Statistica Sinica **14**(2004), 217-229

A TOLERANCE INTERVAL FOR THE NORMAL DISTRIBUTION WITH SEVERAL VARIANCE COMPONENTS

C. T. Liao and H. K. Iyer

National Taiwan University and Colorado State University

Abstract: A tolerance interval procedure is derived from the concept of generalized pivotal quantities usually used to obtain confidence intervals in situations where standard procedures do not lead to useful solutions. We apply the generalized confidence intervals approach and propose a two-sided tolerance interval for the distribution $N(\theta, \sum_{i=1}^{q} h_i\sigma_i^2)$ based on mutually independent statistics $\hat{\theta}$, $S_1^2, \ldots, S_q^2$, where $\hat{\theta}$ is distributed as $N(\theta, \sum_{i=1}^{q} c_i\sigma_i^2)$, h_i and c_i are known constants, and $n_iS_i^2/\sigma_i^2$ are independent chi-squared random variables with n_i df, for $i = 1, \ldots, q$. Some practical examples are given to illustrate the applications of the proposed procedure. A simulation study is conducted to evaluate its frequentist coverage probability. The results indicate that the proposed method may be recommended for use in practical applications. The procedure provided in this paper can be applied to tolerance interval questions arising in arbitrary normal balanced mixed linear model situations.

Key words and phrases: Chi-squared approximation, generalized P-values, generalized confidence intervals, linear models, variance components.

1. Introduction

Let F denote the cumulative distribution of a random variable. An interval $[L(\boldsymbol{Y}), U(\boldsymbol{Y})]$ (or $[L, U]$, for simplicity), based on the data vector $\boldsymbol{Y}$, is called a two-sided β-content, γ-confidence tolerance interval (or (β, γ)-tolerance interval, for short) for F if

$$Pr[F(U(\boldsymbol{Y})) - F(L(\boldsymbol{Y})) \geq \beta] = \gamma.$$

Thus, we can state with confidence coefficient γ that at least a proportion β of the population modeled by F is contained in the interval $[L, U]$.

Two-sided tolerance intervals are widely used in industrial applications where manufactured parts have to meet certain specifications. If the manufacturing process is capable, then a high proportion of the items manufactured will meet the specifications. Two-sided tolerance intervals give us L and U such that we can claim, with a specified degree of confidence γ, that a specified proportion β or more of the manufactured items lie between L and U.

The problem for computing a tolerance interval for the simple case in which F is the normal distribution with unknown mean μ and unknown variance σ^2 has been extensively studied; see, for example, Wald and Wolfowitz (1946), Howe (1969) and Odeh and Owen (1980). For more complex situations, only scattered results are available. The problem of setting a two-sided tolerance interval for the distribution $N(\theta, \sigma_1^2 - \sigma_2^2)$ based on mutually independent statistics $\hat{\theta}$, S_1^2 and S_2^2, where $\hat{\theta} \sim N(\theta, c\sigma_1^2)$, c is a known constant, $n_1 S_1^2/\sigma_1^2$ and $n_2 S_2^2/\sigma_2^2$ are chi-squared random variables with n_1 and n_2 degrees of freedom (df), respectively, was considered in Wang and Iyer (1994). Also some practical examples were given in their paper to illustrate the applications of their proposed procedure. Brown, Iyer and Wang (1997) applied their results to evaluate the bioequivalence of two formulations of a drug using various cross-over designs for data collection. Liao and Iyer (2001) extended the results of Wang and Iyer (1994) and proposed tolerance intervals for the distribution $N(\theta, \sigma_1^2 - \sigma_2^2)$ for the case where the distribution of the statistic $\hat{\theta}$ is $N(\theta, \sum_{i=1}^{q} c_i \sigma_i^2)$, with $q \geq 1$ and the c_i known constants. Their study was motivated by an actual application involving the assessment of the quality of a type of glucose monitoring meter. We will revisit this problem in Section 4.

In this paper, we generalize the problem as follows. We seek a two-sided tolerance interval for a random variable W which has a $N(\theta, \sum_{i=1}^{q} h_i \sigma_i^2)$ distribution. Suppose mutually independent statistics $\hat{\theta}$, $S_1^2, \ldots, S_q^2$ are available, where $\hat{\theta}$ is normally distributed with mean θ and variance $\sum_{i=1}^{q} c_i \sigma_i^2$, h_i and c_i are known constants, and $n_i S_i^2/\sigma_i^2$ are independent chi-squared random variables with n_i df, for $i = 1, \ldots, q$. The solution we propose is based on the concept of *generalized confidence intervals*, see Weerahandi (1993, 1995), and is different from the derivation based on the method given in Wang and Iyer (1994).

The simplest instance of the general problem stated above occurs in the context of a one-way, balanced, random effects model. Consider the model

$$Y_{ij} = \mu + A_i + e_{ij},$$

where Y_{ij} is the jth repeat measurement ($1 \leq j \leq n$) on item i ($1 \leq i \leq a$) selected randomly from a population of items. It is assumed that $A_i \overset{i.i.d.}{\sim} N(0, \sigma_A^2)$, $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$, and all random variables are jointly independent. The true value associated with item i is $\mu + A_i$ and the mean of the entire population of items is μ. In industrial applications one may be interested in a tolerance interval for the distribution of the true values $\mu + A_i$ of the items, in which case one needs a tolerance interval for the distribution $N(\mu, \sigma_A^2)$; or one may be interested in a tolerance interval for the distribution of the measured values, which is $N(\mu, \sigma_A^2 + \sigma_e^2)$. Both of these problems are special cases of the general problem stated above. First consider the distribution of true values.

