

Tolerance Stack Analysis Methods

A Critical Review

Fritz Scholz*
Research and Technology
Boeing Information & Support Services

November 1995

Abstract

This document reviews various ways of performing tolerance stack analyses. This review is limited to assembly criteria which are linear or approximately linear functions of the relevant part dimensions. Beginning with the two extreme cornerstones, namely the arithmetic or worst case tolerance stack and the statistical or RSS tolerance stack method, various compromises or unifying paradigms are presented with their underlying assumptions and rationales. These cover distributions more dispersed than the commonly assumed normal distribution and shifts in the means. Both worst case and statistical stacking of mean shifts are discussed. The latter, in the form presented here, appears to be new. The appropriate methods for assessing nonassembly risk are indicated in each case.

* P.O. Box 3707, MS 7L-22, Seattle WA 98124-2207, e-mail: fritz.scholz@grace.rt.cs.boeing.com

Contents

1	Introduction and Overview	3
2	Notation and Conventions	4
3	Arithmetic Tolerance Stack (Worst Case)	7
4	Statistical Tolerancing (RSS Method)	9
4.1	Statistical Tolerancing With Normal Variation	9
4.2	Statistical Tolerancing Using the CLT	14
4.3	Risk Assessment with Statistical Tolerance Stacking	18
5	Mean Shifts	20
5.1	Arithmetic Stacking of Mean Shifts	24
5.2	Risk Assessment with Arithmetically Stacked Mean Shifts . .	31
5.3	Statistical Stacking of Mean Shifts	34
5.4	Statistical Stacking of Mean Shifts Revisited	40

1 Introduction and Overview

Tolerance stacking analysis methods are described in various texts and papers, see for example Gilson (1951), Mansoor (1963), Fortini (1967), Evans (1975), Cox (1986), Greenwood and Chase (1987), Kirschling (1988), Bjørke (1989), Henzold (1995), and Nigam and Turner (1995). Unfortunately, the notation is often not standard and not uniform, making the understanding of the material at times difficult.

Invariably the discussion includes the two cornerstones, arithmetic and statistical tolerancing. This examination is no exception, since these two methods provide conservative and optimistic benchmarks, respectively. In the basic statistical tolerancing scheme it is assumed that part dimension vary randomly according to a normal distribution, centered at the tolerance interval midpoint and with its $\pm 3\sigma$ spread covering the tolerance interval. For given part dimension tolerances this kind of analysis typically leads to much tighter assembly tolerances, or for given assembly tolerance it requires considerably less stringent part dimension tolerances.

Since practice has shown that the results are usually not quite as good as advertised, one has tried to relax the above distributional assumptions in a variety of ways. One way is to allow other than normal distributions which essentially cover the tolerance interval with a wider spread, but which are still centered on the tolerance interval midpoint. This results in somewhat less optimistic gains than those obtained under the normality assumptions, but usually still much better than those given by arithmetic tolerancing, especially for longer tolerance chains.

Another relaxation concerns the centering of the distribution on the tolerance interval midpoint. The realization that it is difficult to center any process exactly where one wants it to be has led to several mean shift models in which the distribution may be centered anywhere within a certain neighborhood around the tolerance interval midpoint, but usually it is still assumed that the distribution is normal and its $\pm 3\sigma$ spread is still within the tolerance limits. This means that while we allow some shift in the mean we require a simultaneous reduction in variability. The mean shifts are then stacked in worst case fashion. The correspondingly reduced variation of the shifted distributions is stacked statistically. The overall assembly tolerance then becomes (in worst case fashion) a sum of two parts, consisting of an arithmetically stacked mean shift contribution and a term reflecting the sta-

tistically stacked distributions describing the parts variation. It turns out that our corner stones of arithmetic and statistical tolerancing are subparts of this more general model, which has been claimed to unify matters.

However, there is another way of dealing with mean shifts which appears to be new, at least in the form presented here. It takes advantage of statistical stacking of mean shifts and stacking that in worst case fashion with the statistical stacking of the reduced variation in the part dimension distributions. A precursor to this can be found in Desmond's discussion of Mansoor's (1963) paper. However, there it was pointed out that it leads to optimistic results. The reason for this was a flaw in handling the reduction of the part dimension variation caused by the random mean shifts.

When dealing with tolerance stacking under mean shifts one has to take special care in assessing the risk of nonassembly. Typically only one tail of the assembly stack distribution is significant when operating at one of the two worst possible assembly mean shifts. For this reason the method of risk calculations are discussed in detail, where appropriate.

2 Notation and Conventions

The tolerance stacking problem arises because of the inability to produce parts exactly according to nominal. Thus there is the possibility that the assembly of such interacting parts will not function or won't come together as planned. This can usually be judged by one or more assembly criteria, say A_1, A_2, \dots

Here we will be concerned with just one such assembly criterion, say A , which can be viewed as a function of the part dimensions X_1, \dots, X_n , i.e.,

$$A = f(X_1, \dots, X_n) .$$

Here n may be the number of parts involved in the assembly, but n may also be larger than that, namely when some parts contribute more than one dimension to the assembly criterion A . Ideally the part dimensions should be equal to their respective nominal values ν_1, \dots, ν_n . Recognizing the inevitability of part variation from nominal one allows the part dimension X_i to vary over an interval around ν_i . Typically one specifies an interval symmetric around the nominal value, i.e., $I_i = [\nu_i - T_i, \nu_i + T_i]$. However, asymmetric tolerance intervals do occur and in the most extreme form they become unilateral tolerance intervals, e.g., $I_i = [\nu_i - T_i, \nu_i]$ or $I_i = [\nu_i, \nu_i + T_i]$. Most

generally one would specify a tolerance interval $I_i = [c_i, d_i]$ with $c_i \leq \nu_i \leq d_i$. When dealing with a symmetric or unilateral tolerance interval one calls the value T_i the **tolerance**. For the most general bilateral tolerance interval, $I_i = [c_i, d_i]$, one would have two tolerances, namely $T_{1i} = \nu_i - c_i$ and $T_{2i} = d_i - \nu_i$. Although asymmetrical tolerance intervals occur in practice, they are usually not discussed much in the literature. The tolerance stacking principles apply in the asymmetric case as well but the analysis and exposition tends to get messy. We will thus focus our review on the symmetric case.

Sometimes one also finds the term **tolerance range** which refers to the full length of the tolerance interval, i.e., $T'_i = d_i - c_i$. When reading the literature or using any kind of tolerance analysis one should always be clear on the usage of the term **tolerance**.

The function f that shows how A relates to X_1, \dots, X_n is assumed to be smooth, i.e., for small perturbations $X_i - \nu_i$ of X_i from nominal ν_i we assume that $f(X_1, \dots, X_n)$ is approximately linear in those perturbations, i.e.,

$$A = f(X_1, \dots, X_n) \approx f(\nu_1, \dots, \nu_n) + a_1(X_1 - \nu_1) + \dots + a_n(X_n - \nu_n) ,$$

where $a_i = \partial f(\nu_1, \dots, \nu_n) / \partial \nu_i$. Here one would usually treat $\nu_A = f(\nu_1, \dots, \nu_n)$ as the desired nominal assembly dimension.

Often $f(X_1, \dots, X_n)$ is naturally linear, namely

$$A = f(X_1, \dots, X_n) = a_0 + a_1 X_1 + \dots + a_n X_n$$

with known coefficients a_1, \dots, a_n . The corresponding nominal assembly dimension is then

$$\nu_A = a_0 + a_1 \nu_1 + \dots + a_n \nu_n .$$

Note that we can match this linear representation with the previous approximation by identifying

$$a_0 = f(\nu_1, \dots, \nu_n) - a_1 \nu_1 - \dots - a_n \nu_n .$$

In the simplest form the a_i coefficients are all equal to one, i.e.,

$$A = X_1 + \dots + X_n ,$$

or are all of the form $a_i = \pm 1$. This occurs naturally in tolerance path chains, where dimensions are measured off positively in one direction and negatively

in the opposite direction. In that case we would have

$$A = \pm X_1 \pm \dots \pm X_n .$$

We will assume from now on that A is of the form

$$A = a_0 + a_1 X_1 + \dots + a_n X_n$$

with known coefficients a_0, a_1, \dots, a_n . For the tolerance analysis treatment using quadratic approximations to f we refer to Cox (1986). Although this approach is more refined and appropriate for stronger curvature over the variation ranges of the X_i , it also is more complex and not routine. It also has not yet gone far in accommodating mean shifts. This part of the theory will not be covered here.

We note here that not all functions f are naturally smooth. A very simple nonsmooth function f is given by the following:

$$f(X_1, X_2) = \sqrt{X_1^2 + X_2^2}$$

which can be viewed as the distance of a hole center from the nominal origin $(0, 0)$. This function does not have derivatives at $(0, 0)$, its graph in 3-space looks like an upside cone with its tip at $(0, 0, 0)$. There can be no tangent plane at the tip of that cone and thus no linearization. Although we have found these kinds of problems appear in practice when performing tolerance analyses in the context of hole center matching and also in hinge tolerance analysis (Altschul and Scholz, 1994) there seems to be little recognition in the literature of such situations.

Let us return again to our assumption of a linear assembly criterion. The whole point of a tolerance stack analyses is to find out to what extent the assembly dimension A will differ from the nominal value ν_A while the X_i are restricted to vary over I_i . This forward analysis can then be turned around to solve the dual problem. For that problem we specify the amount of variation that can be tolerated for A and the task is that of specifying the part dimension tolerances, T_i , so that desired assembly tolerance for A will be met.

Although this is a review paper, it is far from complete, as should be clear from the above remarks. Nevertheless, it seems to be the most comprehensive review of the subject matter that we are aware of. Many open questions remain.

3 Arithmetic Tolerance Stack (Worst Case)

This type of analysis assumes that all part dimensions X_i are limited to I_i . One then analyzes what range of variation can be induced in A by varying all n part dimensions X_1, \dots, X_n independently (in the nonstatistical sense) over the respective tolerance intervals. Clearly, the largest value of

$$\begin{aligned} A &= a_0 + a_1 X_1 + \dots + a_n X_n \\ &= a_0 + a_1 \nu_1 + \dots + a_n \nu_n + a_1 (X_1 - \nu_1) + \dots + a_n (X_n - \nu_n) \\ &= \nu_A + a_1 (X_1 - \nu_1) + \dots + a_n (X_n - \nu_n) \end{aligned}$$

is realized by taking the largest (smallest) value of $X_i \in I_i = [c_i, d_i]$ whenever $a_i > 0$ ($a_i < 0$). For example, if $a_1 < 0$, then the term $a_1 (X_1 - \nu_1)$ becomes largest positive when we take $X_1 < \nu_1$ and thus at the lower endpoint c_i of I_i . Thus the maximum possible value of A is

$$A_{\max} = \max \{A : X_i \in I_i, i = 1, \dots, n\} = \nu_A + \sum_{a_i > 0} (d_i - \nu_i) a_i + \sum_{a_i < 0} (c_i - \nu_i) a_i .$$

In similar fashion one obtains the minimum value of A as

$$A_{\min} = \min \{A : X_i \in I_i, i = 1, \dots, n\} = \nu_A + \sum_{a_i < 0} (d_i - \nu_i) a_i + \sum_{a_i > 0} (c_i - \nu_i) a_i .$$

If the tolerance intervals I_i are symmetric around the nominal ν_i , i.e., $I_i = [\nu_i - T_i, \nu_i + T_i]$ or $d_i - \nu_i = T_i$ and $c_i - \nu_i = -T_i$, we find

$$A_{\max} = \nu_A + \sum_{a_i > 0} a_i T_i - \sum_{a_i < 0} a_i T_i = \nu_A + \sum_{i=1}^n |a_i| T_i$$

and

$$A_{\min} = \nu_A + \sum_{a_i < 0} a_i T_i - \sum_{a_i > 0} a_i T_i = \nu_A - \sum_{i=1}^n |a_i| T_i .$$

Thus

$$\begin{aligned} A \in [A_{\min}, A_{\max}] &= \left[\nu_A - \sum_{i=1}^n |a_i| T_i, \nu_A + \sum_{i=1}^n |a_i| T_i \right] \\ &= \left[\nu_A - T_A^{\text{arith}}, \nu_A + T_A^{\text{arith}} \right] \end{aligned}$$

where

$$T_A^{\text{arith}} = \sum_{i=1}^n |a_i| T_i \quad (1)$$

is the symmetric assembly tolerance stack. Aside from the coefficients $|a_i|$, which often are one anyway, (1) is a straight sum of the T_i 's, whence the name **arithmetic** tolerance stacking.

