q + \kappa_q^2}}$$

Table 5. Improved coverage factors k. The first column gives the sample size n. Columns 2 and 3 are improved coverage factors k for the cases $\kappa_q \leqslant 0.1$ and $\kappa_q \leqslant 0.5$, respectively, with a confidence level equal to 95%. Column 4 is the corresponding t-table value with $n-1$ degrees of freedom. Columns 5 and 6 are improved coverage factors k for the cases $\kappa_q \leqslant 0.1$ and $\kappa_q \leqslant 0.5$, respectively, with a confidence level equal to 99%. Column 7 is the corresponding t-table value with $n-1$ degrees of freedom.

	Confidence level = 95%			Confidence level = 99%		
	Coverage factor k for		t-table value	Coverage factor k for		t-table value
n	$\kappa_q \leqslant 0.1$	$\kappa_q \leqslant 0.5$	$(n-1)$ df	$\kappa_q \leqslant 0.1$	$\kappa_q \leqslant 0.5$	$(n-1)$ df
2	12.7061	12.7032	12.7062	63.6567	63.6566	63.6567
3	4.3049	4.3580	4.3027	9.9292	10.0327	9.9248
4	3.1846	3.2379	3.1824	5.8476	6.0031	5.8409
5	2.7781	2.8199	2.7764	4.6114	4.7813	4.6041
6	2.5718	2.6028	2.5706	4.0392	4.2065	4.0321
7	2.4478	2.4706	2.4469	3.7140	3.8717	3.7074
8	2.3653	2.3823	2.3646	3.5055	3.6513	3.4995
9	2.3065	2.3193	2.3060	3.3608	3.4945	3.3554
10	2.2625	2.2724	2.2622	3.2548	3.3768	3.2498
11	2.2284	2.2361	2.2281	3.1737	2.2361	3.1693
12	2.2012	2.2073	2.2010	3.1099	2.2073	3.1058
13	2.1790	2.1838	2.1788	3.0583	2.1838	3.0545
14	2.1605	2.1644	2.1604	3.0157	2.1644	3.0123
15	2.1449	2.1481	2.1448	2.9800	2.1481	2.9768
16	2.1316	2.1341	2.1314	2.9496	2.1341	2.9467
17	2.1200	2.1221	2.1199	2.9235	2.1221	2.9208
18	2.1099	2.1117	2.1098	2.9007	2.1117	2.8982
19	2.1010	2.1025	2.1009	2.8808	2.1025	2.8784
20	2.0931	2.0943	2.0930	2.8631	2.0943	2.8609

and $l(n, k)$ is as given in (3). The expression $o(1/n)$ *refers to a term that approaches zero at a rate faster than* $1/n$ *as n tends to infinity. For a proof of this theorem, the reader may refer to [10].*

The term C_n is usually negligible for realistic choices of κ_p, κ_q, and ρ. In fact, it is a relatively simple exercise to show that if $k > \sqrt{3}$

$$4l(n,k)c_1(n,k)\,k^3\,\kappa_q^2 \leqslant C_n \leqslant 4l(n,k)c_2(n,k)\,k^3\,\kappa_q^2, \quad (7)$$

where

$$c_1(n,k) = 4n-4-k^2(n-2) \qquad \text{and} \qquad c_2(n,k) = n-1+k^2.$$

For sample sizes n in the range between 2 and 20, table 5 gives improved coverage factors k that give guaranteed coverage probability very nearly equal to or greater than 95% (respectively, 99%) for two cases: (a) when the value of κ_q may be assumed to be no greater than 0.1, and (b) when the value of κ_q may be assumed to be no greater than 0.5. For comparison, the corresponding t-table values with $n-1$ degrees of freedom are also shown. These coverage factors were obtained by solving for k after setting the bound p_L equal to 0.95 (respectively, 0.99) in equation (2). Observe that the t-table values are quite close to the improved coverage factors k, especially when κ_q is no greater than 0.1.

References

[1] Fieller E C 1940 *J. R. Stat. Soc.* **7** 1–64

[2] Fieller E C 1944 *Q. J. Pharmacy* **17** 117–23

[3] Fieller E C 1954 *J. R. Stat. Soc.* B **16** 175–85

[4] Creasy M A 1954 *J. R. Stat. Soc.* B **16** 186–94

[5] Kappenman R F, Geysser S and Antle C E 1970 *Sankhyā* B **3** 331–40

[6] Buonaccorsi J P 1985 *Commun. Stat.—Theory Methods* **14** 635–50

[7] Buonaccorsi J P and Gatsonis C A 1988 *Biometrics* **44** 87–101

[8] International Organization for Standardization (ISO) 1995 *Guide to the Expression of Uncertainty in Measurement* (Geneva, Switzerland: International Organization for Standardization)

[9] Rao C R 1973 *Linear Statistical Inference and its Applications* 2nd edn (New York: Wiley)

[10] Hannig J 2003 *Asymptotic Bounds for Coverage Probabilities for a Class of Confidence Intervals for the Ratio of Means in a Bivariate Normal Distribution, Technical Report* 2003-3, Department of Statistics, Colorado State University, Fort Collins, Colorado