Write $\bar{Y} = (1/an)\sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}$, $\bar{Y}_i = 1/n\sum_{j=1}^{n} Y_{ij}$, $S_1^2 = n\sum_{i=1}^{a}(\bar{Y}_i - \bar{Y})^2$, $S_2^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2$, $\sigma_1^2 = \sigma_e^2 + n\sigma_A^2$, $\sigma_2^2 = \sigma_e^2$, $\theta = \mu$ and $\hat{\theta} = \bar{Y}$. We then see that we are interested in a tolerance interval for the distribution $N(\theta, h_1\sigma_1^2 + h_2\sigma_2^2)$ with $h_1 = 1/n$ and $h_2 = -1/n$. Furthermore, jointly independent statistics $\hat{\theta}$, S_1^2 and S_2^2 are available with $\hat{\theta} \sim N(\theta, c_1\sigma_1^2 + c_2\sigma_2^2)$ where $c_1 = 1/an$, $c_2 = 0$, $(a-1)S_1^2/\sigma_1^2 \sim \chi_{a-1}^2$, and $a(n-1)S_2^2/\sigma_2^2 \sim \chi_{a(n-1)}^2$. Thus, the premise of our problem is satisfied. Next, consider the distribution of measured values, namely, $N(\mu, \sigma_A^2 + \sigma_e^2)$. Note that $\sigma_A^2 + \sigma_e^2 = h_1\sigma_1^2 + h_2\sigma_2^2$, where $h_1 = 1/n$ and $h_2 = (n-1)/n$. Let $\hat{\theta}$, S_1^2 and S_2^2 be as before. Again, the premise of our general problem is satisfied.

In the next section, we review the concept of *generalized pivotal quantities* and generalized confidence intervals. An approach to computing the tolerance intervals of interest based on the concept of generalized confidence intervals is proposed in Section 3. In Section 4, the glucose meters problem described in Liao and Iyer (2001) and a bioequivalence experiment discussed in Brown, Iyer and Wang (1997) are used to illustrate the application of our proposed procedure. A simulation study is conducted to evaluate the performance of the proposed procedure and the results of this study are discussed in Section 5.

2. Generalized Confidence Intervals

Tsui and Weerahandi (1989) introduced *generalized P-Values* and *generalized test variables* and demonstrated that useful hypothesis tests could be derived for situations where standard methods fail to yield satisfactory solutions. Weerahandi (1993) extended this idea and introduced generalized pivotal quantities and generalized confidence intervals. He demonstrated that generalized confidence intervals provided useful solutions to interval estimation problems where satisfactory solutions were unavailable. Frequentist coverage probabilities of generalized confidence intervals are, typically, functions of nuisance parameters; so generalized confidence intervals do not in general have exact coverage probabilities. Empirical evidence based on examination of large number of situations suggests that coverage probabilities of generalized confidence intervals are, as a rule, sufficiently close to the nominal levels over the entire ranges of nuisance parameters, that the resulting intervals are satisfactory in practical applications.

Generalized tests and generalized confidence intervals are now available in the literature for many applications. The following is a partial list of publications in this area – Weerahandi (1991), Zhou and Mathew (1994), Weerahandi (1995), Khuri, Mathew and Sinha (1998) and Chang and Huang (2000). Generalized P-values and generalized confidence intervals are applicable in a variety of situations, not just normal mixed linear models. However, most of the published implementations of this approach are to linear model problems.

In the following paragraphs we describe the construction of generalized confidence intervals for θ and τ mentioned in the problem statement. These will be used in Section 3 to obtain a two-sided tolerance interval for the distribution $N(\theta, \sum_{i=1}^q h_i\sigma_i^2)$. We follow the convention used by Tsui and Weerahandi (1989) of representing with upper case letters the observable random variables and with lower case letters their realized values.

Let $\tau^2 = \sum_{i=1}^q h_i\sigma_i^2$ and $\sigma^2 = \sum_{i=1}^q c_i\sigma_i^2$. We first give generalized pivotal quantities for θ and τ. Let T represent the observable random variable $\hat{\theta}$ and t denote its observed value, $Z = (T-\theta)/\sigma$ and, for $i = 1,\ldots,q$, $U_i = n_iS_i^2/\sigma_i^2$. Then $Z \sim N(0,1)$ and, for $i = 1,\ldots,q$, $U_i \sim \chi^2(n_i)$. We define

$$R_\theta = t - Z\sqrt{\sum_{i=1}^q \frac{c_i n_i s_i^2}{U_i}} = t - \left(\frac{T-\theta}{\sigma}\right)\sqrt{\sum_{i=1}^q \frac{c_i\sigma_i^2 s_i^2}{S_i^2}}. \tag{2.1}$$

Likewise, let us define

$$R_\tau = \sqrt{\sum_{i=1}^q \frac{h_i n_i s_i^2}{U_i}} = \sqrt{\sum_{i=1}^q \frac{h_i\sigma_i^2 s_i^2}{S_i^2}}. \tag{2.2}$$

From (2.1) and (2.2), it follows that R_θ and R_τ have distributions that are free of model parameters. Furthermore, we see that, when the observed values t and s_i^2 are substituted for the observable random variables T and S_i^2, $i = 1,\ldots,q$, R_θ and R_τ become $r_\theta = \theta$ and $r_\tau = \tau$. Therefore, R_θ and R_τ satisfy the requirements for being generalized pivotal quantities for θ and τ. See Weerahandi (1993) for the relevant definitions and details. Hence, a two-sided α confidence interval for θ and an upper α confidence bound for τ are $\{\theta | R_{\theta,(1-\alpha)/2} \le r_\theta \le R_{\theta,(1+\alpha)/2}\}$ and $\{\tau | r_\tau \le R_{\tau,\alpha}\}$, respectively. The required percentiles $R_{\theta,\alpha}$ and $R_{\tau,\alpha}$ may be determined by the following Monte-Carlo algorithm.

Step 1: Let M be a large positive integer, say 100,000. For i equal to 1 through M, carry out the following steps 2 and 3.