The calculated value T_A^{arith} should then be compared against Q_A , the requirement for successful assembly. If all part dimensions satisfy their individual tolerance requirements, i.e., $X_i \in I_i$, and if T_A^{arith} computed from (1) satisfies $T_A^{\text{arith}} \leq Q_A$, then every assembly involving these parts will fit with mathematical certainty. This is the main strength of this type of tolerance calculation.

In the special case where $|a_i| = 1$ and $T_i = T$ for $i = 1, \dots, n$ we get $T_A^{\text{arith}} = n T$, i.e., the assembly tolerance grows linearly with the number n of part dimensions. If proper assembly requires that $T_A^{\text{arith}} \leq Q_A = .004''$, then the common part tolerances have to be $T = T_A^{\text{arith}}/n \leq .004''/n$. For large n this can result in overly tight part tolerance requirements, which often are not economical. This is the main detracting feature of this form of tolerance stack analysis. It results from the overly conservative approach of stacking the worst case deviations from nominal for all parts. In reality, such worst case stackup should be extremely rare and usually occurs only when it is realized deliberately.

The assumption that all parts satisfy their respective tolerance requirements, $X_i \in I_i$, should not be neglected. Without this there is no 100% guarantee of assembly. In effect this assumption requires an inspection of all parts, typically through simple check gauges. This form of inspection is a lot simpler than that required for statistical tolerancing. For the latter the part measurements X_i themselves are required, at least for samples, in order to demonstrate process stability. Samples of part measurements are more easily amenable to extrapolation and inference about the behavior of the whole population. For samples of go/no-go data this would be a lot more difficult. There may be a cost tradeoff here, namely 100% part checking by inexpensive gauging versus part sampling (less than 100%) with expensive measuring.

Another plus for the arithmetic tolerancing scheme is that the underlying assumptions are very minimal, as can be more appreciated in the next section.

4 Statistical Tolerancing (RSS Method)

The previous section employed a form of tolerance stacking that guards against all allowed variation contingencies by stacking the part tolerances in the worst possible way. It was pointed out that this can only happen when done deliberately, i.e., choosing the worst possible parts for an assembly. If one were to choose parts in a **random** fashion, such a worst case assembly would be extremely unlikely. Typically the part deviations from nominal will tend to average out to some extent and the tolerance stack should not be as extreme as portrayed under the arithmetic tolerance stacking scheme. Statistical tolerancing exploits this type of variation cancellation in a systematic fashion.

We will introduce the method under a certain standard set of assumptions, first assuming a normal distribution describing the part variation, then relaxing this to other distributions by appealing to the central limit theorem (CLT), and we finally address the issue of assessing the risk of nonassembly.

4.1 Statistical Tolerancing With Normal Variation

The following standard assumptions are often made when first introducing the method of statistical tolerancing. These should not necessarily be accepted at face value. More realistic adaptations will be examined in subsequent sections.

Randomness: Rather than assuming that the X_i can fall anywhere in the tolerance interval I_i , even to the point that someone maliciously and deliberately selects parts for worst case assemblies, we assume here that the X_i vary randomly according to some distributions with densities $f_i(x)$, $i = 1, \dots, n$, and cumulative distribution functions

$$F_i(t) = \int_{-\infty}^t f_i(x) dx, \quad i = 1, \dots, n.$$

The idea is that most of the occurrences of X_i will fall inside I_i , i.e., most of the area under the density $f_i(x)$ falls between the endpoints of I_i . As a departure from worst case tolerancing we do accept a certain small fraction of part dimensions that will fall outside I_i . This frees us from having to inspect every part dimension for compliance with

the tolerance interval I_i . Instead we ask/assume that the processes producing the part dimensions are stable (in statistical control) and that these part dimensions fall mostly within the tolerance limits. This is ascertained by sampling only a certain portion of parts and measuring the respective X_i 's.

Independence: The independence assumption is probably the most essential corner stone of statistical tolerancing. It allows for some cancellation of variation from nominal.

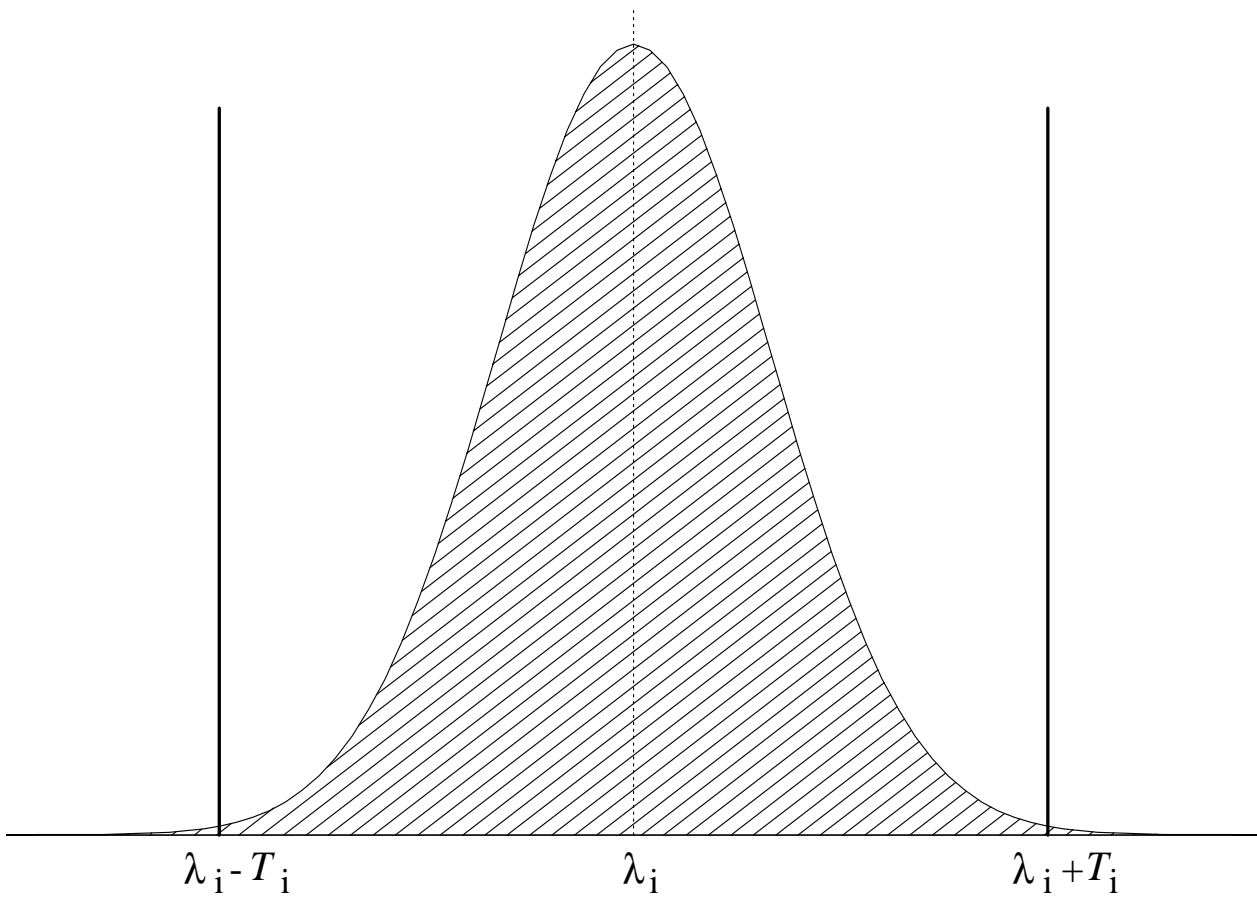
Treating the X_i as random variables, we also demand that these random variables are (statistically) independent. This roughly means that the deviation $X_i - \nu_i$ has nothing to do with the deviation $X_j - \nu_j$ for $i \neq j$. In particular, the deviations will not be predominantly positive or predominantly negative. Under independence we expect to get a mixed bag of negative and positive deviations which essentially allows for some variation cancellation. Randomness alone does not guarantee such cancellation, especially not when all part dimension show random variation in the same direction. This latter phenomenon is exactly what the independence assumption intends to exclude.

Typically the independence assumption is reasonable when part dimensions pertain to different manufacturing/machining processes. However, situations can arise where this assumption is questionable. For example, several similar/same parts (coming from the same process) could be used in the same assembly. Thermal expansion also tends to affect different parts similarly.

Distribution: It would be nice to have data on the part dimension variation, but typically that is lacking at the design stage. For that reason one often assumes that f_i is a normal or Gaussian density over the interval I_i . Since that latter is a finite interval and the Gaussian density extends over the whole real line $R = (-\infty, \infty)$, one needs to strike a compromise. It consists in asking that the area under the density f_i over the interval I_i should represent most of the total area under f_i , i.e.

$$\int_{I_i} f_i(x) dx \approx 1 .$$

Figure 1: Normal Distribution Over Tolerance Interval



In fact, most proposals ask for

$$\int_{I_i} f_i(x) dx = .9973 .$$

This rather odd looking probability (≈ 1) results from choosing f_i to be a Gaussian density with mean $\mu_i = \nu_i$ at the center of the tolerance interval and with standard deviation $\sigma_i = T_i/3$, i.e.,

$$f_i(x) = \frac{1}{\sigma_i} \varphi\left(\frac{x - \mu_i}{\sigma_i}\right) , \quad \text{where} \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

is the standard normal density. We thus have

$$I_i = [\nu_i - T_i, \nu_i + T_i] = [\mu_i - 3\sigma_i, \mu_i + 3\sigma_i]$$

and

$$\begin{aligned} \int_{I_i} f_i(x) dx &= \int_{\mu_i - 3\sigma_i}^{\mu_i + 3\sigma_i} \frac{1}{\sigma_i} \varphi\left(\frac{x - \mu_i}{\sigma_i}\right) dx \\ &= \int_{-3}^3 \varphi(x) dx = \Phi(3) - \Phi(-3) = .9973 , \end{aligned}$$

where

$$\Phi(t) = \int_{-\infty}^t \varphi(x) dx$$

is the standard normal cumulative distribution function.

Thus the odd looking probability .9973 is the result of three assumptions, namely i) a Gaussian density f_i , ii) $\nu_i = \mu_i$, i.e., the part dimension process is centered on the nominal value, and iii) $T_i = 3\sigma_i$. The first two assumptions make it possible that the simple and round factor 3 in $3\sigma_i$ produces the probability .9973. This is a widely accepted choice although others are possible. For example, Mansoor (1963) appears to prefer the factor 3.09 resulting in a probability of .999 for part dimension tolerance compliance.

One reason that is often advanced for assuming a normal distribution is that the deviations from nominal are often the result of many additive contributors which are random in nature, and each of relatively small effect. The central limit theorem (see next section) is then used to claim

normality for the sum of many such contributors. It is assumed that the person producing the part will aim for the nominal part dimension, but for various reasons there will be deviations from the nominal which accumulate to an overall deviation from nominal, which then is claimed to be normal. Thus the values X_i will typically cluster more frequently around the nominal ν_i and will less often result in values far away. This view of the distribution of X_i represents an important corner stone in the method of statistical tolerancing.

Under the above assumptions we can treat the assembly criterion

$$A = a_0 + a_1X_1 + \dots + a_nX_n$$

as a random variable, in fact as Gaussian random variable with mean

$$\mu_A = a_0 + a_1\mu_1 + \dots + a_n\mu_n = a_0 + a_1\nu_1 + \dots + a_n\nu_n = \nu_A$$

and with variance

$$\sigma_A^2 = a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2 = a_1^2 \frac{T_1^2}{3^2} + \dots + a_n^2 \frac{T_n^2}{3^2} = \frac{1}{3^2} (a_1^2T_1^2 + \dots + a_n^2T_n^2) .$$

The first equation states that the mean μ_A of A coincides with the nominal value ν_A of A . This results from the linear dependence of A on the part dimensions X_i and from the fact that the means of all part dimensions coincide with their respective nominals. The above formula for the variance can be rewritten as follows

$$3\sigma_A = \sqrt{a_1^2T_1^2 + \dots + a_n^2T_n^2} .$$

If we call $3\sigma_A = T_A^{\text{RSS}}$, we get the well known RSS-formula for statistical tolerance stacking:

$$T_A^{\text{RSS}} = \sqrt{a_1^2T_1^2 + \dots + a_n^2T_n^2} . \quad (2)$$

Here RSS refers to the root/sum/square operation that has to be performed to calculate T_A^{RSS} . Since A is Gaussian, we can count on 99.73% of all assembly criterion values A to fall within $\pm 3\sigma_A = \pm T_A^{\text{RSS}}$ of its nominal ν_A , or only .27% of all assemblies will fail.

