

Tolerance Intervals in the Two-Way Nested Model With Mixed or Random Effects

Miguel Fonseca

Faculty of Science and Technology, New University of Lisbon
fmig@fct.unl.pt

Thomas Mathew

University of Maryland, Baltimore County
mathew@umbc.edu

João Tiago Mexia

Faculty of Science and Technology, New University of Lisbon
jtm@fct.unl.pt

Roman Zmysłony

Faculty of Mathematics, Computer Sciences and Econometrics,
University of Zielona Góra
rzmyslony@op.pl

Abstract. We address the problem of deriving a one sided tolerance interval in a two-way nested model with mixed or random effects. The generalized confidence interval idea is used in the derivation of our tolerance limit, and the results are obtained by suitably modifying the approach in Krishnamoorthy and Mathew (2004), for the one-way random model. In general, our tolerance limit has to be estimated by Monte Carlo simulation. However, we have been able to develop closed form approximations in some cases. The performance of our tolerance interval is numerically investigated, and the performance is found to be satisfactory. The results are illustrated with an example.

1. Introduction

In the context of a random effects model, the tolerance interval problem has been well investigated for the one-way random model; we refer to the recent article by Krishnamoorthy and Mathew (2004) for some recent results and earlier references. For more general mixed and random effects models, limited results are available; see Bagui, Bhau-mik and Parnes (1996) and Liao and Iyer (2004). The purpose of this article is to derive tolerance intervals for the two-way nested model with mixed or random effects. Such models are used to analyze data from a variety of application areas including chemical, industrial and animal breeding applications. We shall use the ideas in Krishnamoorthy and Mathew (2004) in order to develop the tolerance intervals in this article. In particular, our approach is based on the tolerance interval idea due to Weerahandi (1993); see also Weerahandi (1995). We have derived tolerance intervals for the observable random variable as well as the unobservable random effect (i.e., the “true effect” without the

error term). The latter problem was addressed for the first time by Wang and Iyer (1994) and later by Krishnamoorthy and Mathew (2004).

Let A and B denote the two factors and suppose A has a levels with b_i levels of B nested within the i th level of A . Let Y_{ijk} denote the k th observation corresponding to the j th level of B nested within the i th level of A ; $k = 1, 2, \dots, n_{ij}$. The model is then given by

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + e_{k(ij)}, \quad (1)$$

with $i = 1, \dots, a$, $j = 1, \dots, b_i$, $k = 1, \dots, n_{ij}$. Here μ is the general mean, τ_i is the effect due to the i th level of A , $\beta_{j(i)}$ is the effect due to the j th level of B nested within the i th level of A , and $e_{k(ij)}$ denotes an error term. In the case of the mixed effects model, we make the assumptions that τ_i 's are fixed effects, $\beta_{j(i)}$'s are random effects with $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$, and $e_{k(ij)} \sim N(0, \sigma_e^2)$. For the random effects model, we make the additional distributional assumption that $\tau_i \sim N(0, \sigma_\tau^2)$. Furthermore, all the random variables are assumed to be independent.

If Y is an observation corresponding to a particular level of B nested within the i th level of A and following the model (1), then $Y \sim N(\mu + \tau_i, \sigma_\beta^2 + \sigma_e^2)$ when (1) is a mixed effects model, and $Y \sim N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$ when (1) is a random effects model. The problems we shall address deal with the computation of an upper tolerance limit for the observable random variable Y , and the unobservable "true effect" $\mu + \tau_i + \beta_{j(i)}$, in the case of the mixed effects model. Thus in the mixed effects model, we shall compute an upper tolerance limit for $N(\mu + \tau_i, \sigma_\beta^2 + \sigma_e^2)$ and for $N(\mu + \tau_i, \sigma_\beta^2)$. In the random effects model, our problem is the computation of an upper tolerance for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$ and for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2)$. We shall denote by p the content of the tolerance interval and by γ the coverage of the interval, and refer to the tolerance interval simply as a (p, γ) tolerance interval, and the corresponding tolerance limit as a (p, γ) tolerance limit. Thus if c is a (p, γ) upper tolerance limit for $Y \sim N(\mu + \tau_i, \sigma_\beta^2 + \sigma_e^2)$, then c will be a function of the Y_{ij} 's, and satisfies the condition

$$P_{Y'_{ij}s} \left[P_Y \{ Y \leq c | Y'_{ij}s \} \geq p \right] = \gamma.$$

It is easy to verify that the (p, γ) upper tolerance limits mentioned above are 100 γ % upper confidence limits for the p th percentiles of the appropriate normal distributions. For example, the (p, γ) upper tolerance limit for $N(\mu + \tau_i, \sigma_\beta^2 + \sigma_e^2)$ is the 100 γ % upper confidence limit for the parametric function $\mu + \tau_i + z_p \sqrt{\sigma_\beta^2 + \sigma_e^2}$, where z_p is the p th percentile of the standard normal distribution. The latter parametric function is obviously the p th percentile of $N(\mu + \tau_i, \sigma_\beta^2 + \sigma_e^2)$.