Step 2: Generate a standard normal random deviate Z_i and chi-squared random deviates $U_{1,i}$, …, $U_{q,i}$ with $n_1,\ldots,n_q$ degrees of freedom, respectively. The random deviates are required to be jointly independent.

Step 3: Calculate $R_{\theta,i}$ and $R_{\tau,i}$ using the expressions (2.1) and (2.2) for R_θ and R_τ, respectively.

Let $\hat{\theta}_{(1-\gamma)/2}$ and $\hat{\theta}_{(1+\gamma)/2}$ be the $(1-\gamma)/2$ and $(1+\gamma)/2$ sample percentiles of the collection of values $R_{\theta,1},\ldots,R_{\theta,M}$. Then $[\hat{\theta}_{(1-\gamma)/2}, \hat{\theta}_{(1+\gamma)/2}]$ may be used as a two-sided generalized confidence interval for θ with nominal confidence coefficient γ. Similarly, we may use $\hat{\tau}_\gamma$ as an upper generalized confidence bound for τ with nominal confidence coefficient γ.

3. A Generalized Tolerance Interval for $N(\theta, \sum_{i=1}^q h_i\sigma_i^2)$

Recall that we seek a (β, γ)-tolerance interval for a random variable $W \sim N(\theta, \tau^2)$, where $\tau^2 = \sum_{i=1}^q h_i\sigma_i^2$. We have $\hat{\theta} \sim N(\theta, \sigma^2)$ with σ^2 given by $\sigma^2 = \sum_{i=1}^q c_i\sigma_i^2$, where h_i and c_i are known constants. Furthermore, mutually independent statistics $S_1^2, \ldots, S_q^2$ are available such that they are independent of $\hat{\theta}$ and $n_iS_i^2/\sigma_i^2 \sim \chi^2_{(n_i)}$, for $i = 1, \ldots, q$.

We need to find a margin of error statistic $D = D(S_1^2, \ldots, S_q^2)$ such that

$$Pr_{\hat{\theta}, S_1^2, \ldots, S_q^2}\{Pr_W[\hat{\theta} - D \le W \le \hat{\theta} + D] \ge \beta\} = \gamma.$$

Define $Q(\beta, D) = Pr_{\hat{\theta}, S_1^2, \ldots, S_q^2}\{Pr_W[\hat{\theta} - D \le W \le \hat{\theta} + D] \ge \beta\}$. Also let $Z = (\hat{\theta} - \theta)/\sigma$ which is a standard normal random variable. We thus have $Q(\beta, D) = Pr_{Z, S_1^2, \ldots, S_q^2}\{[\Phi(Z(\sigma/\tau) + (D/\tau)) - \Phi(Z(\sigma/\tau) - (D/\tau))] \ge \beta\}$, where $\Phi(\cdot)$ is the standard normal distribution function. Therefore, the problem is to find D such that $Q(\beta, D) = \gamma$. As in the work of Wald and Wolfowitz (1946), we can first compute $k = k(z, \sigma, \tau, \beta)$ that satisfies

$$\Phi(z\frac{\sigma}{\tau} + k) - \Phi(z\frac{\sigma}{\tau} - k) = \beta. \tag{3.1}$$

Then $Q(\beta, D|Z = z) = Pr_{S_1^2, \ldots, S_q^2}\{(D/\tau) \ge k\} = Pr_{S_1^2, \ldots, S_q^2}\{(D/k) \ge \tau\} = \gamma$.

Obviously, D/k is an upper γ confidence bound for τ. So the value of D may be estimated by the $k\ \hat{\tau}_\gamma$, where $\hat{\tau}_\gamma$ is given in the previous subsection. Finally, the value of k must be computed from (3.1) which satisfies $E_Z[\Phi(Z(\sigma/\tau) + k) - \Phi(Z(\sigma/\tau) - k)] = \beta$. Using the Wald-Wolfowitz (1946) approximation which states that k is closely approximated by the root of the nonlinear equation $\Phi(\phi^{-1} + k) - \Phi(\phi^{-1} - k) = \beta$, where $\phi = \tau/\sigma$. Another approximation is given in Howe (1969), which uses

$$k = \sqrt{1 + \frac{1}{\phi^2}}z_{(1+\beta)/2}. \tag{3.2}$$

The parameter $\phi^2 = \tau^2/\sigma^2$ can be estimated by $\hat{\phi}^2 = \hat{\tau}^2/\hat{\sigma}^2$, where $\hat{\tau}^2 = \sum_{i=1}^q h_is_i^2$ and $\hat{\sigma}^2 = \sum_{i=1}^q c_is_i^2$. There is a possibility that the estimated value $\hat{\tau}$ is a negative number for some situations. If this happens, we replace $\hat{\tau}$ by $\hat{\tau}_\gamma$ in $\hat{\phi}$. The value of $\hat{\tau}_\gamma$ is rarely negative in our experience based on the simulation study reported in Section 5. In case $\hat{\tau}_\gamma \le 0$, we may use the two-sided γ generalized confidence interval $[\hat{\theta}_{(1-\gamma)/2}, \hat{\theta}_{(1+\gamma)/2}]$ as the (β, γ)-tolerance interval.

Based on the above discussion, we propose the following as the (β, γ)-tolerance interval for the distribution $N(\theta, \tau^2)$: (i) When $\hat{\tau}_\gamma > 0$, the required tolerance interval is computed as

$$\hat{\theta} \pm \hat{k}\hat{\tau}_\gamma, \tag{3.3}$$

where $\hat{k}$ is obtained from (3.2) in which ϕ is estimated by $\hat{\phi} = \hat{\tau}/\hat{\sigma}$ if $\hat{\tau} > 0$, and by $\hat{\phi} = \hat{\tau}_\gamma/\hat{\sigma}$ otherwise; (ii) When $\hat{\tau}_\gamma \leq 0$, the required tolerance interval is taken to be the two-sided γ-level generalized confidence interval $[\hat{\theta}_{(1-\gamma)/2}, \hat{\theta}_{(1+\gamma)/2}]$.