What have we gained for the price of tolerating a small fraction of assembly failures? Again the answer becomes most transparent when all part tolerance contributions $|a_i|T_i$ are the same, i.e., $|a_i|T_i = T$. Then we have

$$T_A^{\text{RSS}} = T \sqrt{n}$$

as opposed to $T^{\text{arith}} = Tn$ in arithmetic tolerancing. The factor \sqrt{n} grows a lot slower than n . Even for $n = 2$ we find that $\sqrt{2} = 1.41$ is 29% smaller than 2 and for $n = 4$ we have that $\sqrt{4}$ is 50% smaller than 4.

If proper assembly requires that $T_A^{\text{RSS}} \leq Q_A = .004''$, then the common part tolerance contributions have to be $T = T_A^{\text{RSS}}/\sqrt{n} \leq .004''/\sqrt{n}$. Due to the divisor \sqrt{n} , these part tolerances are much more liberal than those obtained under arithmetic tolerancing.

4.2 Statistical Tolerancing Using the CLT

One assumption used heavily in the previous section is that of a Gaussian distribution for all part dimensions X_i . This assumption has often been challenged, partly based on part data that contradict the normality, partly based on mean shifts that result in an overall mixture of normal distributions, i.e., more smeared out, and last but not least based on the experience that the assembly fallout rate was higher than predicted by statistical tolerancing. We will here relax the normality assumption by allowing more general distributions for the part variations X_i . However, we will insist that the mean μ_i of X_i still coincides with the nominal ν_i . Relaxing this last constraint will be discussed in subsequent sections.

To relax the normality assumption for the part dimensions X_i we appeal to the central limit theorem of probability theory (CLT). In fact, we will now use the following assumptions

1. The X_i , $i = 1, \dots, n$, are statistically independent.
2. The density f_i governing the distribution of X_i has mean $\mu_i = \nu_i$ and standard deviation σ_i .
3. The variability contributions of all terms in the linear combination A become negligible for large n , i.e.,

$$\frac{\max(a_1^2\sigma_1^2, \dots, a_n^2\sigma_n^2)}{a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2} \longrightarrow 0 \quad \text{as } n \rightarrow \infty .$$

Under these three conditions¹ the Lindeberg-Feller CLT states that the linear

¹In fact, they need to be slightly stronger by invoking the more technical Lindeberg condition, see Feller (1966).

combination

$$A = a_0 + a_1X_1 + \dots + a_nX_n$$

has an approximately normal distribution with mean

$$\mu_A = a_0 + a_1\mu_1 + \dots + a_n\mu_n = a_0 + a_1\nu_1 + \dots + a_n\nu_n = \nu_A$$

and with variance

$$\sigma_A^2 = a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2 .$$

Assumption 3 eliminates situations where a small number of terms in the linear combination have so much variation that they completely swamp the variation of the remaining terms. If these few dominating terms have non-normal distributions, it can hardly be expected that the linear combination has an approximately normal distribution.

In spite of the relaxed distributional assumptions for the part dimensions we have that the assembly criterion A is again approximately normally distributed and its mean μ_A coincides with the desired nominal value ν_A (because we deal with a linear combination and since we assumed $\mu_i = \nu_i$). From the approximate normality of A we can count on about 99.73% of all assembly criteria to fall within $[\nu_A - 3\sigma_A, \nu_A + 3\sigma_A]$.

This is almost the same result as before, except for one “minor” point. In the previous section we had assumed a particular relation between the part dimension σ_i and the tolerance T_i , namely we stipulated that $T_i = 3\sigma_i$. This was motivated mainly by the fact that under the normality assumption almost all (99.73%) part dimensions would fall within $\pm 3\sigma_i$ of the nominal $\nu_i = \mu_i$. Without the normality assumption for the parts there is no such high probability assurance for such $\pm 3\sigma_i$ ranges. However, the Camp-Meidell inequality (Encyclopedia of Statistical Sciences, Vol. I, 1982), states that for symmetric and unimodal densities f_i with finite variance σ_i^2 we have

$$P(|X_i - \mu_i| \leq 3\sigma_i) \geq 1 - \frac{4}{81} = .9506.$$

Here symmetry means that $f_i(\nu_i + y) = f_i(\nu_i - y)$ for all y , and thus that $\mu_i = \nu_i$. Unimodality means that $f_i(\nu_i + y) \geq f_i(\nu_i + y')$ for all $0 \leq |y| \leq |y'|$, i.e., the density falls off as we move away from its center, or at least it does not increase. Although this covers a wide family of reasonable distributions, the number .9506 does not carry with it the same degree of certainty as .9973.

We thus do not yet have a natural link between the standard deviation σ_i and the part dimension tolerance T_i . If the distribution of X_i has a finite range, then one could equate that finite range with the $\pm T_i$ tolerance range around ν_i . This is what has commonly been done. In the case of a Gaussian f_i this was not possible (because of the infinite range) and that was resolved by opting for the $\pm 3\sigma_i = \pm T_i$ range. By matching the finite range of a distribution with the tolerance range $[\nu_i - T_i, \nu_i + T_i]$ we obtain the link between σ_i and T_i , and thus ultimately the link between T_A and T_i . Since the spread $2T_i$ of a such finite range distribution can be manipulated by a simple scale change which also affects the standard deviation of the distribution by the same factor it follows that σ_i and T_i will be proportional to each other, i.e., we can stipulate that

$$cT_i = 3\sigma_i ,$$

where c is a factor that is specific to the type of distribution. The choice of linking this proportionality back to $3\sigma_i$ facilitates the comparison with the normal distribution, for which we would have $c = c_N = 1$.

Assuming that the type of distribution (but not necessarily its location and scale) is the same for all part dimensions we get

$$\begin{aligned} T_A^{\text{RSS},c} = 3\sigma_A &= \sqrt{(3a_1\sigma_1)^2 + \dots + (3a_n\sigma_n)^2} \\ &= \sqrt{(ca_1T_1)^2 + \dots + (ca_nT_n)^2} \\ &= c \sqrt{a_1^2T_1^2 + \dots + a_n^2T_n^2} = c T_A^{\text{RSS}} . \end{aligned}$$

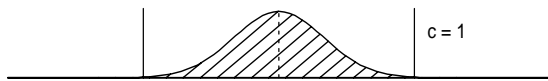
This leads to tolerance stacking formulas that essentially agree with (2), except that an inflation factor, c , has been added. If the distribution type also changes from part to part (hopefully with good justification), i.e., we have different factors c_1, \dots, c_n , we need to use the following more complicated tolerance stacking formula:

$$T_A^{\text{RSS},\mathbf{c}} = \sqrt{(c_1a_1T_1)^2 + \dots + (c_na_nT_n)^2} , \quad (3)$$

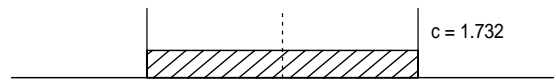
where $\mathbf{c} = (c_1, \dots, c_n)$.

In Table 1 we give a few factors that have been considered in the literature, see Gilson (1951), Mansoor (1963), Fortini (1967), Kirschling (1988), Bjørke (1989), and Henzold (1995). The corresponding distributions are illustrated

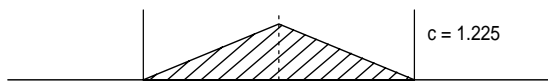
Figure 2: Tolerance Interval Distributions & Factors



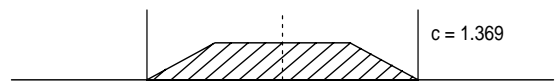
normal density



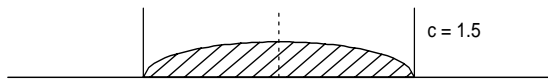
uniform density



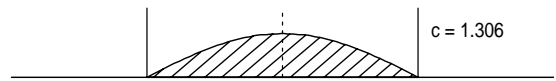
triangular density



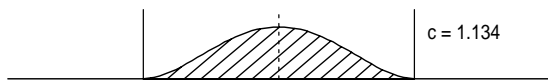
trapezoidal density: $a = .5$



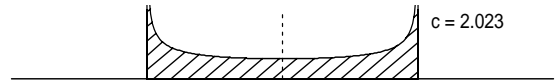
elliptical density



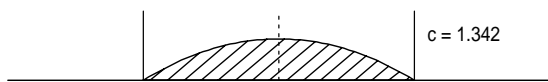
half cosine wave density



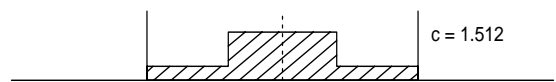
beta density: $a = 3, b = 3$



beta density: $a = .6, b = .6$



beta density: $a = 2, b = 2$ (parabolic)



DIN - histogram density: $p = .7, f = .4$

Table 1: Distributional Inflation Factors

normal	1	uniform	1.732
triangular	1.225	trapezoidal	$\sqrt{3(1+a^2)}/2$
cosine half wave	1.306	elliptical	1.5
beta (symmetric)	$3/\sqrt{2a+1}$		
histogram density (DIN)	$\sqrt{3}\sqrt{(1-p)(1+f)+f^2}$		

in Figure 2. For a derivation of these factors see Appendix A. For the beta density the parameters $a > 0$ and $b > 0$ are the usual shape parameters, a in the trapezoidal density indicates the break point from the flat to the sloped part of the density, and p and f characterize the histogram density (see the last density in Figure 2), namely the middle bar of that density covers the middle portion $\nu_i \pm fT_i$ of the tolerance interval and its area comprises 100

$p\%$ of the total density.

Some factors have little explicit justification and motivation and are presented without proper reference. For example, the factor $c = 1.6$ of Gilson (1951) derives from his crucial empirical formula (2) which is prefaced by “Without going deeply into a mathematical analysis . . .” Evans (1975) seems to welcome such lack of mathematical detail by saying: “None of the results are derived, in the specialized sense of this word, so that it is readable by virtually anyone who would be interested in the tolerancing problem.”

Bender (1962) gives the factor 1.5 based mainly on the fact that production operators will usually give you 2/3 of the true spread ($\pm 3\sigma$ range under a normal distribution) when asked what tolerance limits they can hold and “quality control people recognize that this 2/3 total spread includes about 95% of the pieces.” To make up for these optimistically stated tolerances, Bender suggests the factor $3/2 = 1.5$.

4.3 Risk Assessment with Statistical Tolerance Stacking

In this section we discuss the assembly risk, i.e., the chance that an assembly criterion A will not satisfy its requirement. As in the previous section it is assumed that all part dimensions X_i have symmetric distributions centered

on their nominals, i.e., with means $\mu_i = \nu_i$, and variances σ_i^2 , respectively. The requirement for successful assembly is assumed to be $|A - \nu_A| \leq K_0$, where K_0 is some predetermined number based on design considerations. We are then interested in assessing $P(|A - \nu_A| > K_0)$. According to the CLT we can treat $(A - \nu_A)/\sigma_A = (A - \mu_A)/\sigma_A$ as an approximately standard normal random variable. Thus the assembly risk is

$$\begin{aligned}
P(|A - \nu_A| > K_0) &= P\left(\frac{|A - \nu_A|}{\sigma_A} > \frac{K_0}{\sigma_A}\right) \\
&= P\left(\frac{A - \nu_A}{\sigma_A} < -\frac{K_0}{\sigma_A}\right) + P\left(\frac{A - \nu_A}{\sigma_A} > \frac{K_0}{\sigma_A}\right) \\
&= \Phi\left(-\frac{K_0}{\sigma_A}\right) + 1 - \Phi\left(\frac{K_0}{\sigma_A}\right) = 2\Phi\left(-\frac{K_0}{\sigma_A}\right) \\
&= 2\Phi\left(-\frac{K_0}{T_A^{\text{RSS},c}/3}\right) = 2\Phi\left(-\frac{3K_0}{cT_A^{\text{RSS}}}\right). \quad (4)
\end{aligned}$$

When the requirement K_0 is equal to $cT_A^{\text{RSS}} = c\sqrt{a_1^2T_1^2 + \dots + a_n^2T_n^2}$, then the nonassembly risk is

$$P(|A - \nu_A| > K_0) = 2\Phi(-3) = .0027,$$

the complement of our familiar .9973. The factor c affects to what extent we will be able to fit cT_A^{RSS} into K_0 . If $cT_A^{\text{RSS}} > K_0$ we have to reduce either c or T_A^{RSS} . Since c depends on the distribution that gives rise to it and which portrays our vague knowledge of manufacturing variation, we are left with reducing T_A^{RSS} , i.e., the individual part tolerances. If we do neither we have to accept a higher nonassembly risk which can be computed via formula (4).

