Asymptotic Bounds for Coverage Probabilities for a Class of Confidence Intervals for the Ratio of Means in a Bivariate Normal Distribution

JAN HANNIG *
Department of Statistics
Colorado State University

December 8, 2005

Abstract

Finding 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. Such problems arise frequently in the field of metrology since many quantities of interest are calculated by first measuring more basic quantities and then using the ratio of the basic quantities. In this paper we develop a simple asymptotic formula that allows us to estimate the true coverage of this confidence interval. In particular, the formula allows one to compute a value for k which will result in an associated confidence coefficient very nearly equal to the desired value.

Keywords: Confidence Intervals, ratios, asymptotic evaluations, Gaussian random variables.

1 Introduction

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

*Jan Hannig is an Assistant Professor in the Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877. Phone: (970) 491-7460. Fax: (970) 491-7895. Email: `hannig@stat.colostate.edu`. Jan Hannig's research is supported in part by National Science Foundation under grant DMS-0504737.

attention of numerous researchers (e.g., Fieller (1954), Buonaccorsi & Gatsonis (1988)). Such problems arise frequently in the field of metrology since many quantities of interest are calculated by first measuring more basic quantities and then using the ratio of the basic quantities. For instance, Ohm's law tells us that $I = V/R$ where I is current, V is voltage and R is resistance. Thus, current in a conductor may be "measured" by first measuring voltage and resistance and then taking the ratio. The Guide to the expression of uncertainty in Measurements (commonly referred to as the GUM) available from the International Organization for Standardization (usually referred to as the ISO) provides guidelines for calculating and reporting uncertainties in measurements based on the method of propagation of errors. The ISO GUM (GUM 1995) has become the *de facto* standard for calculating and reporting uncertainties at national measurement laboratories of many countries as well as many industries. Metrologists routinely use confidence intervals of the form $[\hat{\theta} - k\, u_{\hat{\theta}}, \hat{\theta} + k\, u_{\hat{\theta}}]$ to express the uncertainty in an estimated ratio $\hat{\theta}$ where $u_{\hat{\theta}}$ is an estimated standard deviation of $\hat{\theta}$ using the method of propagation of errors, and k is a prespecified constant called the coverage factor. Metrologists routinely use a coverage factor of 2 for calculating a confidence interval for the ratio quantity. This confidence interval is presumed to have a level of confidence equal to 95%. Unfortunately this presumption is often rather far from truth. In this paper we develop a simple asymptotic formula that allows us to estimate the true coverage of this confidence interval. In particular, the formula allows one to compute a value for k which will result in an associated confidence coefficient very nearly equal to the desired value.

2 Main Results

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

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

We are interested in estimating $\theta = \mu_p/\mu_q$. The propagation of error method uses

$$\hat{\theta} = \frac{P}{Q}, \quad \text{with} \quad P = \frac{1}{n}\sum_{i=1}^{n} P_i, \quad Q = \frac{1}{n}\sum_{i=1}^{n} Q_i,$$

as an estimator for θ. A confidence interval

$$(1) \qquad (\hat{\theta} - k\, u_{\hat{\theta}},\ \hat{\theta} + k\, u_{\hat{\theta}})$$

is often used to quantify the uncertainty of the estimator $\hat{\theta}$. The quantity k is a pre-chosen constant and

$$u_{\hat{\theta}} = \frac{P}{\sqrt{n}Q}\left[\frac{S_p^2}{P^2} + \frac{S_q^2}{Q^2} - \frac{2\hat{\rho}S_pS_q}{PQ}\right]^{1/2},$$

where $S_p,\ S_q,\ \hat{\rho}$ are the sample standard deviations and sample correlation respectively (using the $n-1$ denominator).

It is of significant importance for practitioners to know the coverage of the confidence interval in (1). For an ease of notation denote the exact coverage by p_n.

Let $\boldsymbol{S}$ be the sample variance covariance matrix, 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 $\boldsymbol{L}$, $\boldsymbol{L}^\top\boldsymbol{A}\boldsymbol{L}/\boldsymbol{L}^\top\boldsymbol{\Sigma}\boldsymbol{L}$ is distributed as χ^2 with $n-1$ degrees of freedom (Rao (1973), p. 535). In particular, with $\boldsymbol{L}^\top = (1/Q,\ -P/Q^2)$, and conditional on P and Q, the quantity $G = n(n-1)u_{\hat{\theta}}^2/\sigma_{\hat{\theta}}^2$ has a χ^2-distribution with $n-1$ degrees of freedom, where

$$\sigma_{\hat{\theta}}^2 = \sigma_p^2/Q^2 - 2P\sigma_p\sigma_q\rho/Q^3 + P^2\sigma_q^2/Q^4.$$