4. Illustrative Examples

The following practical examples are given to illustrate the proposed procedure.

Example 4.1. The glucose monitoring meter experiment

Liao and Iyer (2001) described a gage study for comparing the quality between a newly developed glucose monitoring meter for in-home use by patients with diabetes (called test meter) and a marked one (called reference meter). The details regarding the experiment are given in their paper. Let X denote a measurement using a test meter and Y denote a measurement using a reference meter. Then X is modeled as $X_{ijkl} = \mu_T + M_i + B_j + L_k + e_{ijkl}$, for $i = 1, \ldots, m$, $j = 1, \ldots, B$, $k = 1, \ldots, L$ and $l = 1, \ldots, E$, where μ_T denotes the expected reading when using a test meter, M_i the effect of test meter i, B_j the effect of the jth blood sample, L_k the effect of the kth strip-lot and e_{ijkl} measurement error. Likewise, $Y_{ijkl} = \mu_R + M'_i + B_j + L_k + e'_{ijkl}$, for $i = 1, \ldots, n$, $j = 1, \ldots, B$, $k = 1, \ldots, L$ and $l = 1, \ldots, E$, where μ_R denotes the expected reading when using a reference meter, M'_i the effect of reference meter i, B_j the effect of the jth blood sample, L_k the effect of the kth strip-lot and e'_{ijkl} measurement error. The effects $M_i, M'_i, B_j, L_k, e_{ijkl}$ and e'_{ijkl} are random effects, normally distributed with zero mean and standard deviations equal to σ_T, σ_R, σ_B, σ_L, σ_e and σ_e, respectively (the variances of e_{ijkl} and e'_{ijkl} are assumed to be equal).

The theoretical mean for the ith test meter when using blood sample j and strip-lot k is equal to $\mu_T + M_i + B_j + L_k$. The theoretical mean reading, averaging over *all* reference meters, for the same blood sample and strip-lot is equal to $\mu_R + B_j + L_k$. This theoretical mean reading is used as the reference value against which the readings from individual test meters will be compared to assess their accuracy. The deviation of the reading obtained using a single test meter from the mean over all reference meters is thus equal to $D_i = \mu_T - \mu_R + M_i$. It is the distribution of the D_i's that is of interest. For the quality control objective, a ($\beta = 0.95$, $\gamma = 0.90$)-tolerance interval is required for the distribution of D_i. A batch of test meters is deemed to have met the quality requirements if the tolerance interval falls completely into the interval $[-5, 5]$.

We now apply the tolerance interval given in Section 3 to this problem. Let

$$\bar{X} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{B}\sum_{k=1}^{L}\sum_{l=1}^{E} X_{ijkl}}{mBLE},$$

$$MS_T = \frac{BLE\sum_{i=1}^{m}(\bar{X}_{i...} - \bar{X})^2}{m-1},$$

$$MSE_T = \frac{\sum_{i=1}^{m}\sum_{j=1}^{B}\sum_{k=1}^{L}\sum_{l=1}^{E}(X_{ijkl} - \bar{X}_{i...} - \bar{X}_{.j..} - \bar{X}_{..k.} + 2\bar{X})^2}{mBLE - m - B - L + 2}.$$

Similarly, let $\bar{Y}$, MS_R and MSE_R denote the corresponding sample mean, mean square for the reference meter effect and error mean square for the model fitted to the reference meters data. Then the statistics $\bar{X}$, $\bar{Y}$, MS_T, MSE_T, MS_R and MSE_R are mutually independent. Let $\sigma_1^2 = \sigma_T^2 + \sigma_e^2/k_0$, $n_1 = m-1$, $S_1^2 = MS_T/k_0$; and $\sigma_2^2 = \sigma_R^2 + \sigma_e^2/k_0$, $n_2 = n-1$, $S_2^2 = MS_R/k_0$, where $k_0 = BLE$. It follows that $n_1S_1^2/\sigma_1^2 \sim \chi^2_{(n_1)}$, $n_2S_2^2/\sigma_2^2 \sim \chi^2_{(n_2)}$. Also $v_1MSE_T/\sigma_e^2 \sim \chi^2_{(v_1)}$ and $v_2MSE_R/\sigma_e^2 \sim \chi^2_{(v_2)}$, where $v_1 = mk_0 - m - B - L + 2$ and $v_2 = nk_0 - n - B - L + 2$. MSE_T and MSE_R are pooled to get $MSE = (v_1MSE_T + v_2MSE_R)/(v_1 + v_2)$, so we have $n_3S_3^2/\sigma_3^2 \sim \chi^2_{(n_3)}$, where $\sigma_3^2 = \sigma_e^2/k_0$, $n_3 = v_1 + v_2$ and $S_3^2 = MSE/k_0$. Let W denote a random variable which has the same distribution as the D_i, $i = 1, \ldots, n$. We are interested in a tolerance interval for $W \sim N(\theta, \tau^2)$, where $\theta = \mu_T - \mu_R$ and $\tau^2 = \sigma_T^2 = \sigma_1^2 - \sigma_3^2$. Also observe that $\hat{\theta} = \bar{X} - \bar{Y} \sim N(\theta, \sigma^2)$, where $\sigma^2 = \sigma_1^2/m + \sigma_2^2/n$.

For the data provided in Liao and Iyer (2001), $m = 44$, $n = 10$, $B = L = E = 3$, $\hat{\theta} = -1.13654$, $s_1^2 = 0.61928$, $s_2^2 = 0.63132$ and $s_3^2 = 0.19052$. A $(\beta = 0.95, \gamma = 0.90)$-tolerance interval for the distribution of D_i is obtained as $[-2.84498, 0.571899]$ which is completely contained in the interval [-5,5]. Therefore, for this example, one concludes that the batch of test meters has satisfied the quality requirement.

