Normal Tolerance Interval Procedures in the tolerance Package

by Derek S. Young

Abstract Statistical tolerance intervals are used for a broad range of applications, such as quality control, engineering design tests, environmental monitoring, and bioequivalence testing. **tolerance** is the only R package devoted to procedures for tolerance intervals and regions. Perhaps the most commonly-employed functions of the package involve normal tolerance intervals. A number of new procedures for this setting have been included in recent versions of **tolerance**. In this paper, we discuss and illustrate the functions that implement these normal tolerance interval procedures, one of which is a new, novel type of operating characteristic curve.

Introduction and overview of the tolerance package

$(1 - \alpha, P)$ statistical tolerance intervals provide bounds to capture at least a specified proportion P of the sampled population with a given confidence level $1 - \alpha$. The quantity P is called the *content* of the tolerance interval and the confidence level $1 - \alpha$ reflects the sampling variability. There is an extensive literature on tolerance intervals with some of the earliest works being Wilks (1941, 1942) and Wald (1943). The texts by Guttman (1970) and Krishnamoorthy and Mathew (2009) are devoted to the theoretical development and application of tolerance intervals, while the text by Hahn and Meeker (1991) discusses their application in the broader context of statistical intervals.

tolerance (Young, 2010) is a popular R package for constructing exact and approximate tolerance intervals and regions. Since its initial release in 2009, the package has grown to include tolerance interval procedures for a large number of parametric distributions, nonparametric settings, and regression models. There are also tolerance region procedures for the multivariate normal and multivariate regression settings. Procedures for more specific settings are also included, such as one-sided tolerance limits for the difference between two independent normal random variables (Hall, 1984) and fiducial tolerance intervals for the function of parameters from discrete distributions (Mathew and Young, 2013). The package also includes some graphical capabilities for visualizing the tolerance intervals (regions) by plotting the limits (regions) on histograms, scatterplots, or control charts of the data.

tolerance has been used for a broad range of applications, including cancer research (Heck et al., 2014), wildlife biology (Pasquaretta et al., 2012), assessing the performance of genetic algorithms (Van der Borght et al., 2014), ratio editing in surveys (Young and Mathew, 2015), air quality assessment (Hafner et al., 2013), and instrumentation testing (Burr and Gavron, 2010). General interest in **tolerance** can be gauged by the **cranlogs** package (Csardi, 2015), which pulls download logs of the RStudio (RStudio Team, 2015) CRAN mirror. Figure 1 shows the daily number of downloads of **tolerance** from the beginning of 2013 to the beginning of 2016. There is clearly a general increasing trend over the years as the average number of daily downloads per year is approximately 5, 10, and 15 in 2013, 2014, and 2015, respectively.

Capabilities of **tolerance** have been discussed in Young (2010, 2014). Even with those varied capabilities, perhaps the most commonly used methods involve the normal distribution. Normal tolerance intervals are often required during design verification or process validation. The utility of normal tolerance intervals is further highlighted in documents published by the EPA (Environmental Protection Agency, 2006), the IAEA (International Atomic Energy Agency, 2008), and standard 16269-6 of the ISO (International Organization for Standardization, 2014). In this paper, we discuss new capabilities in **tolerance** specifically involving normal tolerance intervals. This includes the calculation of exact and equal-tailed normal tolerance intervals, Bayesian normal tolerance intervals, tolerance intervals for fixed-effects ANOVA, and sample size determination strategies. We also introduce novel pseudo-operating characteristic (OC) curves that illustrate how the k-factor, sample size, confidence level, and content level each change relative to one another. Such curves can be useful for planning design tests.

As noted earlier, **tolerance** also includes a function for constructing multivariate normal tolerance regions. The `mvtol.region` function was included with the initial release of **tolerance**. `mvtol.region` includes several Monte Carlo procedures developed in Krishnamoorthy and Mathew (1999) and Krishnamoorthy and Mondal (2006) for finding the k-factor of the multivariate normal tolerance region. The `plottol` function can also be used to plot tolerance ellipses over bivariate normal data and tolerance ellipsoids over trivariate normal data. The latter is accomplished using `plot3d` from the **rgl** package (Adler and Murdoch, 2014). We will not discuss the `mvtol.region` function further since
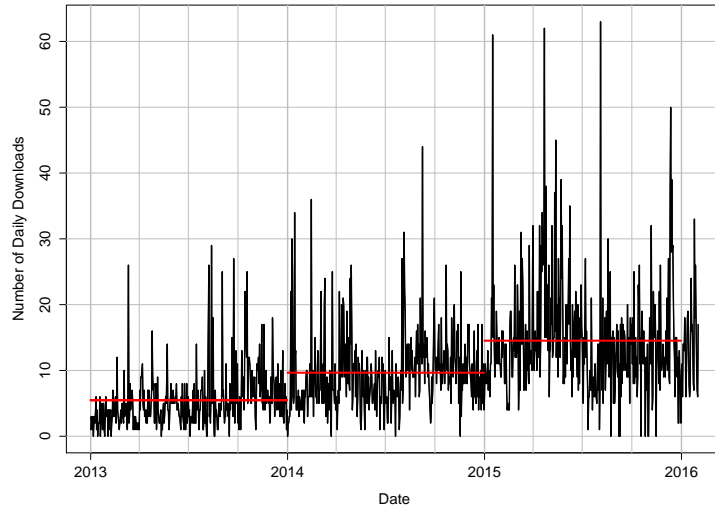
Figure 1: Number of daily downloads for **tolerance** from the RStudio CRAN mirror over a three-year time span (2013–2016).

it is already well-documented (Young, 2010, 2014) and our present focus is on newer capabilities in **tolerance** for normal tolerance intervals.

For our discussion, we assume that the reader has already installed and loaded **tolerance** using the usual commands:

```
> install.packages("tolerance")
> library(tolerance)
```

Normal tolerance intervals - classical and Bayesian

Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a random sample of continuous random variables that have cumulative distribution function F_X, which is parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$. Let $X \sim F_X$, independently of $\mathbf{X}$. In the classical set-up, a $(1 - \alpha, P)$ one-sided *upper tolerance limit* ($U_1(\mathbf{X})$) and one-sided *lower tolerance limit* ($L_1(\mathbf{X})$) satisfy the expressions

$$P_{\mathbf{X}}\left(P_X\left[X \le U_1(\mathbf{X})|\mathbf{X}\right] \ge P\right) = 1 - \alpha \tag{1}$$

and

$$P_{\mathbf{X}}\left(P_X\left[L_1(\mathbf{X}) \le X|\mathbf{X}\right] \ge P\right) = 1 - \alpha, \tag{2}$$

respectively. Similarly, a $(1 - \alpha, P)$ two-sided *tolerance interval*, $(L_2(\mathbf{X}), U_2(\mathbf{X}))$, satisfies

$$P_{\mathbf{X}}\left(P_X\left[L_2(\mathbf{X}) \le X \le U_2(\mathbf{X})|\mathbf{X}\right] \ge P\right) = 1 - \alpha. \tag{3}$$

Sometimes, controlling the proportion in the tails is required, in which case we have a $(1 - \alpha, P)$ *equal-tailed tolerance interval*, $(L_e(\mathbf{X}), U_e(\mathbf{X}))$, that satisfies

$$P_{\mathbf{X}}\left(\{P_X\left[L_e(\mathbf{X}) \le X|\mathbf{X}\right] \le (1 - P)/2\} \cap \{P_X\left[U_e(\mathbf{X}) \ge X|\mathbf{X}\right] \le (1 - P)/2\}\right) = 1 - \alpha. \tag{4}$$

Equal-tailed tolerance intervals ensure that we simultaneously have no more than $(1 - P)/2$ of the sampled population below the lower tolerance limit *and* no more than $(1 - P)/2$ of the sampled population above the upper tolerance limit.