When we invoke the CLT we often treat the resulting approximations as though they are exact. However, it should be kept in mind that in reality we deal with approximations (although typically good ones) and that the accuracy becomes rather limited when we make such calculations involving the extreme tails of the normal distributions. For example, a normal approximation may suggest a probability of .9973, but in reality that probability may be only .98. When making corrections for such extreme tail probabilities, it would seem that one often splits hairs given that these probabilities are only approximate anyway. However, whatever the correct probability might be,

if the approximation suggests a degradation in the tolerance assurance level and if we make an adjustment based on the same approximation, it would seem that we have had some effect. The only problem is that the counteractive measure may not be enough (case a)) or may be more than needed (case b)). If in either of these situations we had done nothing, then we would be much worse off in case a) or are counting on wishful thinking in case b).

5 Mean Shifts

So far we have made the overly simplistic assumption that all part dimensions X_i be centered on their respective nominals ν_i . In practice this is difficult to achieve and often not economical. Such mean shifts may at times be quite deliberate (aiming for maximal material condition, because one prefers rework to scrap), at other times it is caused by tool wear, and often one cannot average out the part manufacturing process exactly at the nominal center ν_i , as hard as one may try. A shift of the distribution of the X_i away from the respective nominal centers will cause a shift also in the assembly criterion A . This in turn will increase the nonassembly risk, since it will shift the normal curve more towards one end of the assembly design requirement $[-K_0, K_0]$.

Some authors, e.g., Bender (1962) and Gilson (1951), have responded to this problem by introducing inflation factors, c , as they were discussed in the previous section, but maintaining a distribution for X_i which is still symmetric around ν_i . In effect, this trades one ill effect, namely the mean shift, against another by assuming a higher variance, but still constraining X_i to the tolerance interval $I_i = [\nu_i - T_i, \nu_i + T_i]$. The remedy (inflation factor c) that accounts for higher variance within I_i will, as a side effect, also be beneficial for dealing with mean shifts, since it causes a tightening of part tolerances and thus a more conservative design. Such a design will then naturally also compensate for some amounts of mean shift. Greenwood and Chase (1987) refer to this treatment of the mean shift problem as using a Band-Aid, since this practice is not specific to the mean shift problem.

A mean shift represents a persistent amount of shift and is thus quite deterministic in its effect, whereas an inflated variance expresses variation that changes from part to part, and thus allows error cancellation. In defense of this latter approach one should mention that sometimes one reacts to off-center means by “recentering” the manufacturing process. Since that will

presumably produce another off-center mean, this iterative “recentering” will just add to the overall variability of the process, i.e., mean shifts are then indeed physically transformed into variability. Whether this “recentering” is a good strategy, is questionable. A shift will typically produce rejected parts only on one side of the tolerance interval, whereas the increased variability due to “recentering” will result in rejects on both sides of the tolerance interval.

We will now discuss some ways of explicitly dealing with mean shifts $\Delta_i = \mu_i - \nu_i$. Although we allow for the possibility of mean shifts we will still maintain the idea of a tolerance interval, i.e., the i^{th} part dimension X_i will still be constrained to the tolerance interval I_i . If the distribution of X_i is assumed to be normal, then its $\pm 3\sigma_i$ range should still fall within I_i , see Figure 3. This means that σ_i has to get smaller as $|\Delta_i|$ gets larger. For fixed tolerance intervals this means that larger mean shifts are only possible with tighter variation. In the extreme this means that the distribution of X_i is shifted all the way to $\nu_i - T_i$ or $\nu_i + T_i$, with no variability at all. This latter scenario is hardly realistic², but it is worth noting since it leads back to worst case tolerancing.

In practice it is not so easy to tighten the variation of a part production process. It is more practical to widen the part dimension tolerance interval I_i or to increase T_i . The tolerance stack up analysis is then performed in terms of these increased T_i . The effect, from an analysis method point of view, is the same. With increased T_i the unchanged σ_i will look reduced relative to T_i . It is only a matter of who pays the price.

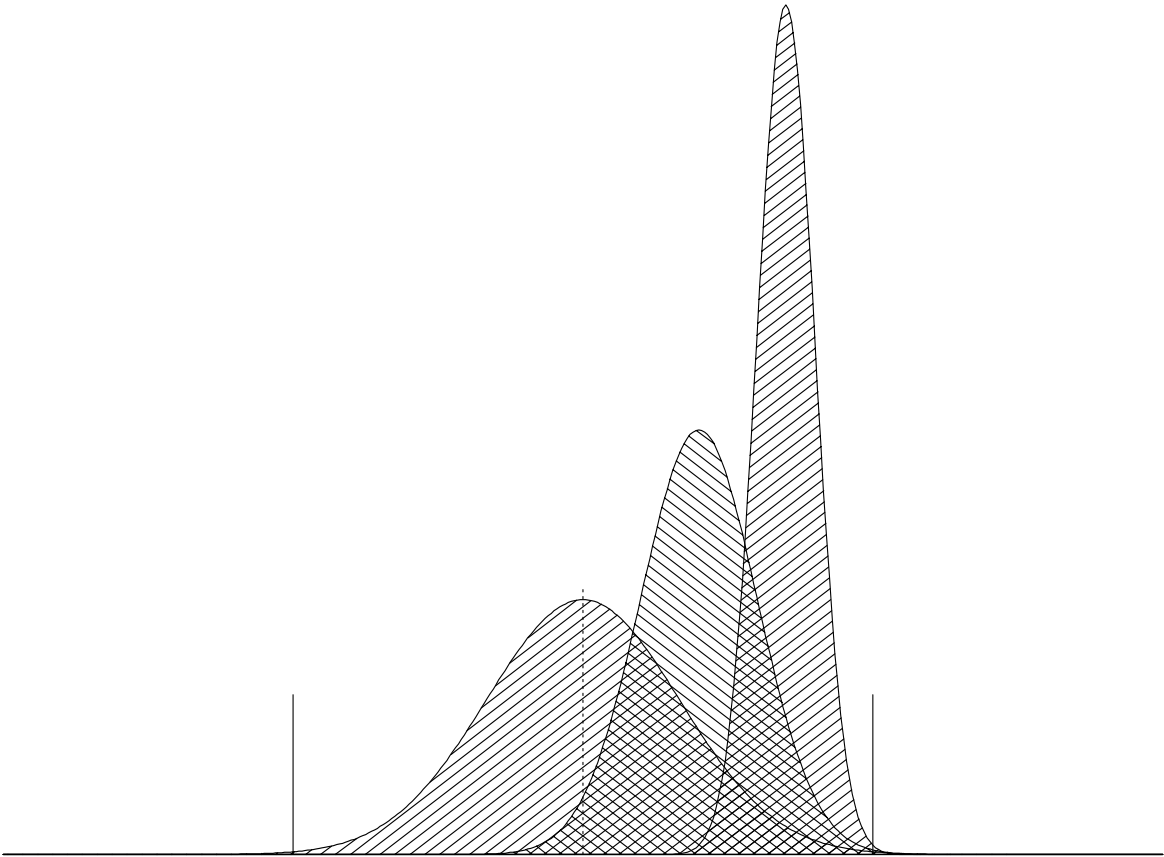
Typically the mean shifts are not known a priori and, as pointed out above, in the extreme case they are unrealistic and lead us right back to worst case tolerancing. To avoid this, the amount of mean shift one is willing to tolerate needs to be limited. Such bounds on the mean shift should be arrived at in consultation with the appropriate manufacturing representatives. For the following discussion it is useful to represent $|\Delta_i|$ as a fraction η_i of T_i , i.e., $|\Delta_i| = \eta_i T_i$, with $0 \leq \eta_i \leq 1$. The bounds on $|\Delta_i|$ can now equivalently be expressed as bounds on the η_i , namely $\eta_i \leq \eta_{0i}$ or

$$|\Delta_i| \leq \eta_{0i} T_i .$$

It is usually more reasonable to state the bounds on $|\Delta_i|$ in proportionality to T_i . One reason for this is that T_i captures to some extent the variability

²It usually is much harder to reduce the variability of X_i than to control its mean.

Figure 3: Shifted Normal Distributions Over Tolerance Interval



of the part dimension process and one is inclined to assume that the same force that is behind this variability is to some extent also responsible for the variation of the mean μ_i , i.e., that there is some proportionality between the two phenomena of variation. Also, once such mean shift bounds are expressed in terms of such proportionality to T_i , one is then more willing to assume a common bound for these proportionality factors, namely $\eta_{01} = \dots = \eta_{0n} = \eta_0$. Having a common bound η_0 for all part dimensions X_i is not necessary, but greatly simplifies the exposition and the practice of adjusting for mean shifts.

We can now view the part dimension X_i as the sum of two (or three) contributions:

$$X_i = \mu_i + \epsilon_i = \nu_i + (\mu_i - \nu_i) + \epsilon_i = \nu_i + \Delta_i + \epsilon_i$$

where μ_i is the mean around which the individual i^{th} part dimensions cluster and ϵ_i is the amount by which X_i deviates from μ_i each time that part gets produced. The variation term ϵ_i is assumed to vary according to some distribution with mean zero and variance σ_i^2 . We can think of the two terms in $\Delta_i + \epsilon_i$ as the total deviation of X_i from the nominal ν_i . Namely, μ_i differs from ν_i by the mean shift Δ_i in a persistent way and then each part dimension will have its own deviation ϵ_i from μ_i . However, this latter deviation will be different from one realization of part dimension X_i to the next. Hence the resulting assemblies will experience different deviations from that part dimension, each time a new assembly is made. However, the contribution Δ_i will be the same from assembly to assembly.

The above representation then leads to a corresponding representation for the assembly criterion:

$$\begin{aligned} A &= a_0 + a_1 X_1 + \dots + a_n X_n \\ &= a_0 + a_1(\mu_1 + \epsilon_1) + \dots + a_n(\mu_n + \epsilon_n) \\ &= a_0 + (a_1 \mu_1 + \dots + a_n \mu_n) + (a_1 \epsilon_1 + \dots + a_n \epsilon_n) \\ &= \mu_A + \epsilon_A = \nu_A + (\mu_A - \nu_A) + \epsilon_A = \nu_A + \Delta_A + \epsilon_A, \end{aligned}$$

where

$$\mu_A = a_0 + a_1 \mu_1 + \dots + a_n \mu_n, \quad \nu_A = a_0 + a_1 \nu_1 + \dots + a_n \nu_n,$$

$$\Delta_A = \mu_A - \nu_A, \quad \text{and} \quad \epsilon_A = a_1\epsilon_1 + \dots + a_n\epsilon_n.$$

Here μ_A is the mean of A , ν_A is the assembly nominal, Δ_A is the assembly mean shift, and ϵ_A captures the variation of A from assembly to assembly, having mean zero and variance

$$\sigma_{\epsilon_A}^2 = a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2.$$

5.1 Arithmetic Stacking of Mean Shifts

The variation of A around the assembly nominal ν_A is the composite of two contributions, namely the assembly mean shift $\Delta_A = \mu_A - \nu_A$ and the assembly variation ϵ_A , which is the sum of n random contributions and thus amenable to statistical tolerance stacking.

The amount by which μ_A may differ from ν_A can be bounded as follows:

$$\begin{aligned} |\mu_A - \nu_A| &= |a_1(\mu_1 - \nu_1) + \dots + a_n(\mu_n - \nu_n)| \\ &= |a_1\Delta_1 + \dots + a_n\Delta_n| \\ &\leq |a_1||\Delta_1| + \dots + |a_n||\Delta_n| \\ &= \eta_1|a_1|T_1 + \dots + \eta_n|a_n|T_n, \end{aligned} \tag{5}$$

where the latter sum reminds of worst case or arithmetic tolerance stacking. In fact, that is exactly what is happening here with the mean shifts, in that we assume all mean shifts go in the most unfavorable direction. The inequality in (5) can indeed be an equality provided all the $a_i\Delta_i$ have the same sign.

The CLT may again be invoked to treat ϵ_A as approximately normal with mean zero and variance $\sigma_{\epsilon_A}^2$, so that we can expect 99.73% of all assembly variations ϵ_A to fall within $\pm 3\sigma_{\epsilon_A}$ of zero. Thus 99.73% of all assembly dimensions A fall within

$$\mu_A \pm 3\sigma_{\epsilon_A}.$$

Since (5) bounds the amount by which μ_A may differ from ν_A we can couple this additively (in worst case fashion) with the previous 99.73% interval bound for A and can claim that at least 99.73% of all assembly dimensions A will fall within

$$\nu_A \pm \left(\sum_{i=1}^n \eta_i |a_i| T_i + 3\sigma_{\epsilon_A} \right). \tag{6}$$

Because of the worst case addition of mean shifts one usually will wind up with less than .27% of assembly criteria A falling outside the interval (6). That percentage is correct when the assembly mean shift is zero. As the assembly mean μ_A shifts to the right or to the left of ν_A , only one of the normal distribution tails will significantly contribute to the assembly out of tolerance rate. That rate is more likely to be just half of .27% or .135%, or slightly above. The shifted and scaled normal densities in Figure 3 illustrate that point as well.