Example 4.2. The individual bioequivalence experiment

Brown, Iyer and Wang (1997) applied the procedure of Wang and Iyer (1994) to evaluate the bioequivalence of two formulations of a drug using various cross-over designs for data collection. One of the designs they gave is the two-period and four-sequence design commonly known as a Balaam design (Balaam (1968)). They used the following model to describe the response from the experiment: $Y_{ijk} = \mu + \kappa_i + \xi_{ij} + \pi_k + \phi_{l[i,k]} + (\xi\phi)_{ijl} + e_{ijk}$, for $i = 1, 2, 3, 4$, $j = 1, \ldots, a$, $k = 1, 2$, $l = 1, 2$. Y_{ijk} denotes the response of the j subject in the i sequence for the kth period and μ is the overall mean, κ_i is the fixed effect of sequence i, ξ_{ij} is the random subject effect, π_k is the fixed effect for the period k, $\phi_{l[i,k]}$ is the direct fixed effect of the lth formulation given in the ith sequence in period k, $(\xi\phi)_{ijl}$ is the interaction of subject ij and drug l, and e_{ijk} is the within-subject random error in observing Y_{ijk}. ξ_{ij}, $(\xi\phi)_{ijl}$ and e_{ijk} are normally distributed with zero mean and standard deviations equal to σ_s, σ_{sd} and σ_e, respectively.

It is of interest to examine the expected difference between the response from the reference and test formulations of an individual at period k. Let this difference be denoted by $\delta_{ij} = \phi_1 - \phi_2 + (\xi\phi)_{ij1} - (\xi\phi)_{ij2}$. One needs a $(\beta = 0.75, \gamma = 0.95)$-tolerance interval for $\delta_{ij} \sim N(\phi_1 - \phi_2, 2\sigma_{sd}^2)$ and check of whether the obtained interval is completely contained in the interval$[0.75, 1.25]$.

Define $d_{ij} = Y_{ij1} - Y_{ij2}$ and $s_{ij} = Y_{ij1} + Y_{ij2}$. Then the following statistics are mutually independent:

$$\hat{\theta} = \frac{1}{4}[(\bar{d_{1.}} - \bar{d_{2.}}) + (\bar{s_{3.}} - \bar{s_{4.}})],$$

$$S_1^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{a}(d_{ij} - \bar{d}_{i.})^2}{2(a-1)},$$

$$S_2^2 = \frac{\sum_{i=3}^{4}\sum_{j=1}^{a}(s_{ij} - \bar{s}_{i.})^2}{2(a-1)},$$

$$S_3^2 = \frac{\sum_{i=3}^{4}\sum_{j=1}^{a}(d_{ij} - \bar{d}_{..})^2}{2a-1},$$

$$S_4^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{a}(s_{ij} - \bar{s}_{..})^2}{2a-1}.$$

Note that in S_3^2, $\bar{d}_{..} = \sum_{i=3}^{4}\sum_{j=1}^{a} d_{ij}/(2a)$ and in S_4^2, $\bar{s}_{..} = \sum_{i=1}^{2}\sum_{j=1}^{a} s_{ij}/(2a)$. Let $\sigma_1^2 = 2\sigma_{sd}^2 + 2\sigma_e^2$, $\sigma_2^2 = 4\sigma_s^2 + 4\sigma_{sd}^2 + 2\sigma_e^2$, $\sigma_3^2 = 2\sigma_e^2$ and $\sigma_4^2 = 4\sigma_s^2 + 2\sigma_{sd}^2 + 2\sigma_e^2$. Then it is easy to check that $n_i S_i^2/\sigma_i^2 \sim \chi^2_{(n_i)}$ for $i = 1, 2, 3, 4$, where $n_1 = n_2 = 2(a-1)$ and $n_3 = n_4 = 2a-1$. Also we have $\hat{\theta} \sim N(\phi_1 - \phi_2, \sigma^2)$, where $\sigma^2 = (\sigma_1^2 + \sigma_2^2)/(8a)$. Therefore, the required tolerance interval can be easily computed using the proposed procedure given in Section 3.

For the log-transformed data provided in Brown, Iyer and Wang (1997), who cited Chow and Liu (1992) as the original source, $\hat{\theta} = 0.1180$, $s_1^2 = 0.0839$, $s_2^2 = 0.5213$, $s_3^2 = 0.1534$, $s_4^2 = 0.2874$ and $a = 6$. A $(\beta = 0.75, \gamma = 0.95)$-tolerance interval for δ_{ij} using the proposed procedure is $[-0.72582, 0.96182]$. After exponentiating, we obtain an interval of $[0.48392, 2.61646]$ which is not completely contained in the interval $[0.75, 1.25]$. So we cannot conclude individual bioequivalence for this example based on the tolerance interval criterion.

5. Simulation Study

To evaluate the performance of the proposed procedure, the following simulation study was carried out based on the glucose monitoring meter experiment of Example 4.1. Without loss of generality, we may assume that $\theta = 0$ and $\sigma_R = 1$. For fixed $B = L = E = 3$ and specified values of m, n, σ_T and σ_e, we generated a normal random deviate Z from the distribution $N(0, \sigma^2)$ and three chi-squared random deviates U_1, U_2 and U_3 with n_1, n_2, n_3 df, using the functions RNORM and RCHISQ, respectively, in the statistical package S-PLUS. The corresponding sample statistics $S_1^2 = U_1\sigma_1^2/n_1$, $S_2^2 = U_2\sigma_2^2/n_2$, $S_3^2 = U_3\sigma_3^2/n_3$ are then generated. We then computed the quantities of $\hat{\tau}_\gamma$, $\hat{\theta}_{(1-\gamma)/2}$ and $\hat{\theta}_{(1+\gamma)/2}$ using the Monte-Carlo algorithm described in Subsection 3.1. When $\hat{\tau}_\gamma > 0$, compute the margin of error $ME = \hat{k}\hat{\tau}_\gamma$ based on (3.7) and let $p = \Phi(Z + ME) - \Phi(Z - ME)$. When $\hat{\tau}_\gamma \leq 0$, then let $p = \Phi(\hat{\theta}_{(1+\gamma)/2}) - \Phi(\hat{\theta}_{(1-\gamma)/2})$, where $\Phi(\cdot)$ is the standard normal distribution function. The procedure was repeated 10,000 times and the fraction of times that p was greater than or equal to β was computed. The results are presented in the following tables.