Let $X_1, \dots, X_n$ be *iid* $\mathcal{N}(\mu, \sigma^2)$; i.e. a normal distribution with unknown mean μ and unknown variance σ^2. Let $\bar{X}$ and S^2 denote the sample mean and sample variance, respectively. The formulas for $(1 - \alpha, P)$ lower and upper normal tolerance limits are

$$L_h(\mathbf{X}) = \bar{X} - k_h(n, \alpha, P)S \quad \text{and} \quad U_h(\mathbf{X}) = \bar{X} + k_h(n, \alpha, P)S, \tag{5}$$

respectively, where $h \in \{1, 2, e\}$. In other words, h is an index specifying whether we want one-sided tolerance limits, two-sided tolerance intervals, or equal-tailed tolerance intervals. $k_1(n, \alpha, P)$ and $k_2(n, \alpha, P)$ are the *k-factors* for these two settings. The *k*-factor ensures that we capture at least a

proportion P of the sampled population with confidence level $(1 - \alpha)$. $k_1(n, \alpha, P)$ is calculated by

$$k_1(n, \alpha, P) = \frac{1}{\sqrt{n}} t_{n-1;1-\alpha}(\sqrt{n} z_P), \tag{6}$$

where $t_{f;q}(\delta)$ is the qth quantile of a noncentral t-distribution with f degrees of freedom and noncentrality parameter δ and z_q is the qth quantile of the standard normal distribution. $k_2(n, \alpha, P)$ is the solution to the integral equation

$$\sqrt{\frac{2n}{\pi}} \int_0^\infty P\left(\chi_{n-1}^2 > \frac{(n-1)\chi_{1;P}^2(z^2)}{k_2(n, \alpha, P)^2}\right) e^{-\frac{1}{2}nz^2} dz = 1 - \alpha, \tag{7}$$

where χ_f^2 is the chi-square random variable with f degrees of freedom and $\chi_{f;q}^2(\delta)$ is the qth quantile of the noncentral chi-squared distribution with f degrees of freedom and noncentrality parameter δ.

Owen (1964) was the first to propose equal-tailed tolerance intervals for the normal distribution. Equal-tailed normal tolerance intervals still take the form of (5), but where the tolerance factor $k_e(n, \alpha, P)$ is found as the solution to the integral equation

$$\left(2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)\right)^{-1} \int_{\frac{(n-1)\vartheta_n^2}{nk_e(n,\alpha,P)^2}}^\infty \left(2\Phi\left(-\vartheta_n + \frac{k_e(n, \alpha, P)\sqrt{nz}}{\sqrt{n-1}}\right) - 1\right) e^{-z/2} z^{\frac{n-1}{2}-1} dz = 1 - \alpha, \tag{8}$$

where $\vartheta_n = \sqrt{n} z_{\frac{1+P}{2}}$ and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. A general discussion comparing the utility of two-sided tolerance intervals versus equal-tailed tolerance intervals is found in Jensen (2009).

The `normtol.int` function in **tolerance** is able to calculate all of the one-sided tolerance limits, two-sided tolerance intervals, and equal-tailed tolerance intervals discussed above. In the past, challenges with computing noncentral distributions necessitated the use of approximations for $k_1(n, \alpha, P)$, $k_2(n, \alpha, P)$, and $k_e(n, \alpha, P)$. For the two-sided tolerance intervals, early versions of **tolerance** simply used various approximations that appeared in the literature over the years for computing the k-factors. These are controlled through the `method` argument and their specific formulas are outlined in Young (2010), which utilized **tolerance** version 0.2.2. Since then, the exact k-factor in (7) and the exact equal-tailed k-factor in (8) have been included. These are implemented by setting `method = "EXACT"` and `method = "OCT"`, respectively. Both of these methods use adaptive quadrature via the `integrate` function as well as box-constrained optimization via the `optim` function. The original approximation methods are still available primarily to retain all functionality of previous versions of **tolerance**.

The dataset that we will use to illustrate most of the procedures in our discussion is a quality control dataset from Krishnamoorthy and Mathew (2009). The data are from a machine that fills plastic containers with a liter of milk. At the end of a particular shift, a sample of $n = 20$ containers was selected and the actual amount of milk in each container was measured using a highly-accurate method. These measurements are as follows:

```
> milk <- c(0.968, 0.982, 1.030, 1.003, 1.046,
+           1.020, 0.997, 1.010, 1.027, 1.010,
+           0.973, 1.000, 1.044, 0.995, 1.020,
+           0.993, 0.984, 0.981, 0.997, 0.992)
```

A quick check of normality with the Shapiro-Wilk test confirms that this is an appropriate assumption:

```
> shapiro.test(milk)

        Shapiro-Wilk normality test

data:  milk
W = 0.96364, p-value = 0.6188
```

For the milk data, the $(0.95, 0.90)$ one-sided tolerance limits, two-sided tolerance interval, and equal-tailed tolerance interval are found as follows:

```
> normtol.int(x = milk, alpha = 0.05, P = 0.90, side = 1)
  alpha   P  x.bar 1-sided.lower 1-sided.upper
1  0.05 0.9 1.0036     0.9610333      1.046167

> normtol.int(x = milk, alpha = 0.05, P = 0.90, side = 2, method = "EXACT", m = 50)
  alpha   P  x.bar 2-sided.lower 2-sided.upper
1  0.05 0.9 1.0036     0.9523519      1.054848
```

```
> normtol.int(x = milk, alpha = 0.05, P = 0.90, side = 2, method = "OCT", m = 50)
  alpha   P  x.bar 2-sided.lower 2-sided.upper
1  0.05 0.9 1.0036       0.9471414      1.060059
```

Note that the equal-tailed tolerance interval is wider than the corresponding two-sided tolerance interval due to the more stringent requirement of controlling the proportions in the tails. For the two-sided tolerance intervals, the additional argument m is used to control the number of subintervals to use for performing the numerical integration via integrate. While not illustrated above, there is an additional argument that can be used if one wishes to construct log-normal tolerance intervals. The argument log.norm is a logical argument set to FALSE by default. If set to TRUE, log-normal tolerance intervals are calculated using the fact that if X is log-normally distributed, then $Y = \log(X)$ is normally distributed. Thus, the normtol.int function simply takes the logarithm of the data in the x argument, constructs the desired normal tolerance limits, and then takes the anti-log of those limits.

