International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 25(1), March, 2025, pp 11-22
DOI: https://doi.org/10.3329/ijss.v25i1.81041

Optimal Prediction and Tolerance Intervals for the Ratio of Dependent Normal Random Variables

K. Krishnamoorthy[1*] and Gboyega Adepoju[2]

[1,2]Department of Mathematics, University of Louisiana at Lafayette, USA

[*]Correspondence should be addressed to K. Krishnamoorthy
(Email: krishna@louisiana.edu)

[Received December 25, 2024; Accepted February 20, 2025]

Abstract

A simple exact method is proposed for computing prediction intervals and tolerance intervals for the distribution of the ratio X_1/X_2 when (X_1, X_2) follows a bivariate normal distribution. The methodology uses the factors available for computing one-sample prediction intervals and tolerance intervals for a univariate normal distribution. Both one-sided and two-sided intervals are constructed, and the two-sided tolerance intervals are obtained with and without imposing the equal-tail requirement. The results are illustrated using two practical applications that call for the computation of prediction intervals and tolerance intervals for the distribution of the ratio X_1/X_2. The first application is an investigation of retroviral contamination in the raw materials used for the manufacture of the influenza vaccine FluMist. The second application is on the cost-effectiveness of a new drug compared to a standard drug.

Keywords: Equal-tailed tolerance interval, Non-central t, Prediction factor, Tolerance factor.

AMS Classification: 62F25.

1. Introduction

This work is motivated by two applications, available in the literature, that call for the computation of prediction intervals and tolerance intervals for the ratio of the marginal random variables in a bivariate normal distribution. The first application is an investigation of retroviral contamination in the raw materials used the manufacture of the influenza vaccine FluMist, and the second application is on the cost-effectiveness of a new drug compared to a standard drug.

A major component of statistical data analysis is the computation of appropriate intervals, motivated by specific applications. These intervals include confidence intervals, prediction intervals and tolerance intervals. As is well known, such intervals can be computed under both parametric and non-parametric scenarios. We recall that a prediction interval is an interval computed using a random sample, satisfying the condition that the interval will include the future values of a random variable with a given confidence level (say, γ). On the other hand, a tolerance interval for a distribution is an interval computed using a random sample, subject to the condition that the interval includes a specified proportion (say, p) or more of the distribution, with a specified confidence level (say, γ). The proportion p associated with a tolerance interval is referred to as its *content* and the interval is often referred to as a (p, γ) tolerance interval. Prediction intervals and tolerance intervals could be one-sided, with only a lower limit or only an upper limit, or they could be two-sided. A (p, γ) tolerance interval has several practical applications; the book by Krishnamoorthy and Mathew (2009) provides a detailed treatment of the topic.

The present work is on the computation of prediction intervals and tolerance intervals in a specific parametric scenario, namely, a bivariate normal setup where interest is focused on the ratio of the random variables. Let $\boldsymbol{X} = (X_1, X_2)'$ be a bivariate normal random vector with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. We want to compute one-sided and two-sided prediction intervals and tolerance intervals for X_1/X_2 based on a sample. If we write $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, the computation of a confidence interval for μ_1/μ_2 is a well-investigated problem. However, we are not aware of any work that addresses the computation of prediction intervals for X_1/X_2. Furthermore, only very limited literature is available for computing tolerance intervals for X_1/X_2; see Zhang et al. (2010) and Flouri et al. (2017). An earlier article by Hall and Sampson (1973) provides an approximate solution assuming that the coefficients of variation are small; the paper also includes an application related to drug development. In their works, Zhang et al. (2010) and Flouri et al. (2017) discuss applications relevant to bioassays and cost-effectiveness analysis where it is necessary to compute tolerance intervals for a ratio random variable of the type X_1/X_2. Ratio random variables also arise in the context of process control problems in manufacturing applications; see Celano et al. (2014), Celano and Castagliola (2016) and Tran et al. (2021) for details and other relevant references. Furthermore, tolerance limits have been recommended to be used as process control limits; see Hamada (2003) and Alqurashi (2021).