So far we have not factored in our earlier assumption that σ_i should decrease as $|\Delta_i|$ increases, so that the part dimension tolerance requirement $X_i \in I_i$ is maintained. If we assume a normal distribution for X_i , this means that we require that the $\pm 3\sigma_i$ ranges around μ_i still be contained within I_i . At the same time this means that the fallout rate will shrink from .27% (for zero mean shift) to .135% as the mean shift gets larger, since only one distribution tail will contribute. Since with zero mean shift one allows .27% fallout, one could have allowed an increase in σ_i so that the single tail fallout would again be .27%. We will not enter into this complication and instead stay with our original interpretation, namely require that the capability index C_{pk} satisfy

$$C_{pk} = \frac{T_i - |\Delta_i|}{3\sigma_i} = \frac{T_i - \eta_i T_i}{3\sigma_i} = \frac{(1 - \eta_i)T_i}{3\sigma_i} \geq 1 .$$

Assuming the highest amount of variability within these constraints, i.e., $C_{pk} = 1$, we have

$$3\sigma_i = (1 - \eta_i)T_i. \quad (7)$$

In view of our initial identification of $3\sigma_i = T_i$ (without mean shift) this equation can be interpreted two ways. Either σ_i needs to be reduced by the factor $(1 - \eta_i)$ or T_i needs to be increased by the factor $1/(1 - \eta_i)$ in order to accommodate a $\pm \eta_i T_i$ mean shift. Whichever way equation (7) is realized, we then have

$$3\sigma_{\epsilon_A} = \sqrt{a_1^2(3\sigma_1)^2 + \dots + a_n^2(3\sigma_n)^2} = \sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2}.$$

With this representation of $3\sigma_{\epsilon_A}$ at least 99.73% of all assembly dimensions A will fall within

$$\nu_A \pm \left(\sum_{i=1}^n \eta_i |a_i| T_i + \sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2} \right). \quad (8)$$

As pointed out above, this compliance proportion is usually higher, i.e., more like 99.865%, as will become clear in the next section on risk assessment.

The above combination of worst case stacking of mean shifts and RSS-stacking of the remaining variability within each tolerance interval was proposed by Mansoor (1963) and further enlarged on by Greenwood and Chase (1987).

In formula (8) the i^{th} shift fraction η_i appears in two places, first in the sum (increasing in η_i) and then under the root (decreasing in η_i). It is thus not obvious that increasing η_i will always make matters worse as far as interval width is concerned. Since

$$\begin{aligned} \frac{\partial}{\partial \eta_j} \left(\sum_{i=1}^n \eta_i |a_i| T_i + \sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2} \right) \\ = |a_j| T_j - \frac{(1 - \eta_j) a_j^2 T_j^2}{\sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2}} \\ \geq |a_j| T_j - \frac{(1 - \eta_j) a_j^2 T_j^2}{\sqrt{(1 - \eta_j)^2 a_j^2 T_j^2}} = 0 \end{aligned}$$

it follows that increasing η_j will widen the interval (8). If all the shift fractions η_j are bounded by the common η_0 , we can thus limit the variation of the assembly criterion A to

$$\nu_A \pm \left(\eta_0 \sum_{i=1}^n |a_i| T_i + (1 - \eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2} \right) = \nu_A \pm T_A^{\Delta, \text{arith}} \quad (9)$$

with at least 99.73% (or better yet with 99.865%) assurance of containing A . The half width of this interval

$$T_A^{\Delta, \text{arith}} = \eta_0 \sum_{i=1}^n |a_i| T_i + (1 - \eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2}$$

is a weighted combination (with weights η_0 and $1 - \eta_0$) of arithmetic and statistical tolerance stacking of the part tolerances T_i . As such it can be viewed as a unified approach, as suggested by Greenwood and Chase (1987), since $\eta_0 = 0$ results in pure statistical tolerancing and $\eta_0 = 1$ results in pure arithmetic tolerancing.

Comparing the two components of this weighted combination it is easily seen (by squaring both sides and noting that all terms $|a_i|T_i$ are nonnegative) that

$$\sum_{i=1}^n |a_i|T_i \geq \sqrt{\sum_{i=1}^n |a_i|^2 T_i^2},$$

where the left side is usually significantly larger than the right. This inequality, which contrasts the difference between arithmetic stacking and RSS stacking, is a simple illustration of the Pythagorean theorem. Think of a rectangular box in n -space, with sides $|a_i|T_i$, $i = 1, \dots, n$. In order to go from one corner of this box to the diametrically opposed corner we can proceed either by going along the edges, traversing a distance $\sum_{i=1}^n |a_i|T_i$, or we can go directly on the diagonal connecting the diametrically opposed corners. In the latter case we traverse the much shorter distance of $\sqrt{\sum_{i=1}^n |a_i|^2 T_i^2}$ according to Pythagoras. The Pythagorean connection was also alluded to by Harry and Stewart (1988), although in somewhat different form, namely in the context of explaining the variance of a sum of independent random variables.

As long as $\eta_0 > 0$, i.e., some mean shift is allowed, we find that this type of stacking the tolerances T_i is of order n . This is seen most clearly when $|a_i|T_i = T$ and $|a_i| = 1$ for all i . Then

$$T_A^{\Delta, \text{arith}} = n\eta_0 T + \sqrt{n}(1 - \eta_0)T = nT \left(\eta_0 + \frac{1 - \eta_0}{\sqrt{n}} \right),$$

which is of order n , although reduced by the factor η_0 . Thus the previously noted possible gain in the compliance rate, namely 99.73% \nearrow 99.865%, is typically more than offset by the order n growth in the tolerance stack when mean shifts are present.

This increased assembly compliance rate could be converted back to 99.73% by placing the factor $2.782/3 = .927$ in front of the square root in formula (9). The value 2.782 represents the 99.73% point of the standard normal distribution. If, due to the allowed mean shift, we only have to worry about one tail of the normal distribution exceeding the tolerance stack limits, then we can reduce our customary factor 3 in $3\sigma_{\epsilon_A}$ to 2.782. To a small extent this should offset the mean shift penalty. The resulting tolerance stack

interval is then

$$\nu_A \pm \left(\eta_0 \sum_{i=1}^n |a_i| T_i + .927 (1 - \eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2} \right) = \nu_A \pm \tilde{T}_A^{\Delta, \text{arith}} . \quad (10)$$

So far we have assumed that the variation of the ϵ_i terms is normal, with mean zero and variance σ_i^2 . This normality assumption can be relaxed as before by assuming a symmetric distribution over a finite interval, J_i , centered at zero. This finite interval, after centering it on μ_i , should still fit inside the tolerance interval I_i . Thus J_i will be smaller than I_i . This reduction in variability is the counterpart of reducing σ_i in the normal model, as $|\Delta_i|$ increases. See Figure 4 for the shifted distribution versions of Figure 2 with the accompanying reduction in variability. Alternatively, we could instead widen the tolerance intervals I_i while keeping the spread of the distributions fixed.

If I_i has half width T_i and if the absolute mean shift is $|\Delta_i|$, then the reduced interval J_i will have half width

$$T'_i = T_i - |\Delta_i| = T_i - \eta_i T_i = (1 - \eta_i) T_i .$$

The density f_i , describing the distribution of ϵ_i over the interval J_i , has variance σ_i^2 and as before we have the following relationship:

$$3\sigma_i = cT'_i = c(1 - \eta_i)T_i ,$$

where c is a factor that depends on the distribution type, see Table 1. Using (6) and

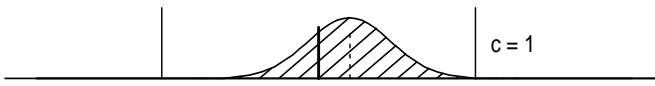
$$\begin{aligned} 3\sigma_{\epsilon_A} &= \sqrt{a_1^2 (3\sigma_1)^2 + \dots + a_n^2 (3\sigma_n)^2} \\ &= c \sqrt{a_1^2 (1 - \eta_1)^2 T_1^2 + \dots + a_n^2 (1 - \eta_n)^2 T_n^2} \end{aligned}$$

formula (8) simply changes to

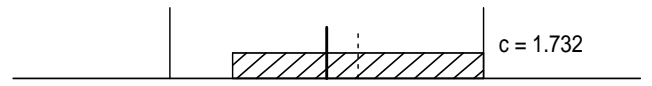
$$\nu_A \pm \left(\sum_{i=1}^n \eta_i |a_i| T_i + c \sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2} \right) , \quad (11)$$

i.e., there is an additional penalty through the inflation factor c . If the part dimension tolerance intervals involve different distributions, then one can accommodate this in a similar fashion as in (3).

Figure 4: Shifted Tolerance Interval Distributions & Factors



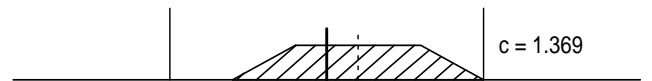
shifted normal density



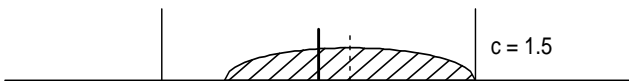
shifted uniform density



shifted triangular density



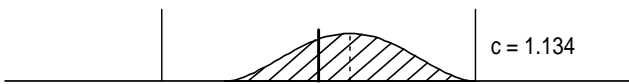
shifted trapezoidal density: $a = .5$



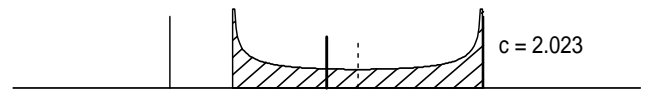
shifted elliptical density



shifted half cosine wave density



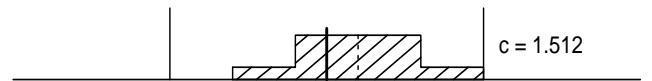
shifted beta density: $a = 3, b = 3$



shifted beta density: $a = .6, b = .6$



shifted beta density: $a = 2, b = 2$ (parabolic)



DIN - histogram density: $p = .7, f = .4$

Here it is not as clear whether increasing η_i will widen interval (11) or not. Taking derivatives as before

$$\begin{aligned} \frac{\partial}{\partial \eta_j} \left(\sum_{i=1}^n \eta_i |a_i| T_i + c \sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2} \right) \\ = |a_j| T_j - c \frac{(1 - \eta_j) a_j^2 T_j^2}{\sqrt{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2}} \geq 0 \end{aligned}$$

if and only if

$$\frac{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2}{(1 - \eta_j)^2 a_j^2 T_j^2} \geq c^2 .$$

This will usually be the case as long as c is not too much larger than one and as long as $(1 - \eta_j)^2 a_j^2 T_j^2$ is not the overwhelming contribution to

$$(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2 .$$

If $\eta_i \leq \eta_0$ then

$$\frac{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2}{(1 - \eta_j)^2 a_j^2 T_j^2} \geq 1 + (1 - \eta_0)^2 \frac{\sum_{i \neq j} a_i^2 T_i^2}{a_j^2 T_j^2} .$$

Here the right side is well above one, unless $a_j^2 T_j^2$ is very much larger than the combined effect of all the other $a_i^2 T_i^2$, $i \neq j$. This situation usually does not arise.

As an example where c is too large consider $n = 2$, $(1 - \eta_1)|a_1|T_1 = (1 - \eta_2)|a_2|T_2$, and $f_1 = f_2 = \text{uniform}$. Then $c = \sqrt{3} = 1.732$ from Table 1 and

$$\frac{(1 - \eta_1)^2 a_1^2 T_1^2 + \dots + (1 - \eta_n)^2 a_n^2 T_n^2}{(1 - \eta_j)^2 a_j^2 T_j^2} = 2 < c^2 = 3 .$$

In that case the above derivative is negative, which means that the interval (11) is widest when there is no mean shift at all. This strange behavior does not carry over to $n = 3$ uniform distributions. Also, it should be pointed out that for $n = 2$ and uniform part dimension variation the CLT does not yet provide a good approximation to the distribution of A , which in that case is triangular.