Users of normal tolerance intervals are often interested in summarizing a variety of possible k-factors for given sample sizes n, confidence levels $1 - \alpha$, and content level P. The K.table function allows the user to specify a vector of possible values for each of these three quantities. A list is then returned whose elements are summarized according to the by argument. For example, suppose we are interested in the $k_1(n, \alpha, P)$ values for all combinations of $n \in \{10, 20\}$, $\alpha \in \{0.01, 0.05\}$, and $P \in \{0.95, 0.99\}$. Moreover, we would like to summarize the list by the levels of P. This is accomplished as follows:

```
> K.table(n = c(10, 20), alpha = c(0.01, 0.05), P = c(0.95, 0.99),
+         side = 1, by.arg = "P")
$`0.95`
          10        20
0.99 3.738315 2.807866
0.95 2.910963 2.396002

$`0.99`
          10        20
0.99 5.073725 3.831558
0.95 3.981118 3.295157
```

For example, the first entry of the first matrix is the k-factor for a one-sided $(0.99, 0.95)$ tolerance limit when $n = 10$. One can also set side = 2, which requires the user to specify the method argument; e.g. "EXACT" for values of $k_2(n, \alpha, P)$ or "OCT for values of $k_e(n, \alpha, P)$. The by.arg argument can also be set to "alpha" or "n" depending on which quantity you want to represent the elements of the outputted list.

Bayesian tolerance intervals were first presented in Aitchison (1964). For the Bayesian set-up, let $\mathbf{x}$ be a vector of realizations of X, $\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})$ be the likelihood function, $\pi(\boldsymbol{\theta})$ be a prior distribution for $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta}|\mathbf{x})$ be the posterior distribution of $\boldsymbol{\theta}$ given by

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}|\mathbf{x})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{9}$$

Then, $(1 - \alpha, P)$ Bayesian one-sided upper and lower tolerance limits satisfy

$$P_{\boldsymbol{\Theta}}\left(P_X\left[X \leq U_1(\boldsymbol{\theta})|\boldsymbol{\theta}\right] \geq P|X\right) = 1 - \alpha \tag{10}$$

and

$$P_{\boldsymbol{\Theta}}\left(P_X\left[L_1(\boldsymbol{\theta}) \leq X|\boldsymbol{\theta}\right] \geq P|X\right) = 1 - \alpha, \tag{11}$$

respectively, a $(1 - \alpha, P)$ Bayesian two-sided tolerance interval satisfies

$$P_{\boldsymbol{\Theta}}\left(P_X\left[L_2(\boldsymbol{\theta}) \leq X \leq U_2(\boldsymbol{\theta})|\boldsymbol{\theta}\right] \geq P|X\right) = 1 - \alpha, \tag{12}$$

and a $(1 - \alpha, P)$ Bayesian equal-tailed tolerance interval satisfies

$$P_{\boldsymbol{\Theta}}\left(\{P_X\left[L_e(\boldsymbol{\theta}) \leq X|\boldsymbol{\theta}\right] \leq (1 - P)/2\} \cap \{P_X\left[U_e(\boldsymbol{\theta}) \geq X|\boldsymbol{\theta}\right] \leq (1 - P)/2\}|X\right) = 1 - \alpha. \tag{13}$$

Notice that the Bayesian set-up is calculated with respect to the probability measure $P_{\boldsymbol{\Theta}}$ while the classical set-up is calculated with respect to the distribution of the random sample $\mathbf{X}$. We refer the reader to the texts by Guttman (1970) and Krishnamoorthy and Mathew (2009) for more details on both classical and Bayesian tolerance intervals.

The parameters μ and σ^2 are still assumed unknown. We use the conjugate priors $\pi(\mu|\sigma^2)$ and

$\pi(\sigma^2)$, which are

$$\mu|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/n_0) \ \text{ and } \ \sigma^2 \sim \text{Scale-inv-}\chi^2(m_0, \sigma_0^2), \tag{14}$$

respectively, where Scale-inv-$\chi^2(\nu, \tau^2)$ is the scaled inverse chi-squared distribution with ν degrees of freedom and scaling parameter τ^2. Thus, the joint prior density of (μ, σ^2) is $\pi(\mu, \sigma^2) = \pi(\mu|\sigma^2)\pi(\sigma^2)$. The four hyperparameters for this prior structure are $\mu_0 \in \mathbb{R}, \sigma_0^2 > 0, m_0 > 0, n_0 > 0$. m_0 and n_0 are not prior sample size quantities, but are tunable quantities to reflect the prior precision *relative* to the sample size. The joint posterior distribution is then

$$p(\mu, \sigma^2|\mathbf{x}) = p(\mu|\sigma^2)p(\sigma^2), \tag{15}$$

where $p(\mu|\sigma^2)$ and $p(\sigma^2)$ are the distributions

$$\mu|\sigma^2 \sim \mathcal{N}\left(\bar{\bar{x}}, \frac{\sigma^2}{n_0 + n}\right) \ \text{ and } \ \sigma^2 \sim \text{Scale-inv-}\chi^2(m_0 + n - 1, q^2), \tag{16}$$

respectively, such that

$$\bar{\bar{x}} = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n} \ \text{ and } \ q^2 = (m_0 + n - 1)^{-1}\left[m_0\sigma_0^2 + (n-1)s^2 + \frac{n_0 n}{n_0 + n}(\bar{x} - \mu_0)^2\right]. \tag{17}$$

Note that the formulas in the Bayesian set-up are written such that they are conditioned on realizations of the observed data; i.e. $\mathbf{X} = \mathbf{x}$. Furthermore, they are written in terms of the sample estimates of the mean ($\bar{x}$) and variance (s^2). Additional details on the above can be found, for example, in Chapter 3 of Gelman et al. (2013).