Table 1. Simulated confidence coefficients (times 10^4) for the ($\beta = 0.95, \gamma = 0.90$)-tolerance interval, with $m = 5, 10$; $B = L = E = 3$ and $\sigma_R = 1$.

		m											
		5						10					
		n						n					
σ_T	σ_e	5	10	20	40	60	80	5	10	20	40	60	80
	0.5	9799	9821	9769	9506	9359	9258	9407	9548	9430	9252	9111	9137
	1	9744	9673	9682	9629	9646	9631	9504	9636	9579	9361	9269	9214
0.5	2	9138	9071	8951	8901	8891	8875	9529	9533	9462	9440	9451	9487
	4	8954	8928	8924	8861	8843	8828	9030	8995	8899	8931	8893	8896
	8	8971	8926	8896	8902	8895	8914	8964	8992	8998	8927	8951	8934
	0.5	9553	9389	9189	9180	9069	9094	9313	9202	9134	9054	9022	9009
	1	9636	9497	9300	9199	9172	9070	9324	9287	9144	9136	9041	9053
1	2	9667	9671	9635	9573	9493	9480	9509	9429	9212	9151	9104	9083
	4	9040	8865	8886	8872	8911	8822	9480	9486	9432	9413	9445	9436
	8	8945	8878	8808	8807	8934	8847	8958	8936	8907	8860	8857	8849
	0.5	9191	9154	9063	9066	9031	9006	9142	9063	9090	8981	8988	8948
	1	9250	9181	9082	9102	8995	9042	9124	9030	9056	9054	8989	8976
2	2	9307	9185	9125	9073	9069	9120	9146	9081	9161	9040	9004	9022
	4	9611	9542	9507	9503	9448	9465	9307	9162	9109	9106	9070	9068
	8	8895	8828	8865	8870	8883	8851	9398	9459	9473	9473	9454	9444
	0.5	9065	9054	8957	8985	9026	9002	9042	9024	9031	8997	9017	9018
	1	9050	9034	9063	9020	8952	9052	8989	8991	8954	9031	9014	8997
4	2	9074	9090	9001	9033	9025	9044	9070	9008	9041	9004	8948	8973
	4	9151	9049	9085	9055	9135	9075	9070	9052	9000	8990	9034	9039
	8	9576	9503	9495	9440	9389	9394	9179	9069	9097	9034	9056	9110

Table 2. Simulated confidence coefficients (times 10^4) for the ($\beta = 0.95, \gamma = 0.90$)-tolerance interval, with $m = 25, 50$; $B = L = E = 3$ and $\sigma_R = 1$.

		m											
		25						50					
		n						n					
σ_T	σ_e	5	10	20	40	60	80	5	10	20	40	60	80
	0.5	8654	8986	9151	9091	9084	9136	8206	8652	8879	8991	8999	9033
	1	8839	9156	9259	9190	9105	9049	8334	8665	8960	9053	9033	9010
0.5	2	9218	9526	9546	9520	9360	9299	8751	9030	9203	9244	9137	9120
	4	9249	9236	9105	9082	9034	8970	9301	9473	9393	9387	9368	9348
	8	8955	8993	9057	9023	9004	8994	8938	9067	9034	9039	9033	8963
	0.5	8968	9060	9033	9006	9046	9039	8715	8912	9001	8962	9021	8927
	1	9062	9133	9090	9005	8900	9054	8652	8970	9037	9038	9033	9038
1	2	9032	9166	9087	9038	9039	9068	8738	8962	9030	9017	9041	8969
	4	9342	9460	9392	9271	9186	9180	8953	9152	9168	9112	9059	9050
	8	9196	9181	9029	8991	8985	9007	9377	9403	9372	9398	9384	9356
	0.5	9008	9052	8980	8967	9025	9033	8965	9006	9046	9023	8983	9009
	1	9044	9067	8992	8992	8949	9014	8964	9003	9015	8971	9029	8943
2	2	9076	9005	8978	9045	8993	8980	8956	9033	8992	8983	9028	9021
	4	9117	9019	9012	9026	9001	9040	8991	8975	8986	9026	8995	8974
	8	9422	9350	9212	9208	9125	9121	9127	9147	9138	9065	9021	9046
	0.5	9046	8999	8981	8976	9013	8986	9002	9019	9069	8933	8999	9003
	1	9001	8965	8976	8957	8971	8992	8987	8997	8991	8978	9012	9013
4	2	9044	9025	9015	8994	8994	8987	9025	8979	8975	8934	8981	9002
	4	9042	8984	8949	9033	8965	8998	8999	8964	8997	8945	8924	8964
	8	9024	9044	8962	8955	9000	9011	9041	9044	9003	9014	9034	8973

Table 3. Simulated confidence coefficients (times 10^4) for the ($\beta = 0.95, \gamma = 0.90$)-tolerance interval, with $m = 75, 100$; $B = L = E = 3$ and $\sigma_R = 1$.