Inference concerning the variance components in the model (1) has already been addressed in the literature; see the book by Burdick and Graybill (1992) for details on the computation of confidence intervals. Except the error sum of squares, the ANOVA sums of squares under the model (1) are not distributed as multiples of chisquares in the unbalanced case. However, chisquare approximations can be used for the derivation

of approximate confidence intervals. In particular, let $\bar{Y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$ and define

$$\begin{aligned} SS_e &= \sum_{i=1}^a \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 \\ SS_\beta^* &= \sum_{i=1}^a \sum_{j=1}^{b_i} (\bar{Y}_{ij.} - \frac{1}{b_i} \sum_{j=1}^{b_i} \bar{Y}_{ij.})^2. \end{aligned} \quad (2)$$

It is easy to verify that

$$\begin{aligned} E(SS_\beta^*) &= (b. - a)\sigma_\beta^2 + \lambda\sigma_e^2, \\ \text{where } \lambda &= \frac{1}{b. - a} \sum_{i=1}^a \frac{b_i - 1}{b_i} \sum_{j=1}^{b_i} \frac{1}{n_{ij}}, \end{aligned} \quad (3)$$

and $b. = \sum_{i=1}^a b_i$. In order to obtain approximate inferences in the unbalanced case of the model (1), we can use the independent distributions

$$\begin{aligned} U_e &= \frac{SS_e}{\sigma_e^2} \sim \chi_{n..-b.}^2, \text{ exactly,} \\ \text{and } U_\beta^* &= \frac{SS_\beta^*}{\sigma_\beta^2 + \lambda\sigma_e^2} \sim \chi_{b.-a}^2, \text{ approximately,} \end{aligned} \quad (4)$$

where $n.. = \sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij}$ and χ_r^2 denotes the chisquare distribution with r degrees of freedom. The above approximate distribution associated with SS_β^* is given in Burdick and Graybill (1992), and can be derived similar to the derivation in Thomas and Hultquist (1978) for the one-way random model with unbalanced data.

The model (1) given above corresponds to the most general unbalanced data situation; however, we could accomplish our tolerance interval construction only in some restricted cases. In the mixed effects model, we assume that $b_i = b$ and $n_{ij} = n_j$ ($i = 1, \dots, a$, $j = 1, \dots, b$). In the random effects case, our derivations are only for the balanced case, i.e., $b_i = b$ and $n_{ij} = n$ ($i = 1, \dots, a$, $j = 1, \dots, b$). The tolerance interval construction for the mixed effects model is described in the next section, and for the random effects model, the procedure is developed in Section 3. As already pointed out, our methodology for deriving a (p, γ) upper tolerance limit is based on the idea of a generalized confidence interval. For this, we have defined a generalized pivot statistic for the parametric function representing the p th percentile of the relevant normal distribution. The 100γ th percentile of the generalized pivot statistic then gives the (p, γ) upper tolerance limit. The limit can be easily estimated using a Monte Carlo simulation. In some cases, it is also possible to develop a closed form approximation for the tolerance limit. We have numerically investigated the performance of our proposed tolerance limits. The results are also illustrated with an example.

For mixed models with balanced data, Liao and Iyer (2004) have derived two-sided tolerance limits following an approximation due to Wald and Wolfowitz (1946) along

with the generalized confidence interval procedure. In the present paper, the approach used is based entirely on the generalized confidence interval idea.

2. The Mixed Effects Model

We shall consider the model (1) in the special case where $b_i = b$ and $n_{ij} = n_j$ ($i = 1, \dots, a, j = 1, \dots, b$). We shall write $\mu_i = \mu + \tau_i$ and rewrite the model as

$$Y_{ijk} = \mu_i + \beta_{j(i)} + e_{k(ij)}, \quad (5)$$

with $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, n_j$. When $b_i = b$ and $n_{ij} = n_j$, λ in (3) simplifies to

$$\lambda = \frac{1}{b} \sum_{j=1}^b \frac{1}{n_j}.$$

Let

$$W_i = \hat{\mu}_i = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij}.$$

Then

$$W_i \sim N\left(\mu_i, \frac{\sigma_\beta^2 + \lambda\sigma_e^2}{b}\right). \quad (6)$$

It is easy to verify that SS_e , SS_β^* (given in (2)) and the W_i 's are all independently distributed. Our tolerance limits will be based on these random variables. We shall also denote by ss_e , ss_β^* and w_i the respective observed values.