Similar to the classical setting, $(1 - \alpha, P)$ Bayesian lower and upper normal tolerance limits are, respectively,

$$L_h(\bar{x}, s^2) = \bar{\bar{x}} - k_h(n, n_0, m_0, \alpha, P)q \ \text{ and } \ U_h(\bar{x}, s^2) = \bar{\bar{x}} + k_h(n, n_0, m_0, \alpha, P)q, \tag{18}$$

where, again, h is used as an index for one-sided limits, a two-sided interval, or an equal-tailed interval. Note that these limits are expressed in terms the maximum likelihood estimates of μ and σ^2, which occur through how $\bar{\bar{x}}$ and q are defined. Thus, the one-sided k-factor is calculated by

$$k_1(n, n_0, m_0, \alpha, P) = \frac{1}{\sqrt{n_0 + n}}t_{m_0+n-1;1-\alpha}(\sqrt{n_0 + n}z_P), \tag{19}$$

the two-sided k-factor $k_2(n, n_0, m_0, \alpha, P)$ is calculated by finding the solution to

$$\sqrt{\frac{2(n_0 + n)}{\pi}}\int_0^\infty P\left(\chi^2_{m_0+n-1} > \frac{(m_0 + n - 1)\chi^2_{1;P}(z^2)}{k_2(n, n_0, m_0, \alpha, P)^2}\right)e^{-\frac{1}{2}(n_0+n)z^2}dz = 1 - \alpha, \tag{20}$$

and the equal-tailed k-factor $k_e(n, n_0, m_0, \alpha, P)$ is calculated by finding the solution to

$$\frac{2^{-\left(\frac{m_0+n-1}{2}\right)}}{\Gamma\left(\frac{m_0+n-1}{2}\right)}\int_{\frac{(m_0+n-1)\vartheta_{n_0+n}^2}{(n_0+n)k_e(n,n_0,m_0,P)^2}}^\infty \left(2\Phi\left(-\vartheta_{n_0+n} + \frac{k_e(n, n_0, m_0, \alpha, P)\sqrt{n_0 + n}z}{\sqrt{m_0 + n - 1}}\right) - 1\right)$$

$$\times e^{-z/2}z^{\frac{m_0+n-1}{2}-1}dz = 1 - \alpha. \tag{21}$$

Finally, if one considers the non-informative prior distribution

$$\pi(\mu, \sigma^2) \propto \sigma^{-2}, \tag{22}$$

the solutions for the one-sided Bayesian normal tolerance limits and two-sided Bayesian normal tolerance intervals are the same as for the classical setting given in Equations (5)–(8); see Chapter 11 of Krishnamoorthy and Mathew (2009) for the details.

The bayesnormtol.int function for computing Bayesian normal tolerance intervals is new as of **tolerance** version 1.1.1. It was composed to closely mirror the normtol.int function. For the milk data, suppose we use the conjugate prior structure in (14). Assuming we have some historical knowledge about the milk filling process, the following hyperparameter values are used: $\mu_0 = 1.000, \sigma^2 = 0.001$, and $m_0 = n_0 = 20$. Then, the Bayesian $(0.95, 0.90)$ one-sided tolerance limits, two-sided tolerance interval, and equal-tailed tolerance interval are found as follows:

```
> bayesnormtol.int(x = milk, alpha = 0.05, P = 0.90, side = 1,
+                  hyper.par = list(mu.0 = 1.000, sig2.0 = 0.001,
+                                   m.0 = 20, n.0 = 20))
```

```
   alpha   P 1-sided.lower 1-sided.upper
1  0.05 0.9     0.9551936      1.048406


> bayesnormtol.int(x = milk, alpha = 0.05, P = 0.90, side = 2, method = "EXACT",
+               m = 50, hyper.par = list(mu.0 = 1.000, sig2.0 = 0.001,
+                                        m.0 = 20, n.0 = 20))
   alpha   P 2-sided.lower 2-sided.upper
1  0.05 0.9     0.9453603       1.05824


> bayesnormtol.int(x = milk, alpha = 0.05, P = 0.90, side = 2, method = "OCT",
+               m = 50, hyper.par = list(mu.0 = 1.000, sig2.0 = 0.001,
+                                        m.0 = 20, n.0 = 20))
   alpha   P 2-sided.lower 2-sided.upper
1  0.05 0.9     0.9407625      1.062838
```

The bayesnormtol.int function has the arguments x, alpha, P, side, method, and m just as in the normtol.int. However, here we also have hyper.par, which is a list with elements for the four hyperparameters. The output is structured identically to the output obtained using normtol.int, which allows for easy comparison between the classical and Bayesian results.

Fixed-effects ANOVA tolerance intervals

The approach for classical normal tolerance intervals can be easily extended for the balanced fixed-effects ANOVA model

$$Y_{ij...kl} = \theta + \alpha_i + \beta_j + \ldots + \gamma_k + \epsilon_{ij...kl}, \tag{23}$$

where θ is the grand mean, $\alpha_i, \beta_j, \ldots, \gamma_k$ are the factor effects each subject to the constraint that the summation of the effects over the respective index is equal to 0, $\epsilon_{ij...kl}$ are iid $\mathcal{N}(0, \sigma^2)$ error terms, and the indices are $i = 1, \ldots, a, j = 1, \ldots, b, \ldots, k = 1, \ldots, c$, and $l = 1, \ldots, n$. The approach discussed below is for the classical setting. Currently, the **tolerance** package does not have a function for calculating Bayesian ANOVA tolerance intervals.

Let $\mathbf{Y} \in \mathbb{R}^{ab\cdots cn}$ be a vector of all of the measured responses in (23), which are iid $\mathcal{N}(0, \sigma^2)$. The formulas for the tolerance limits in the fixed-effects ANOVA setting are:

$$
\begin{aligned}
L_{i;h}(\mathbf{Y}) &= \bar{Y}_{i\cdot\ldots\cdot\cdot} - k_h(n_i, f, \alpha, P)\sqrt{MSE} & \text{and} && U_{i;h}(\mathbf{Y}) &= \bar{Y}_{i\cdot\ldots\cdot\cdot} + k_h(n_i, f, \alpha, P)\sqrt{MSE} \\
L_{j;h}(\mathbf{Y}) &= \bar{Y}_{\cdot j\ldots\cdot\cdot} - k_h(n_j, f, \alpha, P)\sqrt{MSE} & \text{and} && U_{j;h}(\mathbf{Y}) &= \bar{Y}_{\cdot j\ldots\cdot\cdot} + k_h(n_j, f, \alpha, P)\sqrt{MSE} \\
&\vdots & && &\vdots \\
L_{k;h}(\mathbf{Y}) &= \bar{Y}_{\cdot\ldots\cdot k\cdot} - k_h(n_k, f, \alpha, P)\sqrt{MSE} & \text{and} && U_{k;h}(\mathbf{Y}) &= \bar{Y}_{\cdot\ldots\cdot k\cdot} + k_h(n_k, f, \alpha, P)\sqrt{MSE}
\end{aligned} \tag{24}
$$

Conceptually, the formulas in (24) are similar to those in (5). We take a point estimate of the mean at each factor level (i.e. the quantities $\bar{Y}_{i\cdot\ldots\cdot\cdot}, \bar{Y}_{\cdot j\ldots\cdot\cdot}, \ldots, \bar{Y}_{\cdot\ldots\cdot k\cdot}$) and then add or subtract the k-factor times the standard deviation. The standard deviation is now estimated by the root mean square error, $\sqrt{MSE}$.