From here, the coverage probability associated with the interval (1) is equal to

$$\begin{aligned} p_n &= P\left[\hat{\theta} - ku_{\hat{\theta}} \le \theta \le \hat{\theta} + ku_{\hat{\theta}}\right] \\ &= P\left[G \ge n(n-1)(\hat{\theta}-\theta)^2/k^2\sigma_{\hat{\theta}}^2\right]. \end{aligned}$$

Using the standard conditioning argument of probability, we have

$$\begin{aligned} p_n &= \mathrm{E}\left[\Pr\left(G \ge n(n-1)(\hat{\theta}-\theta)^2/k^2\sigma_{\hat{\theta}}^2 \,\middle|\, P, Q\right)\right] \\ &= \mathrm{E}\left[1 - G_{n-1}\left(n(n-1)(\hat{\theta}-\theta)^2/k^2\sigma_{\hat{\theta}}^2\right)\right] \end{aligned}$$

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

The expression for p_n 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 $(P_s,\, Q_s)^\top$ is given by

$$\begin{pmatrix} P_s \\ Q_s \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \kappa_P^2 & \rho\kappa_P\kappa_Q \\ \rho\kappa_P\kappa_Q & \kappa_Q^2 \end{pmatrix}\right)$$

Expressing p_n in terms of P_s and Q_s, we get

$$p_n = \mathrm{E}_{P_s, Q_s} \left[1 - G_{n-1} \left((n-1) Y / k^2 \right) \right]$$

where

$$Y = \frac{(P_s - Q_s)^2}{\kappa_P^2 - 2\rho\kappa_P\kappa_Q Y_s + \kappa_Q^2 Y_s^2} \tag{2}$$

and

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

Then, for each given value of the parameter vector (κ_P, κ_Q, ρ), and sample size n, p_n can be evaluated using numerical double integration or estimated by a Monte-Carlo approach. However an analytical expression for this probability is more desirable as one can use such an expression to determine a more appropriate coverage factor k. While an exact analytical expression is unavailable, we derive an asymptotic expansion for p_n as $n \to \infty$. This is given in Theorem 1. This expansion does lead to an analytical approximation for p_n. Simulation studies reported in Hannig, Wang & Iyer (2003) lead us to conclude that the expression is excellent for situations encountered in metrological applications.

Theorem 1. *Let T_{n-1} have a t distribution with $n-1$ degrees of freedom.*

$$p_n = P(T_{n-1}^2 < k^2) + C(n, k, \kappa_P, \kappa_Q, \rho) + o\left(\frac{1}{n}\right). \tag{3}$$

with

$$C(n, k, \kappa_P, \kappa_Q, \rho) = l(n, k) \left[(n-1)k^3(4\kappa_Q^2 a + 5c^2) - k^5\{(n-3)c^2 - 4\kappa_Q^2 a\} \right], \tag{4}$$

where

$$a = \frac{2\kappa_P^2(1-\rho^2)}{\kappa_P^2 - 2\rho\kappa_P\kappa_Q + \kappa_Q^2} - 1, \qquad c = \frac{2\kappa_Q(\kappa_Q - \rho\kappa_P)}{\sqrt{\kappa_P^2 - 2\rho\kappa_P\kappa_Q + \kappa_Q^2}}, \tag{5}$$

and

$$l(n, k) = \frac{1}{8\sqrt{\pi}} \frac{\Gamma(n/2)}{\Gamma((n+1)/2)} \frac{(n-1)^{(n+1)/2}}{(n-1+k^2)^{(n+4)/2}}.$$

and $\Gamma(x)$ is the gamma function.

Remark 1. Consider (4). By applying Stirling's approximation to the gamma function (see Abramowitz & Stegun (1970) formula 6.1.37) we can get the following asymptotically equivalent form

$$C(n,k,\kappa_P,\kappa_Q,\rho) = \phi(k)\frac{k^3(4\kappa_Q^2 a + 5c^2) - k^5c^2}{4n} + o\left(n^{-1}\right), \tag{6}$$

where ϕ is a standard normal density. The reason we prefer the more complicated form of (4) to (6) is a better performance for small n.