In most cases we will find that the above derivatives are nonnegative and that the maximum interval width subject to $\eta_i \leq \eta_0$ is indeed achieved at

$\eta_i = \eta_0$ for $i = 1, \dots, n$. We would then have that A conservatively falls within

$$\nu_A \pm \left(\eta_0 \sum_{i=1}^n |a_i| T_i + c(1 - \eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2} \right) = \nu_A \pm T_A^{c, \Delta, \text{arith}} \quad (12)$$

with at least 99.73% (or 99.865%) assurance. This latter percentage derives again from the CLT applied to ϵ_A . Taking advantage of the 99.865% we could again introduce the reduction factor .927 in (12) and use

$$\nu_A \pm \left(\eta_0 \sum_{i=1}^n |a_i| T_i + .927c(1 - \eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2} \right) = \nu_A \pm \tilde{T}_A^{c, \Delta, \text{arith}} \quad (13)$$

with at least 99.73% assurance.

5.2 Risk Assessment with Arithmetically Stacked Mean Shifts

This section parallels Section 6 on the same subject, except that here we account for mean shifts. These cause the normal distribution of A to move away from the center of the assembly requirement interval given by $|A - \nu_A| \leq K_0$. The probability of satisfying this assembly requirement is now

$$\begin{aligned} P(|A - \nu_A| \leq K_0) &= P(|A - \mu_A + \mu_A - \nu_A| \leq K_0) \\ &= P\left(\left|\frac{A - \mu_A}{\sigma_{\epsilon_A}} + \frac{\Delta_A}{\sigma_{\epsilon_A}}\right| \leq \frac{K_0}{\sigma_{\epsilon_A}}\right) \\ &= P\left(\left|Z + \frac{\Delta_A}{\sigma_{\epsilon_A}}\right| \leq \frac{K_0}{\sigma_{\epsilon_A}}\right) \\ &\geq P\left(\left|Z - \frac{\sum_{i=1}^n \eta_i |a_i| T_i}{\sigma_{\epsilon_A}}\right| \leq \frac{K_0}{\sigma_{\epsilon_A}}\right) \end{aligned}$$

where Z is a standard normal random variable and in the inequality we replaced Δ_A by one of the two worst case assembly mean shifts for a fixed set of $\{\eta_1, \dots, \eta_n\}$, namely by $-\sum_{i=1}^n \eta_i |a_i| T_i$, see (5). Replacing σ_{ϵ_A} by the corresponding remaining assembly variability, i.e.,

$$\sigma_{\epsilon_A} = \frac{c}{3} \sqrt{\sum_{i=1}^n (1 - \eta_i)^2 a_i^2 T_i^2}, \quad (14)$$

we get

$$P(|A - \nu_A| \leq K_0) \geq P(|Z - C(\eta_1, \dots, \eta_n)| \leq W(\eta_1, \dots, \eta_n)) , \quad (15)$$

where

$$C(\eta_1, \dots, \eta_n) = \frac{\sum_{i=1}^n \eta_i |a_i| T_i}{(c/3) \sqrt{\sum_{i=1}^n (1 - \eta_i)^2 a_i^2 T_i^2}} \geq 0$$

and

$$W(\eta_1, \dots, \eta_n) = \frac{K_0}{(c/3) \sqrt{\sum_{i=1}^n (1 - \eta_i)^2 a_i^2 T_i^2}} .$$

Assuming that we impose the restriction $\eta_i \leq \eta_0$ for $i = 1, \dots, n$, it seems plausible that the worst case lower bound for $P(|A - \nu_A| \leq K_0)$ in (15) is attained when $\eta_i = \eta_0$ for $i = 1, \dots, n$. Clearly $C(\eta_1, \dots, \eta_n)$ is increasing in each η_i , but so is $W(\eta_1, \dots, \eta_n)$. Thus the interval

$$J = [C(\eta_1, \dots, \eta_n) - W(\eta_1, \dots, \eta_n), C(\eta_1, \dots, \eta_n) + W(\eta_1, \dots, \eta_n)]$$

not only shifts further to the right of zero as we increase η_i , but it also gets wider. Thus it is not clear whether the probability lower bound $P(Z \in J)$ decreases as we increase η_i .

Using common part tolerances $|a_i|T_i = T$ for all i , we found no counterexample to the above conjecture during limited simulations for $c = 1$ (normal case). These simulations consisted of randomly choosing (η_1, \dots, η_n) with $0 \leq \eta_i \leq \eta_0$.

However, for $c > 1$ there are counterexamples. For example, when $c = 1.5$ and again assuming common part tolerances $|a_i|T_i = T$ for all i , $\eta_0 = .2$, $n = 3$, and $K_0 = n\eta_0 T + c(1 - \eta_0)T\sqrt{n}$ as $T_A^{c, \Delta, \text{arith}}$ in (12), then $P(Z \in J) = .99802$ when $\eta_1 = \dots = \eta_n = 0$ and $P(Z \in J) = .99865$ when $\eta_1 = \dots = \eta_n = \eta_0 = .2$. As n increases this counterexample disappears. One may argue that the amount by which the conjecture is broken in this particular example is negligible, since the ratio $\rho = .99802/.99865 = .999369$ is very close to one.

Other simulations with $c > 1$ show that the conjecture seems to hold (assuming common part tolerances, $\eta_0 = .2$ and $n \geq 3$) as long as $c \leq 1.2905$. For $n = 2$ the conjecture seems to hold as long as $c \leq 1.05$. Also observed in all these simulations was that by far the closest competitor to our conjecture arises for $\eta_1 = \dots = \eta_n = 0$.

Although any violation of the conjecture may be only mild, from a purely mathematical point of view its validity cannot hold without further restrictions. It is possible that this violation is only an artifact of using a normal approximation in situations (small n) where it may not be appropriate.

Proceeding as though $\eta_i = \eta_0$ for $i = 1, \dots, n$ provides us with a worst case probability lower bound we have

$$\begin{aligned}
P(|A - \nu_A| \leq K_0) &\geq P\left(\left|Z - \frac{3\eta_0 \sum_{i=1}^n |a_i| T_i}{c(1-\eta_0) \sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right| \leq \frac{3K_0}{c(1-\eta_0) \sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right) \\
&= \Phi\left(\frac{3(\eta_0 \sum_{i=1}^n |a_i| T_i + K_0)}{c(1-\eta_0) \sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right) - \Phi\left(\frac{3(\eta_0 \sum_{i=1}^n |a_i| T_i - K_0)}{c(1-\eta_0) \sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right). \quad (16)
\end{aligned}$$

If we take

$$K_0 = T_A^{c,\Delta,\text{arith}} = \eta_0 \sum_{i=1}^n |a_i| T_i + c(1-\eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2}, \quad (17)$$

as in (12) then (16) becomes

$$\begin{aligned}
P(|A - \nu_A| \leq K_0) &\geq \Phi\left(3 + \frac{6\eta_0 \sum_{i=1}^n |a_i| T_i}{c(1-\eta_0) \sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right) - \Phi(-3) \\
&\geq \Phi\left(3 + \frac{6\eta_0}{c(1-\eta_0)}\right) - \Phi(-3) \\
&\approx 1 - \Phi(-3) = .99865. \quad (18)
\end{aligned}$$

Here we have treated the term

$$\Phi\left(3 + \frac{6\eta_0}{c(1-\eta_0)}\right)$$

(the argument of Φ already strongly reduced by the second \geq above) as one, provided η_0 is not too small.

If instead we take

$$K_0 = \tilde{T}_A^{c,\Delta,\text{arith}} = \eta_0 \sum_{i=1}^n |a_i| T_i + .927c(1 - \eta_0) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2},$$

we would get

$$\begin{aligned} P(|A - \nu_A| \leq K_0) &\geq \Phi\left(3(.927) + \frac{6\eta_0}{c(1 - \eta_0)}\right) - \Phi((-3) .927) \\ &\approx 1 - \Phi(-2.782) = .9973. \end{aligned}$$

If we are concerned with the validity of the assumed conjecture, we could calculate $P(Z \in J)$ for the closest competitor $\eta_1 = \dots = \eta_n = 0$ (according to our simulation experience), i.e.,

$$\begin{aligned} P(Z \in J) &= \Phi(C(0, \dots, 0) + W(0, \dots, 0)) - \Phi(C(0, \dots, 0) - W(0, \dots, 0)) \\ &= \Phi\left(\frac{3K_0}{c\sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right) - \Phi\left(\frac{-3K_0}{c\sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right) \end{aligned} \quad (19)$$

and compare the numerical result of (19) with (16), taking the smaller of the two. In particular, with K_0 as in (17) the expression (19) becomes

$$P(Z \in J) = 2\Phi\left(3(1 - \eta_0) + \frac{3\eta_0 \sum_{i=1}^n |a_i| T_i}{c\sqrt{\sum_{i=1}^n a_i^2 T_i^2}}\right) - 1. \quad (20)$$

We can then compare our previous lower bound (18) with (20) and take the smaller of the two as our conservative assessment for the probability of successful assembly.

5.3 Statistical Stacking of Mean Shifts³

In the previous section the mean shifts were stacked in worst case fashion. Again one could say that this should rarely happen in practice. To throw some light on this let us contrast the following two causes for mean shifts.

³A more careful reasoning will be presented in the next section. The treatment here is given mainly to serve as contrast and to provide a reference on related material in the literature.

In certain situations we may be faced with an inability to properly center the manufacturing processes in which case it would be reasonable to view the process means μ_i as randomly varying around the respective nominal values ν_i , although once μ_i is realized, it will stay fixed. For some such μ_i the mean shifts $\mu_i - \nu_i$ will come out positive and for some they will come out negative, resulting in some cancellation of mean shifts when stacked. Hence the absolute assembly mean shift (if all part mean shifts are of the type just described) should be considerably less than attained through worst case or arithmetic stacking of mean shifts.

In other situations the mean shifts are a deliberate aspect of part manufacture. This happens when one makes parts to maximum material condition, preferring rework to scrap. Depending on how these maximum material condition part dimensions interact we could quite well be affected by the cumulative mean shifts in the worst possible way. As an example for this, consider the interaction of hole and fastener, where aiming for maximum material condition on both parts (undersized hole and oversized fastener) could bring both part dimensions on a collision course. Even with such underlying motives it seems rather unlikely that a manufacturing process can be centered at exactly the mean shift boundary. Thus one should view this adverse situation as mainly aiming with the mean shifts in an unfavorable direction (for stacking) but also as still having some random quality about it.

The latter case, in its extreme (without the randomness of the means), was addressed in the previous section. In this section we address the situation of process means μ_i randomly varying around the nominals ν_i . As pointed out before, this suggests applying the method of statistical tolerancing to the means themselves. However, some care has to be taken in treating the one time nature of the random mean situation.

We will assume that the means μ_i themselves are randomly chosen from the permitted mean shift interval $[\nu_i - \eta_i T_i, \nu_i + \eta_i T_i]$. It is assumed that this random choice is governed by a distribution over this interval with mean ν_i and standard deviation τ_i . As before we relate $\eta_i T_i$ to $3\tau_i$ by way of a factor c_μ depending on the distribution type used (see Table 1), i.e.,

$$c_\mu \eta_i T_i = 3\tau_i .$$

By the CLT it follows that

$$\mu_A = a_0 + a_1 \mu_1 + \dots + a_n \mu_n$$

has an approximate normal distribution with mean $\nu_A = a_0 + a_1\nu_1 + \dots + a_n\nu_n$ and standard deviation

$$\begin{aligned}\tau_A &= \sqrt{a_1^2\tau_1^2 + \dots + a_n^2\tau_n^2} \\ &= \sqrt{a_1^2(c_\mu\eta_1T_1/3)^2 + \dots + a_n^2(c_\mu\eta_nT_n/3)^2} \\ &= \frac{c_\mu}{3} \sqrt{a_1^2\eta_1^2T_1^2 + \dots + a_n^2\eta_n^2T_n^2}.\end{aligned}$$

99.73% of all sets of random process means $\{\mu_1, \dots, \mu_n\}$ will result in μ_A being within $\pm 3\tau_A$ of ν_A , i.e., $|\mu_A - \nu_A| \leq 3\tau_A$. Since ϵ_A falls with 99.73% assurance within $\pm 3\sigma_{\epsilon_A}$ (see (14)), we can say with at least 99.595% assurance that

$$|A - \nu_A| = |(\mu_A + \epsilon_A) - \nu_A| \leq |\mu_A - \nu_A| + |\epsilon_A| \leq 3\tau_A + 3\sigma_{\epsilon_A} = T_A^*,$$

where $T_A^* = 3\tau_A + 3\sigma_{\epsilon_A}$. This slightly degraded assurance level follows from

$$\begin{aligned}P(|A - \nu_A| \leq T_A^*) &\geq P(|\mu_A + \epsilon_A - \nu_A| \leq T_A^*, |\mu_A - \nu_A| \leq 3\tau_A) \\ &\geq P(|\epsilon_A - 3\tau_A| \leq T_A^*, |\mu_A - \nu_A| \leq 3\tau_A) \\ &= P(|\mu_A - \nu_A| \leq 3\tau_A) P(|\epsilon_A - 3\tau_A| \leq T_A^*) \\ &= .9973 \left[\Phi\left(\frac{3\tau_A + T_A^*}{\sigma_{\epsilon_A}}\right) - \Phi\left(\frac{3\tau_A - T_A^*}{\sigma_{\epsilon_A}}\right) \right] \\ &= .9973 \left[\Phi\left(3 + \frac{6\tau_A}{\sigma_{\epsilon_A}}\right) - \Phi(-3) \right] \\ &\approx .9973 (1 - \Phi(-3)) = .9973 \cdot .99865 = .99595.\end{aligned}$$

The factorization of the above probabilities follows from the reasonable assumption that the variability of the mean μ_i is statistically independent from the subsequent variability of ϵ_i , the part dimension variation around μ_i . This in turn implies that μ_A and ϵ_A are statistically independent and allows the above factorization.