The k-factor in (24), again, has the subscript h to indicate an index for one-sided limits, a two-sided interval, or an equal-tailed interval. However, the formulas are modified for the ANOVA setting. In formulas (6)–(8), the quantity $(n - 1)$ reflects the degrees of freedom when estimating the sample variance S^2. In the ANOVA setting, this is replaced by the degrees of freedom due to the error; i.e. the degrees of freedom associated with the MSE. Thus, we replace each occurrence of $(n - 1)$ in (6)–(8) with f, the error degrees of freedom. Moreover, all occurrences of the sample size n are replaced with the number of observations at each factor level; i.e. $n_i, n_j, \ldots, n_k$. Note that the tolerance intervals presented are only accurate for balanced (or nearly-balanced) ANOVA settings.

We analyze the well-known dataset that resulted from an experimental design regarding the effects of wool type and the amount of tension applied to a loom of yarn on the number of warp breaks that occur on that loom of yarn (Tippett, 1950). The first factor is wool type, which has two levels: A or B. The second factor is tension level, which has three levels: low (L), medium (M), or high (H). The six treatments (i.e. factor level combinations) are randomly assigned to one of 54 looms of yarn. Thus, we have $n = 9$ replicates per treatment. Suppose we want to construct $(0.85, 0.90)$ equal-tailed tolerance intervals for each factor level's mean. Below is how we would do this in the **tolerance** package:

```
> lm.out <- lm(breaks ~ wool + tension, data = warpbreaks)
> out <- anovatol.int(lm.out, data = warpbreaks, alpha = 0.10,
+                     P = 0.85, side = 2, method = "OCT")
```

```
These are 90%/85% 2-sided tolerance intervals.
> out
$wool
     mean  n        k 2-sided.lower 2-sided.upper
A 31.03704 27 1.886857      9.117165      52.95691
B 25.25926 27 1.886857      3.339387      47.17913

$tension
     mean  n        k 2-sided.lower 2-sided.upper
L 36.38889 18 1.948567     13.7521219      59.02566
M 26.38889 18 1.948567      3.7521219      49.02566
H 21.66667 18 1.948567     -0.9701003      44.30343
```

In the `anovatol.int` function, we have similar arguments as in the `normtol.int` function, except now we input an object of class `"lm"` and we also tell the function the name of the original dataset using the data argument. The output is a list summarizing the tolerance interval results for each factor level. For example, the $(0.90, 0.85)$ equal-tailed tolerance interval for the medium tension applied to the yarn is about $(3.75, 49.03)$.
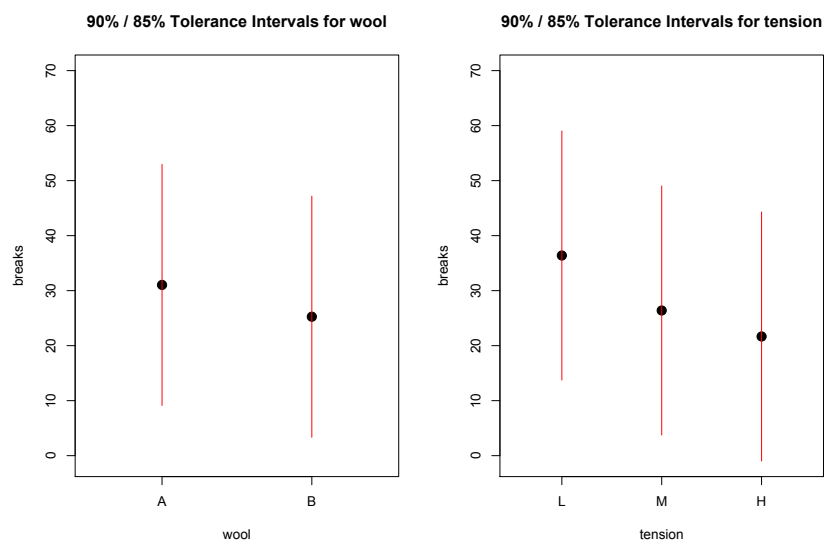
Figure 2: Plot of the $(0.85, 0.90)$ equal-tailed tolerance intervals for the yarn strength data.

We can also produce a figure of the above output by using the `plottol` function as follows:

```
> plottol(out, x = warpbreaks)
```

The above produces the plot in Figure 2. This figure has a separate panel for each factor. The *y*-axis is the response and the *x*-axis is the levels of the respective factor. The solid black point is the factor level mean and the red lines extend to the lower and upper tolerance limits calculated earlier. Such a figure provides a relative comparison of the tolerance intervals for each factor level.

Sample size determination strategies

As noted in Faulkenberry and Weeks (1968), an important question for statistical practitioners is "What sample size should be used to determine the tolerance limits?" Those same authors addressed this problem by developing an approach to ensure that the calculated tolerance intervals are "close" to the quantiles that result in a content level at least as large as P. Their solution was developed for sample size determination of one-sided tolerance limits and two-sided tolerance intervals, but it is not applicable to equal-tailed tolerance intervals. In order to briefly present their approach, let $C_{\mu,\sigma}(\mathbf{X})$ denote any of the three inner probabilities in Equations (1)–(3) or any of the three analogous inner probabilities for the Bayesian set-up in Equations (10)–(12). To ensure the "goodness" of the tolerance limits (interval), one must choose an arbitrary $P' > P$ and small $\delta > 0$ to determine a sample size n^* such that

$$P_{\mathbf{X}}\left\{ C_{\mu,\sigma}(\mathbf{X}) \geq P' \right\} \geq \delta \tag{25}$$

or

$$P_{\boldsymbol{\theta}} \left\{ C_{\mu,\sigma}(\mathbf{X}) \geq P' \right\} \geq \delta \tag{26}$$

for the classical or Bayesian setting, respectively.

The `norm.ss` function for sample size determination of normal tolerance limits (intervals) is new as of **tolerance** version 1.1.1. The function finds the minimum sample size n^* for the approach due to Faulkenberry and Weeks (1968) discussed above. For our example, suppose the quality engineer overseeing the milk-filling process wants to submit a future sample of liters of milk to the highly-accurate measurement method. Per the company's guidelines, the engineer needs to know the minimum sample size to construct a $(0.95, 0.90)$ two-sided tolerance interval such that $P_{\mathbf{X}} \left\{ C_{\mu,\sigma}(\mathbf{X}) \geq 0.97 \right\} \geq 0.10$. This is calculated as follows:

```
> norm.ss(alpha = 0.05, P = 0.90, delta = 0.10, P.prime = 0.97,
+         side = 2, m = 50, method = "FW")
  alpha   P delta P.prime  n
1  0.05 0.9   0.1    0.97 60
```

Thus, the engineer would need a minimum sample size of $n^* = 60$ to ensure that there is only a small probability $\delta = 0.10$ that the $(0.95, 0.90)$ tolerance interval will have a content exceeding $P = 0.90$ by $(P' - P) = 0.07$.

In the `norm.ss` function, the argument `method` is set to `"FW"`. There are two additional sample size determination strategies that can be calculated, which are controlled through the `method` argument. Both of these strategies assume there is some historical data and specification limits for the process at-hand. We briefly illustrate these strategies below and refer the reader to Young et al. (2016) for further details.