2.1. An upper tolerance limit for $N(\mu_i, \sigma_\beta^2 + \sigma_e^2)$

As already noted, a (p, γ) upper tolerance limit for $N(\mu_i, \sigma_\beta^2 + \sigma_e^2)$ is a $100\gamma\%$ upper confidence limit for $\mu_i + z_p \sqrt{\sigma_\beta^2 + \sigma_e^2}$. In order to obtain a generalized confidence limit for $\mu_i + z_p \sqrt{\sigma_\beta^2 + \sigma_e^2}$, we shall first define a generalized pivot statistic. Following the procedure in Krishnamoorthy and Mathew (2004), let

$$\begin{aligned} T_{1p} &= w_i - \frac{\sqrt{b}(W_i - \mu_i)}{\sqrt{SS_\beta^*}} \sqrt{\frac{ss_\beta^*}{b}} \\ &+ z_p \left[\frac{\sigma_\beta^2 + \lambda\sigma_e^2}{SS_\beta^*} ss_\beta^* + (1 - \lambda) \frac{\sigma_e^2}{SS_e} ss_e \right]^{1/2} \\ &\approx_d w_i - \frac{Z}{\sqrt{U_\beta^*}} \sqrt{\frac{ss_\beta^*}{b}} + z_p \left[\frac{ss_\beta^*}{U_\beta^*} + (1 - \lambda) \frac{ss_e}{U_e} \right]^{1/2}, \end{aligned} \quad (7)$$

where U_e and U_β^* are defined in (4), and \approx_d stands for ‘‘approximately distributed as’’. The distribution in the last line of (7) is only approximate since the chisquare distribution of U_β^* in (4) is only approximate, and we will proceed as though U_β^* has a chisquare

distribution. A generalized pivot statistic should satisfy the following conditions: (i) given the observed data, the distributions of the generalized pivot statistic should be free of any unknown parameters, and (ii) the observed value of the generalized pivot statistic (obtained by replacing the random variables by their respective observed values) should be the parameter of interest. It is readily verified that T_{1p} in (7) satisfies the first condition approximately. From the first expression for T_{1p} , it is easy to see that the observed value of T_{1p} is $\mu_i + z_p \sqrt{\sigma_\beta^2 + \sigma_e^2}$, the parameter of interest. The 100γ th percentile of T_{1p} will give an approximate $100\gamma\%$ generalized upper confidence limit for $\mu_i + z_p \sqrt{\sigma_\beta^2 + \sigma_e^2}$, or, equivalently, an approximate (p, γ) upper tolerance limit for $N(\mu_i, \sigma_\beta^2 + \sigma_e^2)$.

It is easy to estimate the percentile $T_{1p}(\gamma)$ of T_{1p} by Monte Carlo simulation. Note that once we have the data, the quantities w_i , ss_e and ss_β^* can be computed, and these are to be treated as fixed while simulating the percentile. We shall denote the percentile so obtained by $T_{1p}(\gamma)$. It is also possible to develop an approximation for the percentile, following the arguments in Krishnamoorthy and Mathew (2004). For this, let $t_{r;\alpha}(\delta)$ denote the 100α th percentile of a non-central t distribution with r degrees of freedom and non-centrality parameter δ , and $F_{r_1, r_2; \alpha}$ denote the 100α th percentile of an F distribution with (r_1, r_2) degrees of freedom. Following the derivation in Krishnamoorthy and Mathew (2004), an approximation, say $T_{1p}(\gamma)^*$, for the (p, γ) upper tolerance limit is given by

$$T_{1p}(\gamma)^* = w_i + t_{a(b-1); \gamma}(\delta_1) \sqrt{\frac{ss_\beta^*}{ab(b-1)}},$$

$$\text{where } \delta_1 = z_p \sqrt{b + \frac{b(b-1)(1-\lambda)ss_e}{(n.-b)ss_\beta^*} F_{a(b-1), a(n.-b); 1-\gamma}}, \quad (8)$$

with $n. = \sum_{j=1}^b n_j$.