		m											
		75						100					
		n						n					
σ_T	σ_e	5	10	20	40	60	80	5	10	20	40	60	80
	0.5	7951	8301	8628	8833	8968	9019	7866	8217	8532	8828	8802	8926
	1	8013	8406	8728	8918	8956	8937	7860	8283	8635	8841	8908	8960
0.5	2	8374	8811	8992	9090	9121	9122	8274	8541	8911	9019	9060	9028
	4	9228	9435	9526	9598	9667	9634	9056	9404	9525	9621	9667	9635
	8	8998	9087	9091	9031	8999	8961	9026	9169	9030	9020	8965	9060
	0.5	8405	8828	8955	8980	9060	9017	8274	8781	8906	8954	8948	8950
	1	8431	8852	8937	8991	9004	9022	8316	8751	8973	8966	8981	8995
1	2	8502	8862	8990	9031	9028	8982	8298	8743	8940	9020	8995	9033
	4	8718	9011	9094	9056	9084	9051	8551	8918	9046	9048	9040	9044
	8	9319	9484	9616	9606	9623	9634	9175	9476	9592	9636	9563	9477
	0.5	8914	8964	9006	9004	9057	8974	8894	8992	8946	9005	9046	9007
	1	8918	8921	8985	9039	8959	9000	8886	8995	8989	8972	8947	8985
2	2	8943	8926	8947	9075	8987	8962	8878	8988	8958	9031	8967	8993
	4	8883	8972	8982	8932	9001	8992	8757	9006	9029	8985	9007	9029
	8	8936	9053	9091	9066	9063	9060	8764	9003	8988	9053	9002	9044
	0.5	8987	8976	8948	8983	8964	9036	9014	9004	9057	9059	9006	9010
	1	8977	9014	8991	8952	9009	8963	8987	9035	8977	8964	8976	8970
4	2	8969	9023	9018	9066	9047	8966	8957	9031	8990	8970	8979	9036
	4	9025	8976	8969	8982	9008	8961	9020	9010	8988	8952	9014	8948
	8	8990	9005	9003	9038	8994	9031	8971	9006	8964	8997	8965	8992

For most parameter combinations, the constructed tolerance intervals are successful in maintaining the confidence level close to the stated value of $\gamma = 0.90$. Nonetheless, the results indicate that when σ_T is smaller than σ_R and m is small, the proposed tolerance interval can be conservative. On the other hand, when σ_T is smaller than σ_R and m is large but n is small, the proposed tolerance interval appears to be somewhat liberal. Fortunately, in most practical situations, σ_T is usually larger than σ_R because the reference meters tend to have much higher precision than the test meters. If the number of reference meters is at least 10, then the results indicate that the proposed approach can be satisfactory for practical use.

6. Concluding Remarks

Liao and Iyer (2001) conducted the same simulation study to evaluate their procedure. For most of the parameter combinations the coverage probabilities of the two methods are nearly the same. This indicates that they have similar performance. Nonetheless, both Wang-Iyer (1994) and Liao-Iyer (2001) methods are problem specific and are derived for a specific family of distributions under consideration. Clearly, the procedure provided in this study can be applied to tolerance interval questions arising in arbitrary normal balanced mixed linear model situations.

It may be of interest to compare the widely used Satterthwaite approximation, (see Graybill (1976)), with the generalized confidence intervals for obtaining the tolerance intervals. Therefore, we made the following replacements for the procedure given in Section 3. Let $\hat{\tau}_\gamma^2 = \hat{f}_1\hat{\tau}^2/\chi^2_{\hat{f}_1,1-\gamma}$, where the df $\hat{f}_1 = (\sum_{i=1}^q h_i s_i^2)^2/(\sum_{i=1}^q (h_i s_i^2)^2/n_i)$; $\hat{\theta}_{(1-\gamma)/2} = \hat{\theta} - t_{\hat{f}_2,(1+\gamma)/2}\hat{\sigma}$ and $\hat{\theta}_{(1+\gamma)/2} = \hat{\theta} + t_{\hat{f}_2,(1+\gamma)/2}\hat{\sigma}$, where the df $\hat{f}_2 = (\sum_{i=1}^q c_i s_i^2)^2/(\sum_{i=1}^q (c_i s_i^2)^2/n_i)$. Here $\chi^2_{\hat{f}_1,1-\gamma}$ is the $1-\gamma$ percentile of the chi-squared distribution with $\hat{f}_1$ df; and $t_{\hat{f}_2,(1+\gamma)/2}$ is the $(1+\gamma)/2$ percentile of the Student's t-distribution with $\hat{f}_2$ df. Then we conducted the same simulation study based on the glucose monitoring meter experiment. We report the following partial results.

The simulation results indicate that the proposed generalized tolerance intervals may outperform those found by the Satterthwaite approximation, particularly when the value of τ is small. Otherwise, both methods can have similar performance. Note that results for certain parameter combinations are not reported above because the performances of the methods being compared were very similar.

We note that the construction of tolerance intervals using the concept of the generalized confidence intervals can also be applied to the one-sided (β, γ)-tolerance interval for the random variable $W \sim N(\theta, \tau^2)$, where $\tau^2 = \sum_{i=1}^q c_i\sigma_i^2$, which corresponds to a statistic U satisfying $Pr\{Pr[W \leq U] \geq \beta\} = \gamma$. It is obvious that U is simply equal to the upper γ-level generalized confidence bound for $\theta + \tau z_\beta$.