The first alternative strategy is a simple *back-of-the-envelope* calculation. We consider the problem of designing a study to demonstrate that a process or product falls within the specification limits (S_L, S_U). We are interested in the minimum sample size necessary such that a $(1 - \alpha, P)$ one-sided lower tolerance limit exceeds S_L, a $(1 - \alpha, P)$ one-sided upper tolerance limit falls below S_U, or a $(1 - \alpha, P)$ two-sided tolerance interval is contained within (S_L, S_U). In other words, this requires finding the minimum sample size n^* such that

$$S_L < \mu - k_1(n, \alpha, P)\sigma; \tag{27}$$
$$S_U > \mu + k_1(n, \alpha, P)\sigma; \text{ or} \tag{28}$$
$$\mu \pm k_e(n, \alpha, P)\sigma \subset (S_L, S_U), \tag{29}$$

for one-sided upper tolerance limits, one-sided lower tolerance limits, or equal-tailed tolerance intervals, respectively. As emphasized in Young et al. (2016), this approach is intended simply for planning purposes and it does not guarantee any specific bounds relative to the nominal coverage probability. Note that (29) is for an equal-tailed tolerance interval since we posit values for μ and σ and, thus, the resulting tolerance interval would be built around a (hypothetically) true center of the normal population.

For our example, suppose that the quality engineer is overseeing the launch of a new process for filling the one-liter containers of milk, which is intended to be more accurate than the previous process. The company set specification limits at $(0.990, 1.010)$. For determining the minimum sample size necessary to construct a $(0.95, 0.90)$ two-sided tolerance interval that is within the specification limits, the engineer assumes the mean and variance from the data of the original process. This calculation can then be done as follows:

```
> norm.ss(alpha = 0.05, P = 0.90, side = 2, spec = c(0.990, 1.010),
+         method = "DIR", hyper.par = list(mu.0 = 1.004, sig2.0 = 0.001))
  alpha   P delta P.prime n
1  0.05 0.9               5
```

Thus, the minimum sample size is $n^* = 5$. This calculation was done by setting `method = "DIR"`, entering the specification limits in the `spec` argument, and entering the assumed μ and σ^2 in the argument `hyper.par`.

The second alternative strategy presented in Young et al. (2016) is a method for providing data-dependent values of the precision quantities P' and δ in the approach due to Faulkenberry and Weeks (1968). The approach assumes there is information on historical data, a set of current data, and specification limits that can be used for calculating values of P' and δ. The approach is intended to be used when there is no practical guidance for setting these values other than using *rule-of-thumb* quantities suggested in Faulkenberry and Weeks (1968).

For the milk filling process, suppose the engineer has historical measurements, which have a combined mean 0.994 liters and variance 0.002. Suppose the specification limits on the original process
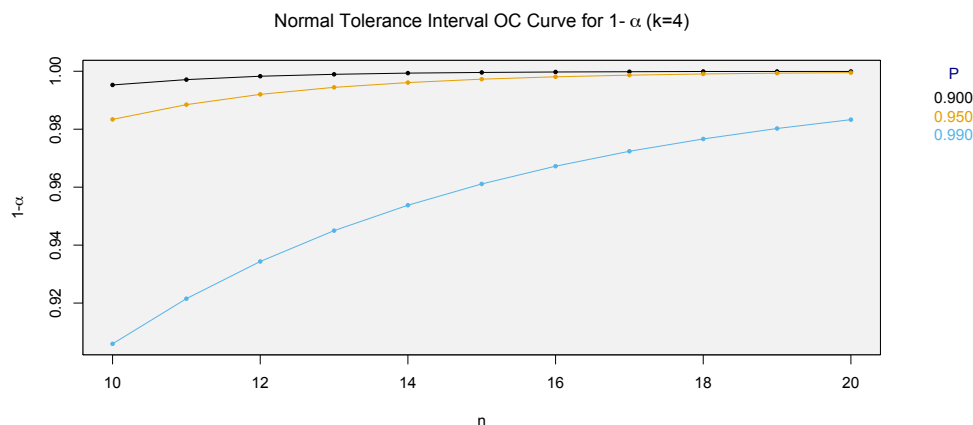
Figure 3: OC curves for $1 - \alpha$ given the set of content levels $P \in \{0.90, 0.95, 0.99\}$, sample sizes $n = 10, 11, \ldots, 20$, and a two-sided k-factor of 4.

are $(0.900, 1.100)$ and that the engineer needs a minimum sample size to construct a $(0.95, 0.90)$ two-sided tolerance interval to show that the process meets the specification limits. However, the engineer is unsure about levels to choose for δ and P'. We can use the `norm.ss` function as follows:

```
> norm.ss(x = milk, alpha = 0.05, P = 0.90, side = 2, spec = c(0.900, 1.100),
+         method = "YGZO", hyper.par = list(mu.0 = 0.994, sig2.0 = 0.002))
  alpha  P    delta   P.prime  n
1  0.05 0.9 0.1807489 0.9733307 42
```

Thus, the engineer would need a minimum sample size of $n^* = 42$ to ensure that there is only a probability of about $\delta = 0.181$ that the $(0.95, 0.90)$ tolerance interval will have a content exceeding $P = 0.90$ by about $(P' - P) = 0.073$.

OC curves involving k-factors

Sometimes, engineers and industrial statisticians are interested in understanding how the confidence level or content level changes as a function of n for a given level of the k-factor. If one has normally distributed data that they intend to demonstrate meets certain specification limits, then it is important to understand the type of values for $1 - \alpha$ and P that one can reasonably expect to use. In this section, we present OC curves for such planning purposes.

The first type of OC curve is used when one specifies a range of values of the sample size n and a target value of the k-factor. Then, one can either specify a set of $1 - \alpha$ values and solve for P or one can specify a set of P values and solve for $1 - \alpha$. The values of n are plotted on the x-axis and the value being solved for – either P or $1 - \alpha$ – is plotted on the y-axis. The different OC curves pertain to the set of specified values – either $1 - \alpha$ or P. Since too many curves can become cumbersome, we have placed an upper limit of 10 curves that can be overlaid on a given plot. Also, the colors used for the curves were chosen using a colorblind-friendly palette that was established by Okabe and Ito (2002).

Suppose a company is designing a product and the engineer needs to collect enough data so that the resulting two-sided tolerance interval will have a k-factor of 4. Content levels under consideration are $P \in \{0.90, 0.95, 0.99\}$ while the possible number of samples that can be used for the test are $n = 10, 11, \ldots, 20$. In order to determine the resulting confidence levels that can be obtained under these conditions, the engineer can construct an OC curve for $1 - \alpha$ using the following code:

```
norm.OC(k = 4, alpha = NULL, P = c(0.90, 0.95, 0.99), n = 10:20,
        side = 2, method = "EXACT", m = 25)
```

The resulting plot is given in Figure 3. For example, if the engineer chooses $n = 15$, then they can construct a two-sided tolerance interval with $k = 4$ and content level of $P = 0.99$ with confidence level near 0.96. However, if the engineer wishes to decrease the content of the tolerance interval to $P = 0.95$ or $P = 0.90$, then a confidence level very near 1 can be achieved.