2.2. An upper tolerance limit for $N(\mu_i, \sigma_\beta^2)$

The generalized pivot is now given by

$$T_{2p} = w_i - \frac{\sqrt{b}(W_i - \mu_i)}{\sqrt{SS_\beta^*}} \sqrt{\frac{ss_\beta^*}{b}}$$

$$+ z_p \left[\frac{\sigma_\beta^2 + \lambda \sigma_e^2}{SS_\beta^*} ss_\beta^* - \lambda \frac{\sigma_e^2}{SS_e} ss_e \right]_+^{1/2}$$

$$\approx_d w_i - \frac{Z}{\sqrt{U_\beta^*}} \sqrt{\frac{ss_\beta^*}{b}} + z_p \left[\frac{ss_\beta^*}{U_\beta^*} - \lambda \frac{ss_e}{U_e} \right]_+^{1/2}, \quad (9)$$

where for any real number c , $c_+ = \max(c, 0)$. It can once again be checked that T_2 satisfies the properties required of a generalized pivot statistic. Hence the 100γ th percentile

$T_{2p}(\gamma)$ of T_{2p} gives an approximate (p, γ) upper tolerance limit for $N(\mu_i, \sigma_\beta^2)$. Similar to $T_{1p}(\gamma)^*$, an approximation for the 100γ th percentile of T_{2p} is given by

$$T_{2p}(\gamma)^* = w_i + t_{a(b-1); \gamma}(\delta_2) \sqrt{\frac{ss_\beta^*}{ab(b-1)}},$$

$$\text{where } \delta_2 = z_p \sqrt{\left[b - \lambda \frac{b(b-1)ss_e}{(n.-b)ss_\beta^*} F_{a(b-1), a(n.-b); 1-\gamma} \right]_+}. \quad (10)$$

2.3. Numerical results

In order to assess the performance of our tolerance intervals, the coverage probabilities were estimated based on 10,000 simulated samples for $p = .90$ and $\gamma = .95$, using the R statistical software. Without loss of generality we assumed $\mu = 0$ and $\sigma_e^2 = 1$. The estimated coverages given in Table 1 below are expressed as a function of $\rho = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_e^2}$. The simulations were carried out under two set ups: (i) a balanced model with $a = b = n = 5$, and (ii) an unbalanced model with $a = b = 5$ and $n_{i1} = 5, n_{i2} = 7, n_{i3} = 9, n_{i4} = 11, n_{i5} = 13, i = 1, \dots, 5$. The results are given in Table 1 below, where $\hat{T}_{1p}(\gamma)$ and $\hat{T}_{2p}(\gamma)$ denote Monte Carlo estimates of the tolerance limits, and $T_{1p}(\gamma)^*$ and $T_{2p}(\gamma)^*$ denote the approximations given in (8) and (10), respectively.

Table 1. Estimated coverages of the upper tolerance limits for $N(\mu_i, \sigma_\beta^2 + \sigma_e^2)$ and $N(\mu_i, \sigma_\beta^2)$

Table 1a. Coverages of the upper tolerance limits for $N(\mu_i, \sigma_\beta^2 + \sigma_e^2)$

Balanced Model					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{1p}(\gamma)$	0.9583	0.9548	0.9507	0.9495	0.9502
$T_{1p}(\gamma)^*$	0.8899	0.9187	0.9317	0.9457	0.9480
Unbalanced Model					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{1p}(\gamma)$	0.9479	0.9512	0.9473	0.9471	0.9453
$T_{1p}(\gamma)^*$	0.8711	0.9153	0.9366	0.9437	0.9475

Table 1b. Coverages of the upper tolerance limits for $N(\mu_i, \sigma_\beta^2)$

Balanced Model					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{2p}(\gamma)$	0.9364	0.9445	0.9485	0.9510	0.9481
$T_{2p}(\gamma)^*$	0.9925	0.9755	0.9648	0.9560	0.9527
Unbalanced Model					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{2p}(\gamma)$	0.9412	0.9482	0.9492	0.9458	0.9471
$T_{2p}(\gamma)^*$	0.9874	0.9721	0.9601	0.9527	0.9535

From the numerical results in Table 1, it is clear that $\hat{T}_{1p}(\gamma)$ and $\hat{T}_{2p}(\gamma)$ are very satisfactory upper tolerance limits, providing coverages very close to the nominal level. The approximation $T_{1p}(\gamma)^*$ provides coverages below the nominal level for small values of ρ and the approximation $T_{2p}(\gamma)^*$ provides coverages above the nominal level for small values of ρ . However, as ρ gets large, both the approximations perform satisfactorily.

3. The Random Effects Model

Our results in this section apply only to the case of balanced data. Thus let $b_i = b$ and $n_{ij} = n$, and the model is given by

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + e_{k(ij)}, \quad (11)$$

$k = 1, 2, \dots, n$; $j = 1, 2, \dots, b$ and $i = 1, 2, \dots, a$. We assume $\tau_i \sim N(0, \sigma_\tau^2)$, $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$, $e_{ijk} \sim N(0, \sigma_e^2)$, and all the random variables are independent. Define $Y_{ij.} = \sum_{k=1}^n Y_{ijk}$, $Y_{i..} = \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ and $Y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ and consider the usual ANOVA sums of squares given by

$$\begin{aligned} SS_\tau &= \frac{1}{bn} \sum_{i=1}^a (Y_{i..} - \bar{Y}_{...})^2 \\ SS_\beta &= \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b (Y_{ij.} - \bar{Y}_{i..})^2 \\ SS_e &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij.})^2 \end{aligned}$$