In the above numerical calculation we approximated

$$\Phi\left(3 + \frac{6\tau_A}{\sigma_{\epsilon_A}}\right) = \Phi\left(3 + \frac{6c_\mu\sqrt{\sum_{i=1}^n a_i^2\eta_i^2T_i^2}}{c\sqrt{\sum_{i=1}^n a_i^2(1-\eta_i)^2T_i^2}}\right) \approx 1,$$

which is very reasonable when the η_i are not too small. For example, when the η_i are all the same, say $\eta_i = \eta$, and assuming $c_\mu \geq c$ and $\eta \geq .2$ then

$$\Phi\left(3 + \frac{6c_\mu\sqrt{\sum_{i=1}^n a_i^2\eta_i^2T_i^2}}{c\sqrt{\sum_{i=1}^n a_i^2(1-\eta_i)^2T_i^2}}\right) = \Phi\left(3 + \frac{6c_\mu\eta}{c(1-\eta)}\right) \geq \Phi(4.5) = .999997.$$

Concerning the 99.595% assurance level it is worthwhile contemplating its meaning, since we mixed two types of random phenomena, namely that of a one time shot of setting the part process means and that of variation as it happens over and over for each set of parts to make an assembly. If the one time shot is way off, as it can happen with very small chance or by worst case mean shifts, then it is quite conceivable that all or a large proportion of the assemblies will exceed the $\pm(3\tau_A + 3\sigma_{\epsilon_A})$ limits. This bad behavior will persist until something is done to bring the process means closer to the respective nominals.

If however, $|\mu_A - \nu_A| \leq 3\tau_A$ as it happens with high chance, then it makes little sense to take that chance into account for all subsequent assemblies. Given $|\mu_A - \nu_A| \leq 3\tau_A$, it will stay that way from now on and the fraction of assemblies satisfying $|A - \nu_A| \leq 3\tau_A + 3\sigma_{\epsilon_A}$ is at least .9973.

Using our previous representations of τ_A and σ_{ϵ_A} in terms of the part tolerances T_i we get

$$\begin{aligned} T_A^* &= c_\mu\sqrt{a_1^2\eta_1^2T_1^2 + \dots + a_n^2\eta_n^2T_n^2} \\ &\quad + c\sqrt{a_1^2(1-\eta_1)^2T_1^2 + \dots + a_n^2(1-\eta_n)^2T_n^2}. \end{aligned} \quad (21)$$

For $c = c_\mu = 1$, with normal variation for ϵ_i and the means μ_i , this was already proposed by Desmond in the discussion of Mansoor(1963). In his reply Mansoor states having considered this dual use of statistical tolerancing himself, but rejects it as producing too optimistic tolerance limits. Note that Mansoor's reluctance may be explained as follows. For $c = c_\mu = 1$ and $\eta_1 = \dots = \eta_n = \eta_0$ the above formula (21) reduces to

$$T_A^* = \sqrt{a_1^2T_1^2 + \dots + a_n^2T_n^2},$$

which coincides with the ordinary statistical tolerancing formula (2) without mean shift allowances. Note however, that in our dual use of statistical tolerancing we may have given up a little by reducing 99.73% to 99.595% in the assembly tolerance assurance.

The fact that this strong connection between (21) and (2) was not noticed by Desmond and Mansoor is probably related to their notation and to their more general setup (not all η_i are the same). Desmond suggested (in Mansoor's notation and assuming $|a_i| = 1$)

$$T_{\text{prob}} = \sqrt{\sum_{i=1}^n (T_i - t_i)^2} + \sqrt{\sum_{i=1}^n t_i^2},$$

where T_i is the tolerance for the part dimension (as we also use it) and t_i is the tolerance for the mean shift. Relating this to our notation we get $t_i = \eta_i T_i$ and Desmond's formula reads

$$T_{\text{prob}} = \sqrt{\sum_{i=1}^n (1 - \eta_i)^2 T_i^2} + \sqrt{\sum_{i=1}^n \eta_i^2 T_i^2}.$$

When these fractions η_i are all the same, as they reasonably may be, we get

$$T_{\text{prob}} = (1 - \eta_0) \sqrt{\sum_{i=1}^n T_i^2} + \eta_0 \sqrt{\sum_{i=1}^n T_i^2} = \sqrt{\sum_{i=1}^n T_i^2}.$$

More generally, we have the following inequality

$$\sqrt{\sum_{i=1}^n (1 - \eta_i)^2 T_i^2} + \sqrt{\sum_{i=1}^n \eta_i^2 T_i^2} \geq \sqrt{\sum_{i=1}^n T_i^2},$$

with equality exactly when $\eta_1 = \dots = \eta_n$. This is seen from an application of the Cauchy-Schwarz inequality, i.e.,

$$\sqrt{\sum_{i=1}^n (1 - \eta_i)^2 T_i^2} \sqrt{\sum_{i=1}^n \eta_i^2 T_i^2} \geq \sum_{i=1}^n \eta_i (1 - \eta_i) T_i^2$$

or

$$2 \sqrt{\sum_{i=1}^n (1 - \eta_i)^2 T_i^2} \sqrt{\sum_{i=1}^n \eta_i^2 T_i^2} \geq \sum_{i=1}^n T_i^2 (1 - \eta_i^2 - (1 - \eta_i)^2),$$

with equality exactly when for some factor β we have $(1 - \eta_i)T_i = \beta\eta_iT_i$ for all i , and hence when $\eta_1 = \dots = \eta_n$. Rearranging terms we get

$$\sum_{i=1}^n (1 - \eta_i)^2 T_i^2 + \sum_{i=1}^n \eta_i^2 T_i^2 + 2 \sqrt{\sum_{i=1}^n (1 - \eta_i)^2 T_i^2} \sqrt{\sum_{i=1}^n \eta_i^2 T_i^2} \geq \sum_{i=1}^n T_i^2$$

which is simply the square of the claimed inequality.

Since statistical tolerance stacking is used both on the stack of mean shifts and on the part variation it will not surprise that the assembly tolerance stack T_A^* typically grows on the order of \sqrt{n} . This is the case when the $|a_i|T_i$ and mean shift fractions η_i are comparable and it is seen most clearly when $|a_i|T_i = T$ and $\eta_i = \eta$ for all $i = 1, \dots, n$. In that case (21) simplifies to

$$T_A^* = \sqrt{n} (c_\mu \eta + c(1 - \eta)) T .$$

Returning to the general form of (21), but making the reasonable assumption $\eta_1 = \dots = \eta_n = \eta$, we get

$$\begin{aligned} T_A^*(\eta) &= (c_\mu \eta + c(1 - \eta)) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2} \\ &= (c + \eta(c_\mu - c)) \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2} \end{aligned} \quad (22)$$

which then can be taken as assembly tolerance stack with assurance of at least 99.595%. If we bound the mean shift by the condition $\eta \leq \eta_0$ and if $c_\mu > c$, then $T_A^*(\eta)$ is bounded by $T_A^*(\eta_0)$. If $c_\mu \leq c$ we have that $T_A^*(0) \geq T_A^*(\eta)$, i.e., the worst case arises when all the variability comes from part to part variability and none from mean shifts. Note however that in this case the part variability is spread over the full tolerance interval I_i .

One reasonable choice for c and c_μ is $c = 1$ and $c_\mu = \sqrt{3} = 1.732$, using normal part dimension variation and a fairly conservative uniform mean shift variation. Under these assumption and limiting the mean shift to a fraction $\eta_0 = .2$, stacking formula (22) becomes

$$T_A^*(.2) = 1.146 \sqrt{a_1^2 T_1^2 + \dots + a_n^2 T_n^2}$$

which seems rather mild. An explanation for this follows in the next section.

5.4 Statistical Stacking of Mean Shifts Revisited

In the previous section we exploited the cancellation of variation in the part mean shifts. It was assumed there that the variation of the mean shifts $\Delta_i = \mu_i - \nu_i$ should be limited to the respective intervals $[-\eta_i T_i, \eta_i T_i]$ and that any such mean shift should be accompanied by a reduction in part variability, i.e., in σ_i . During our treatment we assumed this to mean that the original $\sigma_i = T_i/3$ (in its worst case form) be reduced by the factor $1 - \eta_i$. This is not correct. It would be correct, if the mean shift was at the boundary of the allowed interval, i.e., $\Delta_i = \pm \eta_i T_i$. Since we exploit the fact that the mean shifts vary randomly over their respective intervals, we should also allow for the fact that some of the mean shifts are not all that close to the boundary of the allowed intervals. This in turn allows the standard deviation of ϵ_i to be larger. Recall that we allow mean shifts as long as we maintain $C_{pk} \geq 1$. In the previous section we took advantage of mean shift variation cancellation but kept the residual part to part variability smaller than might actually occur. This fallacy was probably induced by the treatment of arithmetic stacking of mean shifts. It also explains why the tolerance stacking formulas of the previous section appeared optimistic. In this section we will rectify this defect. It appears that the results presented here are new.

Recall that the part dimension measurement was modeled as

$$X_i = \nu_i + \Delta_i + \epsilon_i ,$$

where the mean shift $\Delta_i = \mu_i - \nu_i$ was treated as a random variable having a distribution over the interval $[-\eta_i T_i, \eta_i T_i]$ with mean zero and standard deviation τ_i . The latter relates to T_i via

$$3\tau_i = c_\mu \eta_i T_i ,$$

where c_μ depends on the type of distribution that is assumed for the Δ_i , see Table 1. In the following let

$$Y_i = \frac{\Delta_i}{\eta_i T_i} \quad \text{so that} \quad \Delta_i = Y_i \eta_i T_i ,$$

where Y_i is a random variable with values in $[-1, 1]$, having mean zero and standard deviation $\tau_i/(\eta_i T_i) = c_\mu/3$.

Conditional on $|\Delta_i|$, the part to part variation term ϵ_i is assumed to vary over the interval

$$J_i(|\Delta_i|) = [-(T_i - |\Delta_i|), (T_i - |\Delta_i|)] = [-(1 - \eta_i Y_i)T_i, (1 - \eta_i Y_i)T_i].$$

This is the part that is different from our previous treatment, where we assumed that ϵ_i varies over

$$J_i(\eta_i T_i) = [-(T_i - \eta_i T_i), (T_i - \eta_i T_i)] = [-(1 - \eta_i)T_i, (1 - \eta_i)T_i],$$

i.e., we replaced $|\Delta_i|$ by its maximally permitted value $\eta_i T_i$. This artificially reduced the resulting σ_i for ϵ_i and led to optimistic tolerance stacking formulas. Given Y_i , the standard deviation $\sigma_i = \sigma_i(Y_i)$ of ϵ_i is such that

$$3\sigma_i(Y_i) = cT_i(1 - |Y_i|\eta_i),$$

where as before c is a constant that takes into account the particular distribution of ϵ_i . See Table 1 for constants c corresponding to various distribution types. Assuming the above form of $\sigma_i(Y_i)$ in relation to $T_i(1 - |Y_i|\eta_i)$ is conservative, since the actual $\sigma_i(Y_i)$ could be smaller. The above form is derived from $C_{pk} = 1$, but the actual requirement says $C_{pk} \geq 1$.