Table 4. Simulated confidence coefficients (times 10^4) for the ($\beta = 0.95, \gamma = 0.90$) tolerance intervals using the generalized confidence intervals and the Satterthwaite approximation procedures. The parameters $B = L = E = 3$ and $\sigma_R = 1$.

		m											
		10						75					
		n						n					
σ_T	σ_e	5	10	20	40	60	80	5	10	20	40	60	80
		9313[a]	9202	9134	9054	9022	9009	8405	8828	8955	8980	9060	9017
	0.5	9336[b]	9196	9159	9040	9077	9019	8441	8852	8956	9000	8975	9013
		0[c]	0	0	0	0	0	0	0	0	0	0	0
		9324	9287	9144	9136	9041	9053	8431	8852	8937	8991	9004	9022
	1	9384	9403	9272	9258	9130	9165	8424	8853	9002	8979	9009	9066
		0	0	0	0	0	0	0	0	0	0	0	0
		9509	9429	9212	9151	9104	9083	8502	8862	8990	9031	9028	8982
1	2	9858	9898	9908	9894	9895	9890	8633	8932	9021	9121	9106	9078
		14	5	12	5	12	10	0	0	0	0	0	0
		9480	9486	9432	9413	9445	9436	8718	9011	9094	9056	9084	9051
	4	9492	9480	9471	9490	9462	9479	9022	9299	9482	9449	9461	9445
		514	520	529	510	538	521	0	0	0	0	0	0
		8958	8936	8907	8860	8857	8849	9319	9484	9616	9606	9623	9634
	8	7506	7118	7091	7087	7118	7074	9687	9744	9741	9721	9772	9729
		2931	2883	2909	2913	2882	2926	289	251	258	279	228	271
		9042[a]	9024	9031	8997	9017	9018	8997	8976	8948	8983	8964	9036
	0.5	9064[b]	9008	8999	9054	8969	8985	8976	9023	8928	9015	9060	8999
		0[c]	0	0	0	0	0	0	0	0	0	0	0
		8989	8991	8954	9031	9014	8997	8977	9014	8991	8952	9009	8963
	1	9031	9026	8991	9021	8992	8956	8995	8967	8985	8969	9013	8998
		0	0	0	0	0	0	0	0	0	0	0	0
		9070	9008	9041	9004	8948	8973	8969	9023	9018	9066	9047	8966
4	2	9064	9076	9036	9017	9016	9044	9023	8986	9053	8949	8977	8955
		0	0	0	0	0	0	0	0	0	0	0	0
		9070	9090	9001	9033	9025	9044	9025	8976	8969	8982	9008	8961
	4	9155	9168	9132	9111	9104	9155	9012	9034	9042	9015	9025	9041
		0	0	0	0	0	0	0	0	0	0	0	0
		9179	9069	9097	9034	9056	9110	8990	9005	9003	9038	8994	9031
	8	9896	9903	9876	9895	9886	9880	9085	9045	9124	9092	9076	9159
		12	6	19	8	5	9	0	0	0	0	0	0

[a]Tolerance intervals constructed by using the generalized confidence intervals.
[b]Tolerance intervals constructed by using the Satterthwaite approximation.
[c]The number of times that $\hat{\tau}$ is less then or equal to 0 over the 10000 stimlations.

Finally, to the best of our knowledge, there appear to be no satisfactory two-sided tolerance interval procedures available in the literature for general unbalanced data situations. Bagui, Bhaumik and Parnes (1996) do discuss procedures for one-sided tolerance limits in m-way random effects ANOVA models.

However, their approach is based on the 'plug-in' method whereby tolerance intervals are derived assuming various parameters to be known and then estimates for these parameters are substituted in the results. The coverage probabilities of intervals based on the 'plug-in' method have not been satisfactorily evaluated in general mixed-model situations. In the context of one-way random effects models, our own simulation studies (unpublished) indicate poor performance for the 'plug-in' methods. We are currently investigating other approaches for obtaining satisfactory tolerance intervals in unbalanced mixed models.

References

Bagui, S. C., Bhaumik, D. K. and Parnes, M. (1996). One-sided tolerance limits for unbalanced m-way random effects Anova models. *J. Appl. Statist. Sci.* **3**, 135-148.

Balaam, L. N. (1968). A two-period design with t^2 experimental units. *Biometrics* **24**, 61-73.

Brown, E. B., Iyer, H. K. and Wang, C. M. (1997). Tolerance intervals for assessing individual bioequivalence. *Statist. Medicine* **16**, 803-820.

Chang, Y. P. and Huang, W. T. (2000). Generalized confidence intervals for the largest value of some functions of parameters under normality. *Statist. Sinica* **10**, 1369-1383.

Chow, S. C. and Liu, J. P.(1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, New York.

Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury Press, North Scituate, Massachusetts.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals*. Wiley, New York.

Howe, W. G. (1969). Two-sided tolerance limits for normal distributions-some improvements. *J. Amer. Statist. Assoc.* **64**, 610-620.

Khuri, A. I., Mathew, T. and Sinha, B. K. (1998). *Statistical Tests for Mixed Linear Models*. Wiley, New York.

Liao, C. T. and Iyer, H. K. (2001). A tolerance interval for assessing the quality of glucose monitoring meters. Department of Statistics, Colorado State University, Technical report 2001/14.

Tsui, K. W. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters, *J. Amer. Statist. Assoc.* **84**, 602-607.

Wang, C. M. and Iyer, H. K. (1994). Tolerance intervals for the distribution of true values in the presence of measurement error. *Technometrics* **36**, 162-170.

Wald, A. and Wolfowitz, J. (1946). Tolerance limits for a normal distribution. *Ann. Math. Statist.* **17**, 208-215.

Weerahandi, S. (1991). Testing variance components in mixed models with generalized p-values. *J. Amer. Statist. Assoc.* **86**, 151-153.

Weerahandi, S. (1993). Generalized confidence intervals. *J. Amer. Statist. Assoc.* **88**, 899-905.

Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. Springer-Verlag, New York.

Zhou, L. and Mathew, T. (1994). Some tests for variance components using generalized p-values. *Technometrics* **36**, 394-401.

Department of Agronomy, National Taiwan University, Taipei, Taiwan.

E-mail: ctliao@ccms.ntu.edu.tw

Department of Statistics, Colorado State University, Fort Collins, CO 80523. U.S.A.

E-mail: hari@stat.colostate.edu

(Received October 2002; accepted May 2003)