Then

$$\begin{aligned} U_\tau &= SS_\tau / (bn\sigma_\tau^2 + n\sigma_\beta^2 + \sigma_e^2) \sim \chi_{a-1}^2 \\ U_\beta &= SS_\beta / (n\sigma_\beta^2 + \sigma_e^2) \sim \chi_{a(b-1)}^2 \\ U_e &= SS_e / \sigma_e^2 \sim \chi_{ab(n-1)}^2 \\ Z &= \sqrt{abn}(\bar{Y}_{...} - \mu) / \sqrt{bn\sigma_\tau^2 + n\sigma_\beta^2 + \sigma_e^2} \sim N(0, 1). \end{aligned} \quad (12)$$

We note that $SS_\beta = nSS_\beta^*$, where SS_β^* is given in (2). We also note that $Y_{ijk} \sim N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$.

3.1. An upper tolerance limit for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$

As before, a (p, γ) upper tolerance limit for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$ is the same as a $100\gamma\%$ upper confidence limit for $\mu + z_p \sqrt{\sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2}$. We shall once again obtain such a confidence limit using the generalized confidence interval idea. Thus let $\bar{y}_{...}$, ss_τ , ss_β and ss_e denote the observed values of the corresponding random variables. Define

$$T_{3p} = \bar{y}_{...} - \frac{\sqrt{abn}(\bar{Y}_{...} - \mu)}{\sqrt{SS_\tau}} \times \frac{\sqrt{ss_\tau}}{abn} + z_p \left[\frac{\sigma_e^2}{SS_e} \times ss_e + \frac{1}{n} \left(\frac{n\sigma_\beta^2 + \sigma_e^2}{SS_\beta} \times ss_\beta - \frac{\sigma_e^2}{SS_e} \times ss_e \right) \right]$$

$$\begin{aligned}
& + \frac{1}{bn} \left(\frac{bn\sigma_\tau^2 + n\sigma_\beta^2 + \sigma_e^2}{SS_\tau} \times ss_\tau - \frac{n\sigma_\beta^2 + \sigma_e^2}{SS_\beta} \times ss_\beta \right) \Big]^{1/2} \\
& = \bar{y}_{\dots} - \frac{Z}{\sqrt{U_\tau}} \times \frac{\sqrt{ss_\tau}}{\sqrt{abn}} + z_p \left[\frac{ss_e}{U_e} + \frac{1}{n} \left(\frac{ss_\beta}{U_\beta} - \frac{ss_e}{U_e} \right) + \frac{1}{bn} \left(\frac{ss_\tau}{U_\tau} - \frac{ss_\beta}{U_\beta} \right) \right]^{1/2} \\
& = \bar{y}_{\dots} - \frac{Z}{\sqrt{U_\tau}} \times \frac{\sqrt{ss_\tau}}{\sqrt{abn}} + \frac{z_p}{\sqrt{bn}} \left[\frac{ss_\tau}{U_\tau} + (b-1) \frac{ss_\beta}{U_\beta} + b(n-1) \frac{ss_e}{U_e} \right]^{1/2}. \tag{13}
\end{aligned}$$

The 100γ th percentile of T_{3p} will provide the required $100\gamma\%$ upper confidence limit for $\mu + z_p \sqrt{\sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2}$, and hence a (p, γ) upper tolerance limit for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$.

The above percentile, say $T_{3p}(\gamma)$ is easily estimated by Monte Carlo simulation. In order to develop an approximation for this percentile, similar to the approximation in (8), we need to use a further approximation for the distribution of $(b-1) \frac{ss_\beta}{U_\beta} + b(n-1) \frac{ss_e}{U_e}$ in (13). We shall use the approximation

$$(b-1) \frac{ss_\beta}{U_\beta} + b(n-1) \frac{ss_e}{U_e} \approx_d \frac{c}{\chi_f^2}, \tag{14}$$

where the constant c and the degrees of freedom f have to be determined by equating the first and second moments. For this, we use the fact that $E[1/\chi_r^2] = 1/(r-2)$ and $E[1/(\chi_r^2)^2] = 1/[(r-2)(r-4)]$. Equating the first and second moments of both expressions in (14), we get the following equations:

$$\begin{aligned}
\frac{c}{f-2} & = \frac{(b-1)ss_\beta}{a(b-1)-2} + \frac{b(n-1)ss_e}{ab(n-1)-2} = e_1 \quad (\text{say}) \\
\frac{c^2}{(f-2)(f-4)} & = \frac{(b-1)^2 ss_\beta^2}{[a(b-1)-2][a(b-1)-4]} + \frac{b^2(n-1)^2 ss_e^2}{[ab(n-1)-2][ab(n-1)-4]} \\
& \quad + \frac{2b(b-1)(n-1)ss_\beta ss_e}{[a(b-1)-2][ab(n-1)-2]} = e_2 \quad (\text{say}). \tag{15}
\end{aligned}$$