The methodology in Zhang et al. (2010) is based on the *generalized variables approach*, and the authors address the development of a one-sided tolerance limit only. Here we shall not elaborate on their approach. On the other hand, Flouri et al. (2017) address the computation of both one-sided tolerance limits and two-sided tolerance intervals;

they appeal to non-parametric methods for computing the tolerance limits based on a parametric bootstrap sample, and then apply a bootstrap calibration so as to have better coverage probabilities. Thus the methodology is computationally demanding. Our work presents a simple and accurate approach for computing exact one-sided and two-sided prediction intervals and tolerance intervals for the ratio X_1/X_2 in the bivariate normal scenario. Our intervals are in fact exact; i.e., they guarantee the coverage probability. However, our approach produces bona fide prediction limits (or tolerance limits) only when the marginal lower prediction limits (respectively, marginal lower tolerance limits) are positive. We expect this to be the case when the support of the marginal distributions is mostly on the positive part of the real line. In addition to deriving the prediction intervals and tolerance intervals, we have also illustrated our methodology by applying it to the two applications that motivated our work, and mentioned at the beginning of this section.

2. Prediction Intervals and Tolerance Intervals

Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ be a sample from a bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Furthermore, let $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ denote the sample mean vector and sample covariance matrix, respectively. Write

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \bar{\boldsymbol{X}} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix} S_1^2 & \widehat{\rho}S_1S_2 \\ \widehat{\rho}S_1S_2 & S_2^2 \end{pmatrix},$$

where ρ and $\widehat{\rho}$, respectively, denote the population correlation coefficient and the sample correlation coefficient. We start with a derivation of our proposed prediction interval for X_1/X_2 when $\boldsymbol{X} = (X_1, X_2)' \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

2.1. Prediction Intervals

A $100\gamma\%$ prediction interval (PI) for X_1/X_2, say $\big(L(k; \bar{\boldsymbol{X}}, \boldsymbol{S}),\ U(k; \bar{\boldsymbol{X}}, \boldsymbol{S})\big)$ satisfies

$$P_{\bar{\boldsymbol{X}}, \boldsymbol{S}, X_1, X_2}\left(L(k; \bar{\boldsymbol{X}}, \boldsymbol{S}) \le \frac{X_1}{X_2} \le U(k; \bar{\boldsymbol{X}}, \boldsymbol{S})\right) = \gamma, \tag{1}$$

where k is the prediction factor to be determined subject to the above condition. We note that the condition (1) is based on the joint distribution of $(\bar{\boldsymbol{X}}, \boldsymbol{S}, X_1, X_2)$. It turns out that a PI for X_1/X_2 can be deduced from the one for $X_1 - RX_2$, where R is a constant (to be determined). Let $\bar{Y}$ and S_y^2 denote the sample mean and sample variance, respectively, based on $Y_i = X_{1i} - RX_{2i}$, $i = 1, ..., n$, where we write $\boldsymbol{X}_i = (X_{1i}, X_{2i})'$. A $100\gamma\%$ PI for a future observation $Y^* = X_1^* - RX_2^*$ is given by

$$\bar{Y} \pm kS_y = \bar{X}_1 - R\bar{X}_2 \pm k\sqrt{S_1^2 - 2R\widehat{\rho}S_1S_2 + R^2S_2^2}, \tag{2}$$

where $k = t_{n-1;(1+\gamma)/2}\left(1+\frac{1}{n}\right)^{1/2}$, and $t_{m;\alpha}$ denotes the 100α percentile of the t distribution with degrees of freedom (df) m. It should be clear that the quantity under the square root sign in (2) is simply the estimated variance of $X_1 - RX_2$. Equating the two endpoints of the above interval to zero, and solving the resulting equations for R, we find prediction limits for X_1/X_2. The limits are given by
$(L(k;\bar{X}_1,\bar{X}_2,S_1,S_2), U(k;\bar{X}_1,\bar{X}_2,S_1,S_2))$

$$= \frac{\bar{X}_1\bar{X}_2 - k^2\widehat{\rho}S_1S_2 \mp \sqrt{\left(\bar{X}_1\bar{X}_2 - k^2\widehat{\rho}S_1S_2\right)^2 - \left(\bar{X}_1^2 - k^2S_1^2\right)\left(\bar{X}_2^2 - k^2S_2^2\right)}}{\left(\bar{X}_2^2 - k^2S_2^2\right)}. \tag{3}$$