Remark 2. The term $C(n,k,\kappa_P,\kappa_Q,\rho)$ is usually negligible for realistic choices of κ_P, κ_Q and ρ. In fact, it is a relatively simple calculus exercise to show that if $k > \sqrt{3}$

$$\kappa_Q^2\ 4l(n,k)k^3c_1(n,k) \le C(n,k,\kappa_P,\kappa_Q,\rho) \le \kappa_Q^2\ 4l(n,k)k^3c_2(n,k) \tag{7}$$

where

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

The bound (7) works for all values κ_P, κ_Q and ρ.

3 Proof

First we will get a precise asymptotic expansion of Y as defined in (2). The Taylor series approximation to Y is

$$Y = Y_0 + \frac{Y_1}{\sqrt{n}} + \frac{Y_2}{n} + R_n, \tag{8}$$

$$Y_0 = Z^2, \quad Y_1 = -cZ^3, \quad Y_2 = -\kappa_Q^2 Z^3 V, \quad R_n = O\left(n^{-3/2}\right), \tag{9}$$

where

$$Z = \frac{P_s - Q_s}{\sqrt{\kappa_P^2 - 2\rho\kappa_P\kappa_Q + \kappa_Q^2}},$$

$$V = \frac{[\kappa_Q\kappa_P^2(1-4\rho^2) + 6\kappa_P\kappa_Q^2\rho - 3\kappa_Q^3]P_s + (\kappa_Q^3 - 3\kappa_P^2\kappa_Q + 2\kappa_P^3\rho)Q_s}{\kappa_Q(\kappa_P^2 - 2\rho\kappa_P\kappa_Q + \kappa_Q^2)^{3/2}}$$

and c was defined in (5). Here and in what follows if the O or o notation involves random variables it holds for each fixed ω.

Simple calculation shows that

$$\left(\begin{array}{c} Z \\ V \end{array}\right) \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \frac{2\kappa_P^2(1-\rho^2)}{\kappa_P^2-2\rho\kappa_P\kappa_Q+\kappa_Q^2}-1 \\ \frac{2\kappa_P^2(1-\rho^2)}{\kappa_P^2-2\rho\kappa_P\kappa_Q+\kappa_Q^2}-1 & 1 \end{array}\right)\right),$$

whence there is $Z' \sim N(0,1)$ independent of Z such that $V = aZ + bZ'$ and

$$a = \frac{2\kappa_P^2(1-\rho^2)}{\kappa_P^2 - 2\rho\kappa_P\kappa_Q + \kappa_Q^2} - 1, \quad b = \frac{2\kappa_P(\kappa_Q - \kappa_P\rho)\sqrt{1-\rho^2}}{\kappa_P^2 - 2\rho\kappa_P\kappa_Q + \kappa_Q^2}.$$

Plugging this into (9) we get $Y_2 = -\kappa_Q^2(aZ^4 + bZ^3Z')$.

Now we calculate:

$$\begin{aligned} E\left[1 - G_{n-1}\left(\frac{n-1}{k^2}Y\right)\right] =& E\left[1 - G_{n-1}\left(\frac{n-1}{k^2}Z^2\right)\right] \\ &+ E\left[G_{n-1}\left(\frac{n-1}{k^2}Z^2\right) - G_{n-1}\left(\frac{n-1}{k^2}Y\right)\right] \end{aligned} \tag{10}$$

Consider the first part of the right-hand-side (r.h.s.) of (10) first. Let C be a random variable having chi-square distribution with $n-1$ degrees of freedom. For any s

$$1 - G_{n-1}\left(\frac{n-1}{k^2}s\right) = P\left(C > \frac{n-1}{k^2}s\right).$$

From here

$$\begin{aligned} E\left[1 - G_{n-1}\left(\frac{n-1}{k^2}Z^2\right)\right] =& EP\left(C > \frac{n-1}{k^2}Z^2 \Big| Z\right) \\ =& P\left(\frac{Z^2}{C/(n-1)} < k^2\right) \\ =& P(T^2 < k^2), \end{aligned} \tag{11}$$

where T has a t distribution with $n-1$ degrees of freedom.

Consider now the second part of the r.h.s. of (10). By a direct calculation using a second order Taylor's series approximation to the function $f(x) = G_{n-1}\left(\frac{n-1}{k^2}x\right)$ at the point Z^2 one can, after taking an expectation, arrive to formula (4). Unfortunately, proving that the remainder of the Taylor's series is of the order $o(n^{-1})$ directly appears to be rather delicate.