Solving, we get

$$c = \frac{2e_1 e_2}{e_2 - e_1^2}, \quad f = 2 \left[1 + \frac{e_2}{e_2 - e_1^2} \right]. \tag{16}$$

We now have the approximation

$$\begin{aligned}
T_{3p} & \approx_d \bar{y}_{\dots} - \frac{Z}{\sqrt{U_\tau}} \times \frac{\sqrt{ss_\tau}}{\sqrt{abn}} + \frac{z_p}{\sqrt{bn}} \left[\frac{ss_\tau}{U_\tau} + \frac{c}{U} \right]^{1/2}, \quad (\text{where } U \sim \chi_f^2) \\
& = \bar{y}_{\dots} - \frac{\sqrt{ss_\tau}}{\sqrt{abn(a-1)}} \frac{1}{\sqrt{U_\tau/(a-1)}} \left[-Z + z_p \sqrt{a} \left(1 + \frac{c(a-1)}{dss_\tau} F_0 \right)^{1/2} \right],
\end{aligned}$$

where $F_0 = \frac{U_\tau/(a-1)}{U/f} \approx_d F_{(a-1),f}$, the central F distribution with $(a-1, f)$ degrees of freedom. Now using the approximation in Krishnamoorthy and Mathew (2004) that

lead to (8), we replace F_0 by the $100(1 - \gamma)$ th percentile of $F_{(a-1),f}$. This finally gives

$$\begin{aligned} T_{3p} &\approx_d \bar{y}_{\dots} - \frac{\sqrt{ss_{\tau}}}{\sqrt{abn(a-1)}} \frac{1}{\sqrt{U_{\tau}/(a-1)}} \left[-Z + z_p \sqrt{a} \left(1 + \frac{c(a-1)}{f \times ss_{\tau}} F_{(a-1),f;1-\gamma} \right)^{1/2} \right] \\ &= \bar{y}_{\dots} - \frac{\sqrt{ss_{\tau}}}{\sqrt{abn(a-1)}} t_{a-1}(\delta_3), \end{aligned}$$

where $t_{a-1}(\delta_3)$ denotes a non-central t random variable with $(a - 1)$ degrees of freedom and non-centrality parameter δ_3 given by

$$\delta_3 = z_p \sqrt{a} \left(1 + \frac{c(a-1)}{f \times ss_{\tau}} F_{(a-1),f;1-\gamma} \right)^{1/2}. \quad (17)$$

Hence an approximate 100γ th percentile of T_{3p} is given by

$$T_{3p}(\gamma)^* = \bar{y}_{\dots} + t_{a-1;\gamma}(\delta_3) \frac{\sqrt{ss_{\tau}}}{\sqrt{abn(a-1)}}. \quad (18)$$

3.2. An upper tolerance limit for $N(\mu, \sigma_{\tau}^2 + \sigma_{\beta}^2)$

A (p, γ) upper tolerance limit for $N(\mu, \sigma_{\tau}^2 + \sigma_{\beta}^2)$ is the same as a $100\gamma\%$ upper confidence limit for $\mu + z_p \sqrt{\sigma_{\tau}^2 + \sigma_{\beta}^2}$. Define the generalized pivot statistic

$$\begin{aligned} T_{4p} &= \bar{y}_{\dots} - \frac{\sqrt{abn}(\bar{Y}_{\dots} - \mu)}{\sqrt{SS_{\tau}}} \times \frac{\sqrt{ss_{\tau}}}{abn} + z_p \left[\frac{1}{n} \left(\frac{n\sigma_{\beta}^2 + \sigma_e^2}{SS_{\beta}} \times ss_{\beta} - \frac{\sigma_e^2}{SS_e} \times ss_e \right) \right. \\ &\quad \left. + \frac{1}{bn} \left(\frac{bn\sigma_{\tau}^2 + n\sigma_{\beta}^2 + \sigma_e^2}{SS_{\tau}} \times ss_{\tau} - \frac{n\sigma_{\beta}^2 + \sigma_e^2}{SS_{\beta}} \times ss_{\beta} \right) \right]_+^{1/2} \\ &= \bar{y}_{\dots} - \frac{Z}{\sqrt{U_{\tau}}} \times \frac{\sqrt{ss_{\tau}}}{\sqrt{abn}} + z_p \left[\frac{1}{n} \left(\frac{ss_{\beta}}{U_{\beta}} - \frac{ss_e}{U_e} \right) + \frac{1}{bn} \left(\frac{ss_{\tau}}{U_{\tau}} - \frac{ss_{\beta}}{U_{\beta}} \right) \right]_+^{1/2} \\ &= \bar{y}_{\dots} - \frac{Z}{\sqrt{U_{\tau}}} \times \frac{\sqrt{ss_{\tau}}}{\sqrt{abn}} + \frac{z_p}{\sqrt{bn}} \left[\frac{ss_{\tau}}{U_{\tau}} + (b-1) \frac{ss_{\beta}}{U_{\beta}} - b \frac{ss_e}{U_e} \right]_+^{1/2} \end{aligned}$$