Clearly, the above limits provide a bona fide prediction interval only when the quantity under the square root sign is positive. As shown in Appendix A, a sufficient condition for this is that the marginal lower prediction limits $\bar{X}_1 - kS_1$ and $\bar{X}_2 - kS_2$ are both positive. We have also shown in Appendix B that the lower prediction limit is more likely positive when both X_1 and X_2 are positive random variables. Furthermore, the PI is equivariant under the scale transformations $X_1 \to c_1X_1$ and $X_2 \to c_2X_2$ and it is invariant when $c_1 = c_2 = c$. We also note that the upper prediction limit for X_2/X_1 is obtained as the reciprocal of the lower prediction limit of X_1/X_2. That is,

$$U(k;\bar{X}_2,\bar{X}_1,S_2,S_1) = 1/L(k;\bar{X}_1,\bar{X}_2,S_1,S_2), \tag{4}$$

where $U(k;\bar{X}_2,\bar{X}_1,S_2,S_1)$ is a $100\gamma\%$ upper prediction limit for X_2/X_1. Noticing that $U(k;\bar{X}_2,\bar{X}_1,S_2,S_1)$ and $U(k;\bar{X}_1,\bar{X}_2,S_1,S_2)$ differ only in the denominator terms, multiplying both sides by $L(k;\bar{X}_1,\bar{X}_2,S_1,S_2)$ and using the result that $(a-b)(a+b) = a^2 - b^2$ the above relation can be readily verified.

2.2. One-Sided Tolerance Limits

A one-sided (p,γ) upper tolerance limit (TL) for the distribution of the ratio X_1/X_2 is a quantity $U(\bar{\boldsymbol{X}},\boldsymbol{S})$ satisfying

$$P_{\bar{\boldsymbol{X}},\boldsymbol{S}}\left\{P\left(\frac{X_1}{X_2} \le U(\bar{\boldsymbol{X}},\boldsymbol{S})\middle|(\bar{\boldsymbol{X}},\boldsymbol{S})\right) \ge p\right\} = \gamma.$$

It is known that a (p,γ) upper TL for any distribution is simply a $100\gamma\%$ upper confidence limit for the pth quantile of the distribution. In view of this, if $R_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the pth quantile of the distribution of X_1/X_2, then $U(\bar{\boldsymbol{X}},\boldsymbol{S})$ also satisfies

$$P_{\bar{\boldsymbol{X}},\boldsymbol{S}}\left\{U(\bar{\boldsymbol{X}},\boldsymbol{S}) \ge R_p(\boldsymbol{\mu},\boldsymbol{\Sigma})\right\} = \gamma.$$

Notice that for a constant R, $X_1 - RX_2 \sim N(\mu_1 - R\mu_2, \sigma_R^2)$, where $\sigma_R^2 = \sigma_1^2 - 2R\rho\sigma_1\sigma_2 + R^2\sigma_2^2$, and so its $100p$ percentile is given by $\mu_1 - R\mu_2 + z_p\sigma_R$, where z_p is the pth quantile of the standard normal distribution. An upper confidence limit for $R_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the pth the percentile of X_1/X_2, can be deduced from an upper CL for $\mu_1 - R\mu_2 + z_p\sigma_R$. Specifically, by equating the upper CL for $\mu_1 - R\mu_2 + z_p\sigma_R$ to zero and then solving the resulting equation for R, we can get an upper confidence limit for $R_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let

$$k_1 = \frac{1}{\sqrt{n}} t_{n-1;\gamma}(z_p\sqrt{n}), \tag{5}$$

where $t_{m;\alpha}(\delta)$ denotes the α quantile of the noncentral t distribution with df $= m$ and the noncentrality parameter δ. It is known that an upper confidence limit for the pth percentile of a normal distribution is given by

$$\text{(sample mean)} + k_1 \times \text{SD}.$$

For example, see Krishnamoorthy and Mathew (2009, Chapter 2). Thus, a $100\gamma\%$ upper confidence limit for $\mu_1 - R\mu_2 + z_p\sigma_D$ is given by

$$\bar{X}_1 - R\bar{X}_2 + k_1\sqrt{S_1^2 - 2R\widehat{\rho}S_1S_2 + R^2S_2^2}. \tag{6}$$