In the `norm.OC` function, the arguments of `side`, `method`, and `m` are, again, passed down to the underlying `K.factor` function. In order to generate Figure 3, we need to specify a single value for `k` (i.e. the k-factor) and at least one value for P. Since we are constructing curves where the sample size is on the x-axis, we need at least two values for n. Note that `alpha` must be left at its default NULL value.
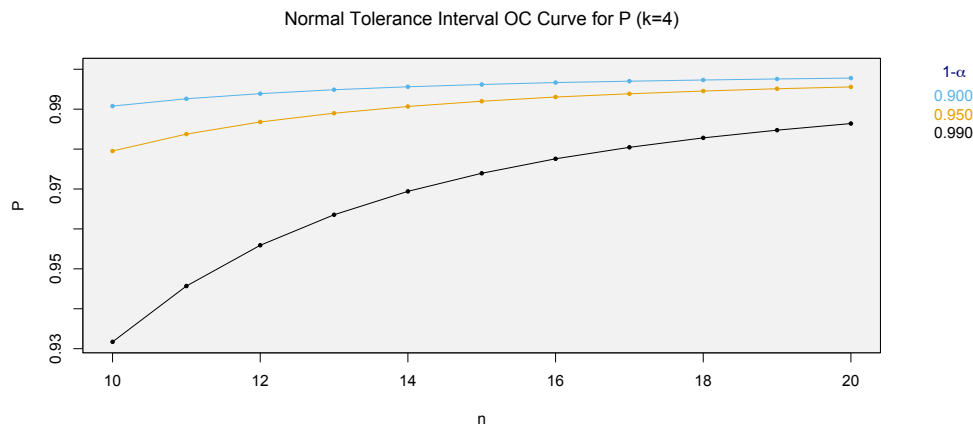
Figure 4: OC curves for P given the set of confidence levels $1 - \alpha \in \{0.90, 0.95, 0.99\}$, sample sizes $n = 10, 11, \ldots, 20$, and a two-sided k-factor of 4.

Suppose now that the same engineer considers confidence levels of $1 - \alpha \in \{0.90, 0.95, 0.99\}$ with the same values of k and n from before. In order to determine the resulting content levels that can be obtained under these conditions, the engineer can construct an OC curve for P using the following code:

```
norm.OC(k = 4, alpha = c(0.01, 0.05, 0.10), P = NULL, n = 10:20,
        side = 2, method = "EXACT", m = 25)
```

The resulting plot is given in Figure 4. For example, if the engineer chooses $n = 12$, then they can construct a two-sided tolerance interval with $k = 4$ and confidence level of $1 - \alpha = 0.99$ that captures about 95% of the sampled population. However, if the engineer wishes to decrease the confidence level of the tolerance interval to $1 - \alpha = 0.95$ or $1 - \alpha = 0.90$, then a tolerance interval that captures about 99% of the sampled population can be achieved. Note that the code using the `norm.OC` function is similar to the previous example, except that now we specify at least one value for `alpha` and leave P must at its default `NULL` value.

Finally, the `norm.OC` function can also be used to construct an OC-curve where the k-factor is calculated for specified values of n, $1 - \alpha$, and P. The different curves will be for each combination of the specified $1 - \alpha$ and P levels. For our example, suppose the engineer is interested in the k-factors for two-sided tolerance intervals for the set of confidence levels $1 - \alpha \in \{0.90, 0.95, 0.99\}$, the set of content levels $P \in \{0.90, 0.95, 0.99\}$, and sample sizes $n = 10, 11, \ldots, 20$. Then, we can specify the respective arguments in the `norm.OC` function while leaving the k argument at its default `NULL` value:

```
norm.OC(k = NULL, P = c(0.90, 0.95, 0.99), alpha=c(0.01,0.05,0.10),
        n = 10:20, side = 2, method = "EXACT", m = 25)
```

The resulting plot is given in Figure 5. This OC-curve allows the user to assess the width of the tolerance interval as n changes for the given $(1 - \alpha, P)$ tolerance levels.

Summary

tolerance is the only R package devoted to the calculation of tolerance intervals and regions. Since its earlier versions (Young, 2010), there have been many updates to the package that include additional parametric tolerance interval procedures, improved nonparametric tolerance interval procedures, and some multivariate tolerance region procedures.

In this paper, we focused on the varied capabilities of **tolerance** pertaining to tolerance intervals for the normal distribution. Many of these procedures have been added to the package since the discussion presented in Young (2010). We discussed the calculation of one-sided normal tolerance limits, exact and equal-tailed normal tolerance intervals, Bayesian normal tolerance intervals, tolerance intervals for fixed-effects ANOVA, and sample size determination strategies. We also introduced novel operating characteristic (OC) curves that illustrate how the k-factor, sample size, confidence level, and content level each change relative to one another. As pointed out throughout our discussion, all of these procedures have a large degree of utility in a variety of practical contexts.

The **tolerance** package continues to expand the functions available for constructing tolerance intervals and regions. We note that some of the updates over the years have been a direct result of
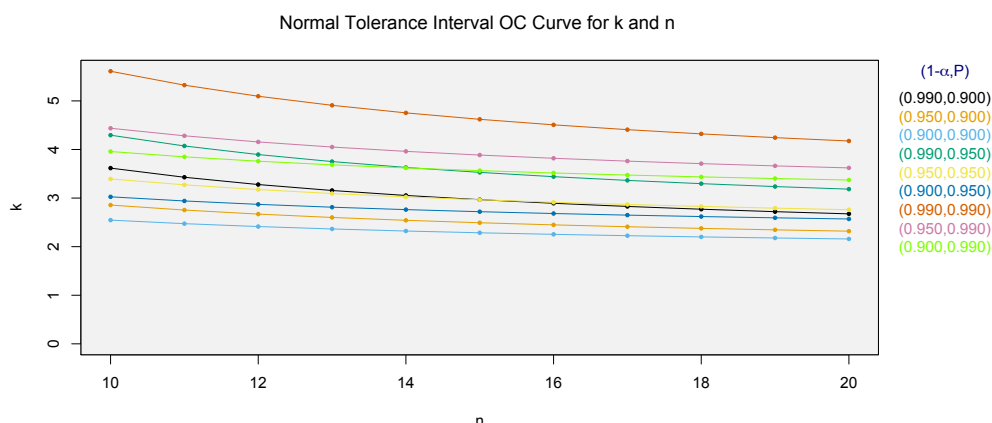
Figure 5: OC curves for the k-factor when given the set of confidence levels $1 - \alpha \in \{0.90, 0.95, 0.99\}$, the set of content levels $P \in \{0.90, 0.95, 0.99\}$, and sample sizes $n = 10, 11, \ldots, 20$.

requests by end users of the package. Thus, one can expect additional capabilities in future versions of **tolerance**, both for the normal setting and other data settings.