The 100γ th percentile of T_4 , say $T_{4p}(\gamma)$, will provide the required $100\gamma\%$ upper confidence limit for $\mu + z_p \sqrt{\sigma_{\tau}^2 + \sigma_{\beta}^2}$, and hence a (p, γ) upper tolerance limit for $N(\mu, \sigma_{\tau}^2 + \sigma_{\beta}^2)$. We have not been able to develop an approximation for this percentile.

3.3. Numerical results

Simulation results similar to those in Table 1 are given in Table 2 below. The simulations were carried out fixing $n = 20$, and the quantities a and b were chosen to be 10 or 20. Once again, we chose $p = .90$ and $\gamma = .95$. We also chose $\sigma_{\beta}^2 = 1$ and

$\sigma_e^2 = 1$ and have expressed the coverage probabilities as a function of $\rho = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2}$. The results in Table 2 are also based on 10,000 simulations, and $\hat{T}_{3p}(\gamma)$ and $\hat{T}_{4p}(\gamma)$ represent the tolerance limits estimated by Monte Carlo. The notation $T_{3p}(\gamma)^*$ represents the approximation given in (18).

Table 2. Estimated coverages of the upper tolerance limits for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$ and $N(\mu, \sigma_\tau^2 + \sigma_\beta^2)$

Table 2a. Coverages of the upper tolerance limits for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2)$

$a = 5, b = 5$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{3p}(\gamma)$	0.9738	0.9703	0.9653	0.9594	0.9523
$T_{3p}(\gamma)^*$	0.8998	0.9171	0.9301	0.9390	0.9463
$a = 5, b = 20$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{3p}(\gamma)$	0.9694	0.9660	0.9611	0.9568	0.9518
$T_{3p}(\gamma)^*$	0.8821	0.9072	0.9250	0.9366	0.9454
$a = 20, b = 5$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{3p}(\gamma)$	0.9716	0.9683	0.9668	0.9647	0.9581
$T_{3p}(\gamma)^*$	0.8684	0.8945	0.9151	0.9313	0.9444
$a = 20, b = 20$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{3p}(\gamma)$	0.9634	0.9611	0.9598	0.9592	0.9573
$T_{3p}(\gamma)^*$	0.8375	0.8824	0.9084	0.9275	0.9443

Table 2b. Coverages of the upper tolerance limits for $N(\mu, \sigma_\tau^2 + \sigma_\beta^2)$

$a = 5, b = 5$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{4p}(\gamma)$	0.9766	0.9723	0.9668	0.9593	0.9521
$a = 5, b = 20$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{4p}(\gamma)$	0.9715	0.9661	0.9600	0.9555	0.9512
$a = 20, b = 5$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{4p}(\gamma)$	0.9730	0.9717	0.9672	0.9624	0.9571
$a = 20, b = 20$					
ρ	0.1	0.3	0.5	0.7	0.9
$\hat{T}_{4p}(\gamma)$	0.9642	0.9628	0.9602	0.9577	0.9553

From the numerical results in Table 2, it follows that both $\hat{T}_{3p}(\gamma)$ and $\hat{T}_{4p}(\gamma)$ exhibit satisfactory performance unless a , b and ρ are small, in which case the tolerance intervals are somewhat conservative. However, the approximation $T_{3p}(\gamma)^*$ can be recommended only for large values of ρ .

4. An Example

To illustrate the use of upper tolerance limits, we will consider a data set from a breeding experiment given in Sahai and Ageel (2000, p. 379). The experiment evaluated the breeding value of five sires in raising pigs. Each sire was mated to two randomly selected dams and the average weight gain of two pigs from each litter were recorded.