Note that the above upper confidence limit is similar to the prediction limit in (2) except for the factor. Therefore, equating (6) to zero, and solving the resulting equation for R, we will get a $100\gamma\%$ upper confidence limit for $R_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which is the pth quantile of the distribution of X_1/X_2. In other words, such a solution for R is a (p, γ) upper TL for the distribution of X_1/X_2. The tolerance limit is given by $U(k_1; \bar{X}_1, \bar{X}_2, S_1, S_2)$, similar to (3), except that a different factor is used. Similarly, a (p, γ) LTL can be obtained as $L(k_1; \bar{X}_1, \bar{X}_2, S_1, S_2)$. These one-sided TLs also possess some natural properties like the prediction limits. They are bona fide limits provided individual lower TLs $\bar{X}_1 - k_1S_1$ and $\bar{X}_2 - k_1S_2$ are positive and equivariant under scale transformations; see appendices A and B. Furthermore, an upper TL for the distribution of X_2/X_1 can be obtained as reciprocal of the lower TL for X_1/X_2. In other words,

$$U(k_1; \bar{X}_2, \bar{X}_1, S_2, S_1) = 1/L(k_1; \bar{X}_1, \bar{X}_2, S_1, S_2).$$

2.3. Two-Sided Tolerance Intervals

There are two types of two-sided tolerance intervals. A two-sided tolerance interval denoted by $\big(L(k_t; \bar{X}_1, \bar{X}_2, S_1, S_2), U(k_t; \bar{X}_1, \bar{X}_2, S_1, S_2)\big)$ is constructed so that it would include at least a proportion p of the population with confidence γ. That is,

$$P_{\bar{\boldsymbol{X}},\boldsymbol{S}}\left\{P_{X_1,X_2}\left(L(k_t; \bar{\boldsymbol{X}}, \boldsymbol{S}) \le \frac{X_1}{X_2} \le U(k_t; \bar{\boldsymbol{X}}, \boldsymbol{S})\middle| \bar{\boldsymbol{X}}, \boldsymbol{S}\right) \ge p\right\} = \gamma, \tag{7}$$

where k_t is the factor to be determined so as to satisfy the probability requirement in (7). The second type of two-sided tolerance intervals are called equal-tailed tolerance intervals or central tolerance intervals. Such intervals are constructed subject to the condition that the interval would include the lower $100(1-p)/2$ percentile and the upper $100(1-p)/2$ percentile with probability γ. In other words, the interval would include the central $100p\%$ of the distribution. Thus an equal-tailed interval, say $\big(L(k_e; \bar{X}_1, \bar{X}_2, S_1, S_2), U(k_e; \bar{X}_1, \bar{X}_2, S_1, S_2)\big)$, satisfies

$$P_{\bar{\boldsymbol{X}},\boldsymbol{S}}\left\{L(k_e; \bar{X}_1, \bar{X}_2, S_1, S_2) \le R_{\frac{1-p}{2}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ \text{ and } \ R_{\frac{1+p}{2}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \le U(k_e; \bar{X}_1, \bar{X}_2, S_1, S_2)\right\} = \gamma, \tag{8}$$

where $R_\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the 100α percentile of the distribution of X_1/X_2, and k_e is the factor to be determined subject to the above probability requirement.

Since $Y = X_1 - RX_2$ follows a univariate normal distribution when R is a constant, the two-sided tolerance factors k_t and k_e can be obtained from what is already available for the normal distribution; see Krishnamoorthy and Mathew (2009, Chapter 2). For instance, let $\bar{Y} \pm k\sqrt{\widehat{\text{var}}(Y)}$ be a two-sided TI for the distribution of Y, where $k = k_t$ for the TI defined in (7) and $k = k_e$ for the TI defined in (8). Since Y_i's are independent normal random variables, k_t is the factor for computing a TI for a normal distribution, and is the solution of the integral equation

$$\sqrt{\frac{2n}{\pi}} \int_0^\infty P\left(\chi^2_m > \frac{m\chi^2_{1;p}(z^2)}{k_t^2}\right) e^{-\frac{1}{2}nz^2} dz = \gamma, \tag{9}$$

where $m = n-1$; see Krishnamoorthy and Mathew (2009, Section 2.3). The two-sided TI is given in (3) with k replaced by k_t.