Therefore we will proceed indirectly and prove the form (6) instead.

$$
\begin{aligned}
&E\left[G_{n-1}\left(\frac{n-1}{k^2}Z^2\right)-G_{n-1}\left(\frac{n-1}{k^2}Y\right)\right]\\
&\qquad=E\left[G_{n-1}\left(\frac{n-1}{k^2}Z^2\right)-\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Z^2-k^2}{k^2}\right)\right]\\
&\qquad+E\left[\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Z^2-k^2}{k^2}\right)-I_{[0,\infty)}(Z^2-k^2)\right]\\
&\qquad+E\left[I_{[0,\infty)}(Z^2-k^2)-I_{[0,\infty)}(Y-k^2)\right]\\
&\qquad+E\left[I_{[0,\infty)}(Y-k^2)-\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Y-k^2}{k^2}\right)\right]\\
&\qquad+E\left[\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Y-k^2}{k^2}\right)-G_{n-1}\left(\frac{n-1}{k^2}Y\right)\right]
\end{aligned}
\tag{12}
$$

We will now investigate the five parts of the r.h.s. of (12) separately.

Define

$$L_n=\sqrt{\frac{n-1}{2}}\frac{Z^2-k^2}{k^2}$$

and notice that the distribution of L_n is more and more diffuse, so the density $f_{L_n}(x)$ of L_n converges to 0 for each x. More specifically, the Taylor series expansion

$$f_{L_n}(x)=\sum_{i=0}^{\infty}\frac{a_{i,n}}{i!}x^i,$$

satisfies $a_{i,n}=O(n^{-i/2})$ and in particular

$$a_{1,n}=-\frac{e^{-\frac{k^2}{2}}(k+k^3)}{(n-1)\sqrt{2\,\pi}}.$$

We will treat the second part of the r.h.s. of (12) first:

$$
\begin{aligned}
&E\left[\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Z^2-k^2}{k^2}\right)-I_{[0,\infty)}(Z^2-k^2)\right]\\
&\qquad=\int_{-\infty}^{\infty}\left[\Phi(x)-I_{[0,\infty)}(x)\right]f_{L_n}(x)\,dx\\
&\qquad=\int_{-\infty}^{\infty}\left[\Phi(x)-I_{[0,\infty)}(x)\right](a_{0,n}+a_{1,n}x)\,dx+o\left(\frac{1}{n}\right)\\
&\qquad=-\frac{a_{1,n}}{2}+o\left(\frac{1}{n}\right)=\frac{e^{-\frac{k^2}{2}}(k+k^3)}{2(n-1)\sqrt{2\,\pi}}+o\left(\frac{1}{n}\right).
\end{aligned}
\tag{13}
$$

To get the asymptotic behavior of the first part of the r.h.s. of (12) notice that Chi-square distribution is obtained as a sum of independent identically distributed random variables. Therefore $G_{n-1}\left((n-1)Z^2/k^2\right)$ will converge to $\Phi\left(L_n\right)$ by central limit theorem. In order to estimate the higher order terms in this convergence we will use Edgeworth's expansion (see for example Abramowitz & Stegun (1970) formula 26.2.48). In particular after some algebra we get

$$
G_{n-1}\left(\frac{n-1}{k^2}Z^2\right)=\Phi(L_n)+\phi(L_n)\left[\frac{\sqrt{2}}{3\sqrt{n}}(L_n^2-1)\right]+O\left(\frac{1}{n}\right),
$$

where the remainder term is uniform in L_n. From here and the asymptotic behavior of $a_{i,n}$ we get

$$
\begin{aligned}
&E\left[G_{n-1}\left(\frac{n-1}{k^2}Y\right)-\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Y-k^2}{k^2}\right)\right]\\
&\qquad=-\int_{-\infty}^{\infty}\left\{\phi(x)\left[\frac{\sqrt{2}}{3\sqrt{n}}(x^2-1)\right]\right\}a_{0,n}\,dx+o\left(\frac{1}{n}\right)\\
&\qquad=o\left(\frac{1}{n}\right).
\end{aligned}
\tag{14}
$$

To obtain the asymptotic behavior of the third part of r.h.s. of (12) notice

$$
EI_{[0,\infty)}(Z^2-k^2)=P(Z^2>k^2),\quad EI_{[0,\infty)}(Y-k^2)=1-P(Y<k^2).
$$

Furthermore

$$
P(Y<k^2)=P\left(Z^2-\frac{c}{\sqrt{n}}Z^3-\frac{\kappa_Q^2}{n}(aZ^4+bZ^3Z')+R_n<k^2\right)
$$