Bibliography

D. Adler and D. Murdoch. *rgl: 3D visualization device system (OpenGL)*, 2014. URL http://CRAN.R-project.org/package=rgl. R package version 0.95.1201. [p1]

J. Aitchison. Bayesian Tolerance Regions. *Journal of the Royal Statistical Society, Series B*, 26(2):161–175, 1964. [p4]

T. Burr and A. Gavron. Pass/Fail Criterion for a Simple Radiation Portal Monitor Test. *Modern Instrumentation*, 1(3):27–33, 2010. [p1]

G. Csardi. *cranlogs: Download Logs from the 'RStudio' 'CRAN' Mirror*, 2015. URL http://CRAN.R-project.org/package=cranlogs. R package version 2.1.0. [p1]

Environmental Protection Agency. *Data Quality Assessment: Statistical Methods for Practitioners*. U.S. Environmental Protection Agency, Washington, DC, USA, 2006. URL http://www.epa.gov/sites/production/files/2015-08/documents/g9s-final.pdf. [p1]

G. D. Faulkenberry and D. L. Weeks. Sample size determination for tolerance limits. *Technometrics*, 10 (2):343–348, 1968. [p7, 8]

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, 3rd edition, 2013. [p5]

I. Guttman. *Statistical Tolerance Regions: Classical and Bayesian*. Charles Griffin and Company, London, 1970. [p1, 4]

S. D. Hafner, C. Howard, R. E. Muck, R. B. Franco, F. Montes, P. G. Green, F. Mitloehner, S. L. Trabue, and C. A. Rotz. Emission of Volatile Organic Compounds from Silage: Compounds, Sources, and Implications. *Atmospheric Environment*, 77:827–839, 2013. [p1]

G. J. Hahn and W. Q. Meeker. *Statistical Intervals: A Guide for Practitioners*. Wiley-Interscience, New York, NY, 1991. [p1]

I. J. Hall. Approximate One-Sided Tolerance Limits for the Difference or Sum of Two Independent Normal Variates. *Journal of Quality Technology*, 16(1):15–19, 1984. [p1]

M. M. Heck, M. Retz, M. Bandu, M. Souchay, E. Vitzthum, G. Weirich, M. Mollenhauer, T. Schuster, M. Autenrieth, H. Kübler, T. Maurer, M. Thalgott, K. Herkommer, J. E. Gschwend, and R. Nawroth. Topography of Lymph Node Metastases in Prostate Cancer Patients Undergoing Radical Prostatectomy and Extended Lymphadenectomy: Results of a Combined Molecular and Histopathologic Mapping Study. *European Urology*, 66(2):222–229, 2014. [p1]

International Atomic Energy Agency. *Safety Report Series No. 52: Best Estimate Safety Analysis for Nuclear Plants: Uncertainty Evaluation*. IAEA Publishing Section, Vienna, Austria, 2008. URL http://www-pub.iaea.org/MTCD/publications/PDF/Pub1306_web.pdf. [p1]

International Organization for Standardization. *ISO 16269-6: Statistical Interpretation of Data – Part 6: Determination of Statistical Tolerance Intervals*. International Organization for Standardization, Geneva, Switzerland, 2014. URL http://www.iso.org/iso/catalogue_detail.htm?csnumber=57191. [p1]

W. A. Jensen. Approximations of Tolerance Intervals for Normally Distributed Data. *Quality and Reliability Engineering International*, 25(5):571–580, 2009. [p3]

K. Krishnamoorthy and T. Mathew. Comparison of approximation methods for computing tolerance factors for a multivariate normal population. *Technometrics*, 41(3):234–249, 1999. [p1]

K. Krishnamoorthy and T. Mathew. *Statistical Tolerance Regions: Theory, Applications, and Computation*. Wiley, Hoboken, NJ, 2009. [p1, 3, 4, 5]

K. Krishnamoorthy and S. Mondal. Improved tolerance factors for multivariate normal distributions. *Communications in Statistics - Simulation and Computation*, 35(2):461–478, 2006. [p1]

T. Mathew and D. S. Young. Fiducial-Based Tolerance Intervals for Some Discrete Distributions. *Computational Statistics and Data Analysis*, 61:38–49, 2013. [p1]

M. Okabe and K. Ito. Color Universal Design (CUD) - How to Make Figures and Presentations that are Friendly to Colorblind People, 2002. URL http://jfly.iam.u-tokyo.ac.jp/color/. [p9]

D. B. Owen. Control of Percentages in Both Tails of the Normal Distribution. *Technometrics*, 6(4): 377–387, 1964. [p3]

C. Pasquaretta, G. Bogliani, L. Ranghetti, C. Ferrari, and A. von Hardenberg. The Animal Locator: A New Method for Accurate and Fast Collection of Animal Locations for Visible Species. *Wildlife Biology*, 18(2):202–214, 2012. [p1]

RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL http://www.rstudio.com/. [p1]

L. H. C. Tippett. *Technological Applications of Statistics*. Wiley, New York, NY, 1950. [p6]

K. Van der Borght, G. Verbeke, and H. van Vlijmen. Multi-Model Inference Using Mixed Effects from a Linear Regression Based Genetic Algorithm. *BMC Bioinformatics*, 15(88):1–11, 2014. [p1]

A. Wald. An Extension of Wilks' Method for Setting Tolerance Limits. *Annals of Mathematical Statistics*, 14(1):45–55, 1943. [p1]

S. S. Wilks. Determination of Sample Sizes for Setting Tolerance Limits. *The Annals of Mathematical Statistics*, 12(1):91–96, 1941. [p1]

S. S. Wilks. Statistical Prediction with Special Reference to the Problem of Tolerance Limits. *The Annals of Mathematical Statistics*, 13(4):400–409, 1942. [p1]

D. S. Young. tolerance: An R Package for Estimating Tolerance Intervals. *Journal of Statistical Software*, 36(5):1–39, 2010. URL http://www.jstatsoft.org/v36/i05/. [p1, 2, 3, 10]

D. S. Young. Computing Tolerance Intervals and Regions Using R. In M. B. Rao and C. R. Rao, editors, *Handbook of Statistics, Volume 32: Computational Statistics with R*, pages 309–338. North Holland - Elsevier, Amsterdam, Netherlands, 2014. [p1, 2]

D. S. Young and T. Mathew. Ratio Edits Based on Statistical Tolerance Intervals. *Journal of Official Statistics*, 31(1):77–100, 2015. [p1]

D. S. Young, C. M. Gordon, S. Zhu, and B. D. Olin. Sample Size Determination Strategies for Normal Tolerance Intervals Using Historical Data. *Quality Engineering*, 28(3):335–349, 2016. [p8]

Derek S. Young
Department of Statistics
University of Kentucky
323 Multidisciplinary Science Building
725 Rose Street
Lexington, KY 40536-0082 USA
derek.young@uky.edu