Similarly, the factor k_e for constructing equal-tailed TI of the form $\bar{Y} \pm k_e\sqrt{\widehat{\text{var}}(Y)}$ is the solution of the integral equation

$$\frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} \int_{\frac{m\delta^2}{k_e^2 n}}^\infty \left(2\Phi\left(-\delta + \frac{k_e\sqrt{nx}}{\sqrt{m}}\right) - 1\right) e^{-x/2} x^{\frac{m}{2}-1} dx = \gamma, \tag{10}$$

where $\Phi(x)$ denotes the standard normal distribution function. Details are once again given in Krishnamoorthy and Mathew (2009, Section 2.3). The equal-tailed TI is given in (3) with k replaced by k_e.

A sufficient condition for the existence of these two-sided TIs is that

$$\bar{X}_i - \text{factor} \times S_i > 0, \ \text{ for } \ i = 1, 2,$$

where the factor is k_t or k_e. In Appendix B, we provided some numerical evidence that the above conditions hold with high probability. The R package by Young (2020) or the PC calculator *StatCalc* that accompanies the book by Krishnamoorthy (2016) can be used to compute the factors k_t and k_e.

3. Examples

We shall now present two examples in order to illustrate our methodology for computing prediction intervals and tolerance intervals for the ratio X_1/X_2. The first example is discussed in Zhang et al. (2010), and both the examples are given in Flouri et al. (2017).

Example 1. This application described in Zhang et al. (2010) is in the context of the manufacture of the influenza vaccine FluMist, and is an investigation of retroviral contamination in the raw materials used in its manufacture. The reverse transcriptase (RT) assay is used to detect contamination, and the presence of the enzyme RNA directed DNA polymerase is indicative of the presence of the retrovirus. Additionally, a large radioactivity count in the RT assay is taken as evidence for a large amount of the enzyme in the sample. A sample is classified as negative by comparing the radioactivity count from the sample with that from a negative control. Operationally, this amounts to verifying if the ratio of the two radioactivity counts is below a threshold. Thus the two radioactivity counts is the random variable of interest, and historical in-control data supports the bivariate normality assumption. For more details, we refer to the article by Zhang et al. (2010).

The bivariate sample under consideration consists of $n = 45$ pairs of radioactivity counts in the negative controls and in the in-control RT assays. The Shapiro-Wilk normality test by Zhang et al. (2010) revealed that the data were distributed as bivariate normal. Furthermore, the sample data gave the following statistics (sample means $\bar{x}_1$ and $\bar{x}_2$, sample variances s_1^2 and s_2^2, and sample correlation coefficient $\hat{\rho}$):

$$\bar{x}_1 = 38.1, \quad s_1^2 = 56.3, \quad \bar{x}_2 = 38.9, \quad s_2^2 = 35.1, \quad \widehat{\rho} = 0.81.$$

Let X_1 and X_2 denote the radioactivity counts in the negative controls and in the in-control RT assays, respectively. The factor for computing a 95% two-sided prediction interval for the ratio X_1/X_2 is obtained as $t_{n-1;1-\alpha/2}\sqrt{1+1/n} = t_{44;0.975}\sqrt{1+1/45} = 2.038$ and the 95% PI (3) is (0.731, 1.218). Similarly, we computed the 99% PI as (0.631, 1.309).

In order to construct a $(.90, .95)$ upper tolerance limit for X_1/X_2, we first find the factor as $t_{44;.95}(z_{.9}\sqrt{45})/\sqrt{45} = 2.0924$. We computed the upper tolerance limits and have reported them in Table 1 along with those based on the two generalized variable methods by Zhang et al. (2010); in the table, these are denoted by GVI and GVII. Compared to the factors based on the generalized variable approach, the exact approach has produced factors that are somewhat smaller, and hence these are to be preferred. However the difference between the tolerance factors based on the two approaches is somewhat insignificant. We recall that the Zhang et al. (2010) approach can produce only one-sided tolerance limits.

The factors for computing (p, .95) two-sided tolerance interval are 2.4116 and 3.1680 when $p = 0.95$ and 0.99, respectively. The factors for computing a $(p, .95)$ equal-tailed tolerance interval are 2.5595 and 3.3005 when $p = 0.95$ and 0.99, respectively.

The corresponding two-sided intervals are also given in Table 1. As expected, the equal-tailed tolerance interval is wider compared to the tolerance interval that does not impose this requirement.