By conditioning on Z' we get

$$P(Y < k^2) = E\left[P(z_1 < Z < z_2|Z')\right],$$

where

$$z_1 = -k + \frac{ck^2}{2\sqrt{n}} - \frac{5c^2k^3}{8n} - \frac{a\kappa_Q^2 k^3}{2n} + \frac{b\kappa_Q^2 k^2 Z'}{2n} + o\left(\frac{1}{n}\right),$$
$$z_2 = k + \frac{ck^2}{2\sqrt{n}} + \frac{5c^2k^3}{8n} + \frac{a\kappa_Q^2 k^3}{2n} + \frac{b\kappa_Q^2 k^2 Z'}{2n} + o\left(\frac{1}{n}\right),$$

where the asymptotic expression for z_1 and z_2 is obtained from approaching the equation $Z^2 = k^2 + \frac{c}{\sqrt{n}} Z^3 + \frac{\kappa_Q^2}{n}(aZ^4 + bZ^3 Z') - R_n$ as quadratic equation $Z^2 = k^2 +$ perturbation and finding the solution by an iterative procedure.

From the Taylor series approximation to the standard normal c.d.f.

$$P(Z < k + x) = \Phi(k) + \phi(k)x - k\phi(k)x^2/2 + o(x^2)$$

we get

$$P(z_1 < Z < z_2|Z') = P(-k < Z < k|Z') + \frac{e^{-\frac{k^2}{2}}(k^3(5c^2 + 4a\kappa_Q^2) - k^5c^2)}{4n\sqrt{2\pi}} + o\left(\frac{1}{n}\right).$$

From here we conclude

$$1 - P(Y < k^2) = P(Z^2 > k^2) - \frac{e^{-\frac{k^2}{2}}(k^3(5c^2 + 4a\kappa_Q^2) - k^5c^2)}{4n\sqrt{2\pi}} + o\left(\frac{1}{n}\right).$$

This implies

$$E\left[I_{[0,\infty)}(Z^2 - k^2) - I_{[0,\infty)}(Y - k^2)\right] = \frac{e^{-\frac{k^2}{2}}(k^3(5c^2 + 4a\kappa_Q^2) - k^5c^2)}{4n\sqrt{2\pi}} + o\left(\frac{1}{n}\right). \tag{15}$$

The calculation of the asymptotic expansion of the last to parts of the r.h.s. of (12) is analogous to the calculation for the first two parts. Define

$$\tilde{L}_n = \sqrt{\frac{n-1}{2}}\frac{Y - k^2}{k^2}$$

Again the distribution of $\tilde{L}_n$ is more and more diffuse, the Taylor series expansion

$$f_{\tilde{L}_n}(x) = \sum_{i=0}^{\infty} \frac{\tilde{a}_{i,n}}{i!} x^i,$$

satisfies $\tilde{a}_{i,n} = O(n^{-i/2})$ and in particular

$$\tilde{a}_{1,n} = -\frac{e^{-\frac{k^2}{2}}(k+k^3)}{(n-1)\sqrt{2\pi}} + o\left(\frac{1}{n}\right).$$

Therefore as before

$$E\left[I_{[0,\infty]} - \Phi\left(\sqrt{\frac{n-1}{2}}\frac{Y-k^2}{k^2}\right)(Z^2-k^2)\right] = -\frac{e^{-\frac{k^2}{2}}(k+k^3)}{2(n-1)\sqrt{2\,\pi}} + o\left(\frac{1}{n}\right). \tag{16}$$

$$E\left[\Phi\left(\sqrt{\frac{n-1}{2}}\frac{Y-k^2}{k^2}\right) - G_{n-1}\left(\frac{n-1}{k^2}Y\right)\right] = o\left(\frac{1}{n}\right). \tag{17}$$

Notice that equations (13) and (16) have opposite signs. Combining equations (10), (11), (12), (13), (14), (15), (16), (17) we get (6). Finally, (4) follows from (6) by Remark 1.

4 Acknowledgments

I would like to thank the anonymous referees for constructive comments that improved final version of this paper.

References

Abramowitz, M. & Stegun, I. A. (1970), *Hanbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover.

Buonaccorsi, J. P. & Gatsonis, C. A. (1988), 'Bayesian inference for ratios of coefficients in a linear model', *Biometrics* **44**(1), 87–101.

Fieller, E. C. (1954), 'Symposium on interval estimation: Some problems in interval estimation', *J. Roy. Statist. Soc. Ser. B.* **16**, 175–185.

GUM (1995), *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization (ISO), Geneva, Switzerland.

Hannig, J., Wang, C. M. & Iyer, H. K. (2003), 'Uncertainty calculation for the ratio of dependent measurements', *Metrologia* **4**, 177–186.

Rao, C. R. (1973), *Linear statistical inference and its applications*, second edn, John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.