Suppose that a two way nested model with mixed effects is applicable where τ_i 's represent the sire effects (fixed) and β_j 's represent the dam effects (random). Computations based on the data gave the values

$$y_{1..} = 2.67, y_{2..} = 2.53, y_{3..} = 2.63, y_{4..} = 2.47, y_{5..} = 2.57, \\ ss_\beta = 0.56, ss_e = 0.39$$

Thus, (0.9, 0.95) upper tolerance limits for $N(\mu_i, \sigma_\beta^2 + \sigma_e^2)$, $i = 1, \dots, 5$, are

i	$\hat{T}_{1p}(\gamma)$	$T_{1p}(\gamma)^*$
1	3.52	3.51
2	3.37	3.38
3	3.49	3.48
4	3.33	3.32
5	3.43	3.42

For the $N(\mu_i, \sigma_\beta^2)$, $i = 1, \dots, 5$, the upper tolerance limits are

i	$\hat{T}_{2p}(\gamma)$	$T_{2p}(\gamma)^*$
1	3.46	3.47
2	3.32	3.34
3	3.42	3.44
4	3.26	3.28
5	3.36	3.38

We notice the close agreement between the Monte Carlo estimate and the approximation.

If both factors are random, we have

$$ss_\tau = 0.05, ss_\beta = 0.56, ss_e = 0.39,$$

and the (.90, .95) upper tolerance limits are

$$\text{For } N(\mu, \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2): \quad \hat{T}_{3p}(\gamma) = 3.08, \quad T_{3p}(\gamma)^* = 2.87 \\ \text{For } N(\mu, \sigma_\tau^2 + \sigma_\beta^2): \quad \hat{T}_{4p}(\gamma) = 2.99$$

We note that $\hat{T}_{3p}(\gamma)$ and $T_{3p}(\gamma)^*$ produce significantly different values. From the numerical results in Table 2, we recall that the tolerance interval based on $T_{3p}(\gamma)^*$ results in poor coverage, especially for small values of $\rho = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2}$. In other words, we expect $T_{3p}(\gamma)^*$ to be smaller than $\hat{T}_{3p}(\gamma)$ for small values of ρ . For our example, the estimate of ρ is $\hat{\rho} = -0.30$ (since $\hat{\sigma}_\tau^2 = -0.01$), leading us to believe that ρ is very small.

5. Concluding Remarks

The tolerance interval problem has been extensively investigated for the univariate normal distribution, and for the one-way random model. This article addresses the problem for a two-way nested model with mixed or random effects. Our approach exploits the property that the computation of a one-sided tolerance limit reduces to the computation of a confidence limit for an appropriate percentile of the distribution; the generalized confidence interval idea is then used to derive the required confidence limit. Our tolerance limit can be easily estimated by Monte Carlo simulation, and numerical results indicate that the proposed tolerance interval exhibits satisfactory performance in terms of providing coverage close to the nominal level. However, the approximations that we have developed for the tolerance limit perform well only in some cases. Note that even though we have addressed the problem of deriving an upper tolerance limit, the computation of a lower tolerance limit can be carried out in an obvious manner. For example, similar to (8), a lower tolerance limit has the approximation, say $\tilde{T}_{1p}(\gamma)$, given by

$$\tilde{T}_{1p}(\gamma) = w_i - t_{a(b-1); \gamma}(\delta_1) \sqrt{\frac{ss_\beta^*}{ab(b-1)}},$$

where the various quantities are as defined in (8).

It should be emphasized that the approach pursued in this article is useful only for the derivation of one-sided tolerance limits. We refer to Liao and Iyer (2000) for the derivation of approximate two-sided tolerance limits in mixed and random effects models.

References

1. Bagui, S. C., Bhaumik, D. K. and Parnes, M. (1996). One-sided tolerance limits for unbalanced m -way random-effects ANOVA models. *Journal of Applied Statistical Science*, **3**, 135–147.
2. Burdick, R. and Graybill, F. A. (1992). *Confidence Intervals on Variance Components*. New York: Marcel Dekker.
3. Krishnamoorthy, K. and Mathew, T. (2004). Generalized confidence limits and one-sided tolerance limits in balanced and unbalanced one-way random models. *Technometrics*, **46**, 44–52.

4. Liao, C. T. and Iyer, H. K. (2004). A tolerance interval for the normal distribution with several variance components. *Statistica Sinica*, **14**, 217-229.
5. Sahai H. and Ageel M. I. (2000). *The analysis of variance. Fixed, random and mixed models*. Boston: Birkhuser
6. Thomas, J. D. and Hultquist, R. A. (1978). Interval estimation for the unbalanced case of the one-way random effects model. *The Annals of Statistics*, **6**, 582-587.
7. Wald, A. and Wolfowitz, J. (1946). Tolerance limits for a normal distribution. *The Annals of Mathematical Statistics*, **17**, 208-215.
8. Wang, C. M. and Iyer, H. K. (1994). Tolerance intervals for the distribution of true values in the presence of measurement errors. *Technometrics*, **36**, 162-170.
9. Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, **88**, 899-905.
10. Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. New York: Springer-Verlag.