Table 1: (p, .95) upper TLs and tolerance intervals for the distribution of X_1/X_2 for Example 1

			Exact method		
p	GV I	GV II	One-sided	Two-sided	Equal-tailed
0.95	1.233	1.230	1.224	(0.678, 1.266)	(0.656, 1.286)
0.99	1.346	1.343	1.334	(0.555, 1.376)	(0.531, 1.397)

Example 2. Ratio parameters and ratios of random variables are encountered in applications dealing with cost-effectiveness analysis. Our focus here is on a specific application in cost-effectiveness analysis where the problem of interest is the computation of tolerance intervals for a ratio random variable applying our methodology. Thus we shall not give a discussion of the area of cost-effectiveness analysis and the relevant statistical methodologies; for details we refer to the book by Willan and Briggs (2006). Our example involves the comparison of two pharmacological agents (a *test drug* and a *reference drug*), and is adapted from Gardiner et al. (2000). In this comparison trial, each of the test drug and the reference drug was administered to 150 patients, and the effectiveness of the drug was assessed using Quality-Adjusted-Life-Years (QALYs). Furthermore, the cost of the treatment was in US dollars. This application first appeared in Sacristan et al. (1995), and the data was later analyzed in Gardiner et al. (2000) and Flouri et al. (2017). Bivariate normality was assumed by these authors for the cost (C) and effectiveness (E) and the problem of interest is the computation of tolerance limits for C/E for the test drug and for the reference drug. As noted by Flouri et al. (2017), the original article by Sacristan et al. (1995) does not report the correlation between the cost and effectiveness. In the absence of correlations, Gardiner et al. (2000) assumed the value 0.7 for the correlation coefficients $\widehat{\rho}_i$ in both groups. In their work, Flouri et al. (2017) used the same value for the correlation coefficients in both groups and illustrated their methods of constructing TIs for C/E. The means and standard deviations for the two groups are as follows:

Table 2: Summary statistics for Example 2

	Sample size	Mean cost	Mean effectiveness	SD cost	SD effectiveness	$\widehat{\rho}$
Test drug	150	200000	8	78400	2.1	0.7
Reference drug	150	80000	5	27343	2	

The factor for computing 90% one-sided PIs for the ratio X_1/X_2 is 1.292. For the test drug, the 90% one-sided lower and upper prediction limits are 15231.0 and 34474.4, respectively. For the reference drug, the 90% one-sided lower and upper prediction limits are 10907.2 and 25779.1, respectively. These one-sided prediction limits are computed using (3).

The one-sided and two-sided tolerance limits for the same example are given in Table 3 and Table 4, respectively. Between the test drug and the reference drug, we notice significant differences among the one-sided tolerance limits, and also among the two-sided tolerance intervals. In fact we note larger tolerance limits for the distribution of C/E for the test drug, compared to those for the standard drug in the case of both one-sided tolerance limits as well as two-sided tolerance intervals. These differences indicate that the two drugs are not similar. This is also the conclusion in Flouri et al. (2017).

Table 3: (0.90, 0.95) one-sided tolerance limits for the ratio C/E for the test drug and for the reference drug for Example 2

Method	Bootstrap		Tolerance	Exact method	
	LTL	UTL	factor	LTL	UTL
Test drug	14262.37	35639.61	1.478	13566.16	36032.33
Reference drug	10058.54	27822.01	1.478	10216.08	28690.79

Table 4: (.90,.95) tolerance intervals for the ratio C/E for the test drug and for the reference drug for Example 2

		Exact method	
Method	Bootstrap TI	Two-sided	Equal-tailed
Test drug	(9441.55, 40017.71)	(10084.12, 39239.81)	(9114.08, 40121.31)
Reference drug	(8505.69, 39680.02)	(8849.58, 37854.98)	(8481.67, 41698.69)

4. Discussion

Interval estimation problems involving ratio of parameters and ratio of random variables present challenges when it comes to developing the necessary methodology that guarantees the desired statistical properties; for example, satisfying the coverage probability requirements and having solutions that are indeed finite intervals. While the interval estimation of the ratio of normal means is a well investigated problem, only very limited literature is available for computing prediction intervals and tolerance intervals for the ratio random variable in a bivariate normal distribution. This article is on the development of such intervals. Some attractive features of our intervals are that they are exact, and are very easy to compute since they have explicit analytic expressions. It is hoped that they will be used in applications that call for the computation of prediction limits and tolerance limits for ratio random variables; for example, in statistical process control problems that are ratio based. However, we have not taken up such applications in the present work.

Acknowledgements

The authors are grateful to two reviewers for providing useful comments and suggestions.

References

[1] Alqurashi, M. (2021). *Some Contributions to Tolerance Intervals and Statistical Process Control*. Doctoral dissertation submitted to the University of Alabama, Tuscaloosa, Alabama.

[2] Celano, G., Castagliola, P., Faraz, A. and Fichera, S. (2014) Statistical performance of a control chart for individual observations monitoring the ratio of two normal variables. *Quality and Reliability Engineering International* 30:1361–1377. DOI: 10.1002/qre.1558.

[3] Celano, G. and Castagliola, P. (2016). Design of a phase II control chart for monitoring the ratio of two normal variables. *Quality and Reliability Engineering International* 32:291–308. DOI: 10.1002/qre.1748

[4] Flouri, M., Zhai, S., Mathew, T., Bebu, I. (2017). Tolerance limits and tolerance intervals for ratios of normal random variables using a bootstrap calibration. *Biometrical Journal* 59:550–566.

[5] Gardiner, J. C., Huebner, M., Jetton, J. and Bradley, C. J. (2000). Power and sample size assessments for test of hypotheses on cost-effectiveness ratios. *Health Economics*, 9, 227–234.

[6] Hall, I. J. and Sampson, C. B. (1973). Tolerance limits for the distribution of the product and quotient of normal variates. *Biometrics* 29:109–119.

[7] Hamada, M. (2003). Tolerance Interval Control Limits for $\bar{X}$, R, and S Charts. *Quality Engineering*, 15, 471-487.

[8] Krishnamoorthy, K. (2016). *Handbook of Statistical Distributions with Applications*, 2nd edition. Chapman & Hall/CRC, Boca Raton, FL.

[9] Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications and Computation*. Hoboken, NJ: John Wiley.

[10] Sacristan, J. A., Day, S. J., Navarro, O., Ramos, J. and Hernandez, J. M. (1995). Use of confidence intervals and sample size calculations in health economic studies. *The Annals of Pharmacotherapy*, 29, 719-725.

[11] Tran, K. D., Khaliq, Q-U-A., Nadi, A. A., Nguyen, T. H. and Tran, K. P. (2021). One-sided Shewhart control charts for monitoring the ratio of two normal variables in short production runs, Journal of Manufacturing Processes, 69, 273-289.

[12] Willan, A. R. and Briggs, A. H. (2006). *Statistical Analysis of Cost-Effectiveness Data*. Wiley, New York.

[13] Young, D. S. (2020). Package "tolerance." Statistical Tolerance Intervals and Regions. R package version 2.0.0. https://cran.r-project.org/web/packages/tolerance/
tolerance.pdf

[14] Zhang, L., Mathew, T., Yang, H., Krishnamoorthy, K. and Cho, I. (2010). Tolerance limits for a ratio of normal random Variables. Journal of Biopharmaceutical Statistics 20:172–184.

Appendix A

To show that the term $T = \left(\bar{X}_1\bar{X}_2 - k^2\widehat{\rho}S_1S_2\right)^2 - \left(\bar{X}_1^2 - k^2S_1^2\right)\left(\bar{X}_2^2 - k^2S_2^2\right)$ under the radical sign in (3) is positive, let $a_i = \bar{X}_i/S_i$, $i = 1, 2$. It is easy to see that the above term is positive if

$$(a_1a_2 - k^2\widehat{\rho})^2 - (a_1^2 - k^2)(a_2^2 - k^2) > 0.$$

After simplification, we see that the above condition simplifies to $(a_1^2 + a_2^2) + k^2\widehat{\rho}^2 - k^2 - 2\widehat{\rho}a_1a_2 > 0$. Using the result $a_1^2 + a_2^2 > 2a_1a_2$, it is sufficient to show that

$$2a_1a_2(1 - \widehat{\rho}) - k^2(1 + \widehat{\rho})(1 - \widehat{\rho}) > 0.$$

Noting that $\widehat{\rho} < 1$, a sufficient condition for the above inequality to hold is $a_1 > k$ and $a_2 > k$. That is, $\bar{X}_1 - k\sqrt{S_{11}} > 0$ and $\bar{X}_2 - k\sqrt{S_{22}} > 0$. Also, under these conditions, the denominator term $\bar{X}_2^2 - k^2S_2^2$ in (3) is also positive because $\bar{X}_2^2 - k^2S_2^2 = (\bar{X}_2 - kS_2)(\bar{X}_2 + kS_2)$.

Appendix B

We now show that the $\bar{X}_1 - kS_1$ is positive with high probability. We first note that $\bar{X}_1 - kS_1 > 0$ if and only if

$$\frac{\bar{X}_1}{S_1} > k = (1 + 1/n)^{1/2}t_{n-1;\frac{1+\gamma}{2}}.$$

It is easy to show that

$$\frac{\bar{X}_1}{S_1} \sim \frac{1}{\sqrt{n}}t_{n-1}\left(n_1\frac{\mu_1}{\sigma_1}\right).$$

Since we postulate normal model for a positive random variable X_1, the population parameters are expected to satisfy $\mu_1 - 3.3\sigma_1 > 0$, or equivalently, $\mu_1/\sigma_1 > 3.3$. Then

$$\begin{aligned} P(\bar{X}_1/S_1 > k) &= P\left(t_{n-1}(\sqrt{n}\mu_1/\sigma_1) > \sqrt{n+1}t_{n-1;\frac{1+\gamma}{2}}\right) \\ &> P\left(t_{n-1}(3.3 \times \sqrt{n}) > \sqrt{n+1}t_{n-1;\frac{1+\gamma}{2}}\right) \\ &= P(n,\gamma), \text{ say.} \end{aligned}$$

To get the above inequality, we used that fact that the noncentral t distribution is stochastically increasing in the noncentrality parameter. The above probability is very close to 1 for many practical values of n and γ. For example, $P(n, .95) = 0.987, .997$ and 0.9999 when $n = 15, 20$ and 30, respectively.

To show that the one-sided tolerance limits are also more likely positive, we need to assess $P(\bar{X}_1/S_1 > k_1)$, where $k_1 = \frac{1}{\sqrt{n}}t_{n-1;\gamma}(z_p\sqrt{n})$; see Eqn (5). As argued earlier for the prediction limits, this probability

$$\begin{aligned} P(\bar{X}_1/S_1 > k_1) &= P\left(\bar{X}_1/S_1 > \frac{1}{\sqrt{n}}t_{n-1;\gamma}(z_p\sqrt{n})\right) \\ &> P\left(t_{n-1}(3.3 \times \sqrt{n}) > t_{n-1;\gamma}(z_p\sqrt{n})\right) \\ &= P_1(n,p,\gamma), \text{say.} \end{aligned}$$

The above probability is also close to 1 for many reasonable cases. For example, $P_1(n, .90, .95) = .996$ and 0.9998 when $n = 15$ and 20, respectively. $P_1(n, .95, .95) = 0.920, 0.979$ and 0.995 when $n = 15, 20$ and 25, respectively.

Recall that the two-sided TI is the interval in (3) with k replaced by k_t, where k_t is determined by Eqn (9). We computed $P_2(n,p,\gamma) = P\left(t_{n-1}(3.3 \times \sqrt{n}) > \sqrt{n}k_t\right) < P(\bar{X}_1/S_1 > k_t)$ along the lines of the preceding paragraphs as follows. $P_2(n, .90, .95) = 0.941, 0.988$ and 0.998 when $n = 15, 20$ and 30, respectively. $P_2(n, .95, .95) = 0.941, 0.963$, 0.985 and 0.994 when $n = 15, 20, 30$ and 40, respectively. The equal-tailed TI in (3) with k replaced by k_e, where k_e is determined by Eqn (10). We computed $P_3(n,p,\gamma) = P\left(t_{n-1}(3.3 \times \sqrt{n}) > \sqrt{n}k_e\right) < P(\bar{X}_1/S_1 > k_e)$ for some values of (n,p,γ) as follows. $P_3(n, .90, .95) = 0.841, 0.948$ and 0.996 when $n = 15, 20$ and 30, respectively. $P_3(n, .95, .95) = 0.594, 0.755, 0.927$ and 0.982 when $n = 15, 20, 30$ and 40, respectively.