With the above notation we can express the assembly deviation from nominal as follows

$$D_A = A - \nu_A = \sum_{i=1}^n a_i \Delta_i + \sum_{i=1}^n a_i \epsilon_i = \sum_{i=1}^n a_i \eta_i T_i Y_i + \sum_{i=1}^n a_i \epsilon_i. \quad (23)$$

The first sum on the right represents the assembly mean shift effect. Since it involves the random terms Y_i , it is, for large enough n , approximately normally distributed with mean zero and standard deviation

$$\tau_A = \sqrt{\sum_{i=1}^n a_i^2 \tau_i^2} = \sqrt{\sum_{i=1}^n a_i^2 \eta_i^2 T_i^2 \left(\frac{c_\mu}{3}\right)^2}.$$

Since this assembly shift is a one time effect, we can bound it by

$$\left| \sum_{i=1}^n a_i \eta_i T_i Y_i \right| \leq 3\tau_A = c_\mu \sqrt{\sum_{i=1}^n a_i^2 \eta_i^2 T_i^2}$$

with 99.73% assurance. The second sum on the right of equation (23) contains a mixture of one time variation (through the Y_i which affect the standard deviations $\sigma_i(Y_i)$) and of variation that is new for each assembly, namely the variation of the ϵ_i , once the Y_i and thus the $\sigma_i(Y_i)$ are fixed.

Given $\mathbf{Y} = (Y_1, \dots, Y_n)$, resulting in fixed $(\sigma_1(Y_1), \dots, \sigma_n(Y_n))$, we can treat $\sum_{i=1}^n a_i \epsilon_i$ as normal (assuming the ϵ_i to be normal with $c = 1$) or, for

large enough n and appealing to the CLT, as approximately normal with mean zero and standard deviation

$$\sigma_A(\mathbf{Y}) = \sqrt{\sum_{i=1}^n a_i^2 T_i^2 (1 - |Y_i| \eta_i)^2 \left(\frac{c}{3}\right)^2}.$$

Conditional on \mathbf{Y} we can expect 99.73% of all assemblies to satisfy

$$\left| \sum_{i=1}^n a_i \epsilon_i \right| \leq 3\sigma_A(\mathbf{Y}) = c \sqrt{\sum_{i=1}^n a_i^2 T_i^2 (1 - |Y_i| \eta_i)^2}.$$

$\sigma_A(\mathbf{Y})$ is largest when all $Y_i = 0$, i.e.,

$$\sigma_A(\mathbf{Y}) \leq \sigma_A(\mathbf{0}) = \frac{c}{3} \sqrt{\sum_{i=1}^n a_i^2 T_i^2}.$$

Regardless of the value of \mathbf{Y} we thus have

$$\left| \sum_{i=1}^n a_i \epsilon_i \right| \leq 3\sigma_A(\mathbf{0}) = c \sqrt{\sum_{i=1}^n a_i^2 T_i^2}$$

with 99.73% assurance.

Since the mean shift variation, reflected in \mathbf{Y} , is a one time effect, we stack the two contributions $\sum_{i=1}^n a_i \epsilon_i$ and $\sum_{i=1}^n a_i \eta_i T_i Y_i$ in worst case fashion, i.e., arithmetically. This means we bound both terms statistically and then add the bounds. Thus we obtain the following conservative tolerance stacking formula

$$\begin{aligned} T_A^{**} &= 3\tau_A + 3\sigma_A(\mathbf{0}) \\ &= c_\mu \sqrt{\sum_{i=1}^n a_i^2 \eta_i^2 T_i^2} + c \sqrt{\sum_{i=1}^n a_i^2 T_i^2}. \end{aligned} \quad (24)$$

This stacking formula is conservative for three reasons: 1) we stacked the tolerances of the two contributions $\sum_{i=1}^n a_i \epsilon_i$ and $\sum_{i=1}^n a_i \eta_i T_i Y_i$ in worst case fashion, 2) we used the conservative upper bounds $\sigma_i(0)$ on the standard deviations $\sigma_i(Y_i)$, and 3) we assumed that the $\sigma_i(Y_i)$ filled out the remaining space left free by the mean shift, i.e., $3\sigma_i(Y_i) = cT_i(1 - |Y_i|\eta_i)$.

When the η_i , which bound the fractions of mean shift, are all the same, say $\eta_i = \eta_0$, then (24) reduces to

$$T_A^{**} = (\eta_0 c_\mu + c) \sqrt{\sum_{i=1}^n a_i^2 T_i^2} . \quad (25)$$

This is again the RSS formula with an inflation factor. However, it is not motivated by increasing the variation around the nominal center. Instead it is motivated by statistically stacking the mean shifts, and stacking that in worst case fashion with the statistically toleranced part to part variation.

When the $|a_i|T_i$ are all the same or approximately comparable, one sees again that T_A^{**} is of order \sqrt{n} . This is the main gain in statistically tolerancing the mean shifts. Recall that worst case mean shift stacking led to assembly tolerances of order n .

As a particular example, let $\eta_0 = .2$ and consider normal part to part variation for all parts, i.e., $c = 1$, and uniform mean shift variation⁴ over the interval $[-\eta_0 T_i, \eta_0 T_i]$, i.e., $c_\mu = \sqrt{3}$. Then (25) becomes

$$T_A^{**} = 1.346 \sqrt{\sum_{i=1}^n a_i^2 T_i^2} .$$

A less conservative and more complicated approach to the above tolerancing problem looks at the sum of both contributions $\sum_{i=1}^n a_i \epsilon_i$ and $\sum_{i=1}^n a_i \eta_i T_i Y_i$ conditionally given \mathbf{Y} . Conditional on \mathbf{Y} , the sum of these two contributions, namely $D_A = A - \nu_A$, has an approximate normal distribution with mean

$$\mu_{D_A}(\mathbf{Y}) = \sum_{i=1}^n a_i \eta_i T_i Y_i$$

and standard deviation

$$\sigma_A(\mathbf{Y}) = \sqrt{\sum_{i=1}^n a_i^2 T_i^2 (1 - |Y_i| \eta_i)^2 \left(\frac{c}{3}\right)^2} .$$

Thus the interval $I_0(\mathbf{Y})$, given by

$$\mu_{D_A}(\mathbf{Y}) \pm 3\sigma_A(\mathbf{Y}) ,$$

⁴This assumption is conservative in the sense that among all symmetric and unimodal distributions over that interval the uniform distribution is the one with highest variance.

will bracket D_A with conditional probability .9973 for a fixed value of \mathbf{Y} . For some values of \mathbf{Y} the interval $I_0(\mathbf{Y})$ will slide far to the right and for some it will slide far to the left. Note that changing the signs of all Y_i flips the interval around the origin. Thus for every interval with extreme right endpoint T_B there is a corresponding interval with extreme left endpoint $-T_B$.

The basic remaining problem is to find a high probability region B for \mathbf{Y} , say $P(\mathbf{Y} \in B) = .9973$, such that for all $\mathbf{Y} \in B$ we have

$$-T_B \leq \mu_{D_A}(\mathbf{Y}) - 3\sigma_A(\mathbf{Y}) \quad \text{and} \quad \mu_{D_A}(\mathbf{Y}) + 3\sigma_A(\mathbf{Y}) \leq T_B .$$

Here T_B denotes the most extreme upper endpoint of $I_0(\mathbf{Y})$ as \mathbf{Y} varies over B . It is also assumed that B is such that with $\mathbf{Y} \in B$ we also have $-\mathbf{Y} \in B$, so that $-T_B$ is the most extreme lower endpoint of $I_0(\mathbf{Y})$ as \mathbf{Y} varies over B .

The event $\mathbf{Y} \notin B$ is very rare (with .27% chance), i.e., about .27% of all assembly setups (with part processes feeding parts to that assembly) will have mean shifts extreme enough so that $\mathbf{Y} \notin B$.

Let us look at the right endpoint of $I_0(\mathbf{Y})$

$$\begin{aligned} \mu_{D_A}(\mathbf{Y}) + 3\sigma_A(\mathbf{Y}) &= \sum_{i=1}^n a_i \eta_i T_i Y_i + c \sqrt{\sum_{i=1}^n a_i^2 T_i^2 (1 - |Y_i| \eta_i)^2} \\ &= \sqrt{\sum_{i=1}^n a_i^2 T_i^2} \left(\sum_{i=1}^n \eta_i w_i Y_i + c \sqrt{\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_i)^2} \right) \end{aligned}$$

with weights

$$w_i = \frac{a_i T_i}{\sqrt{\sum_{j=1}^n a_j^2 T_j^2}} \quad \text{so that} \quad \sum_{i=1}^n w_i^2 = 1 .$$

Note that $\sqrt{\sum_{i=1}^n a_i^2 T_i^2}$ is the usual RSS tolerance stacking formula. Thus we may view the factor

$$F_U(\mathbf{Y}) = \sum_{i=1}^n \eta_i w_i Y_i + c \sqrt{\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_i)^2}$$

as an RSS inflation factor. Similarly one can write

$$\mu_{D_A}(\mathbf{Y}) - 3\sigma_A(\mathbf{Y}) = \sqrt{\sum_{i=1}^n a_i^2 T_i^2} \left(\sum_{i=1}^n \eta_i w_i Y_i - c \sqrt{\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_i)^2} \right)$$

$$= F_L(\mathbf{Y}) \sqrt{\sum_{i=1}^n a_i^2 T_i^2}$$

with

$$F_L(\mathbf{Y}) = \sum_{i=1}^n \eta_i w_i Y_i - c \sqrt{\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_i)^2}$$

having the same distribution as $-F_U(\mathbf{Y})$ since the Y_i are assumed to have a symmetric distribution on $[-1, 1]$.

At this point $F_U(\mathbf{Y})$ and $F_L(\mathbf{Y})$ still depend on the conditioning mean shift variables \mathbf{Y} . By bounding $F_U(\mathbf{Y})$ with high probability from above by F_0 and $F_L(\mathbf{Y})$ from below by $-F_0$, for example $P(F_U(\mathbf{Y}) \leq F_0) = P(F_L(\mathbf{Y}) \geq -F_0) = .99865$, we would then have the following more refined, mean shift adjusted tolerance stack formula

$$T_A^\diamond = F_0 \sqrt{\sum_{i=1}^n a_i^2 T_i^2}.$$

We will find later that the inflation factor F_0 in front of the RSS term is again independent of n and thus the growth of T_A^\diamond is dictated by the growth of the RSS contribution which is of order \sqrt{n} .

The complication with obtaining the distributions of $F_U(\mathbf{Y})$ and $F_L(\mathbf{Y})$ is that the two summands in each are correlated (involving the same Y_i) and that the second sum is under a square root. Thus it is difficult to get the exact distribution of $F_U(\mathbf{Y})$ or $F_L(\mathbf{Y})$ analytically. Also, for small n we should not appeal to the CLT. To shed some light on this we performed some simulations under the following reasonable and simplifying assumptions. Namely, $\eta_1 = \dots = \eta_n = \eta_0$, Y_i is uniform over the interval $[-1, 1]$, i.e., $c_\mu = \sqrt{3}$, and ϵ_i is normal, i.e., $c = 1$. Furthermore, we let the $|a_i|T_i$ be all the same, so that $w_i = 1/\sqrt{n}$ for all i . Then

$$F_U(\mathbf{Y}) = \eta_0 \sum_{i=1}^n \frac{Y_i}{\sqrt{n}} + \sqrt{\sum_{i=1}^n \frac{1}{n} (1 - |Y_i| \eta_0)^2}$$

and we simulated the distribution of $F_U(\mathbf{Y})$ with 10,000 replications for each $n = 2, \dots, 10, 12, 15, 20$. This exercise was repeated for a different choice of $|a_i|T_i$, leading to linearly increasing weights $w_i = m \cdot i$ with $m =$

Figure 5: Distribution of RSS Factors, $n = 2, 3, 4$

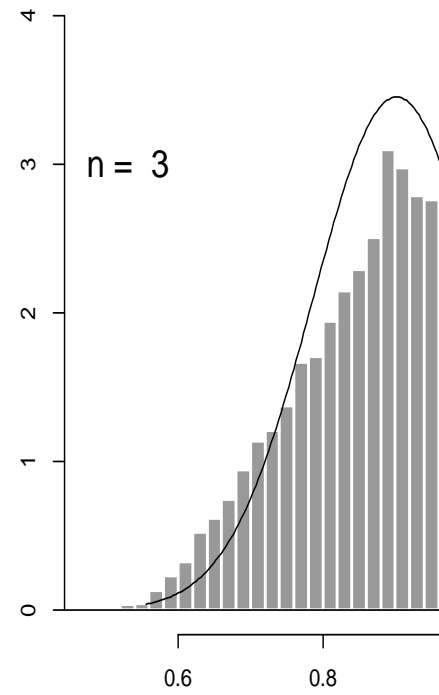
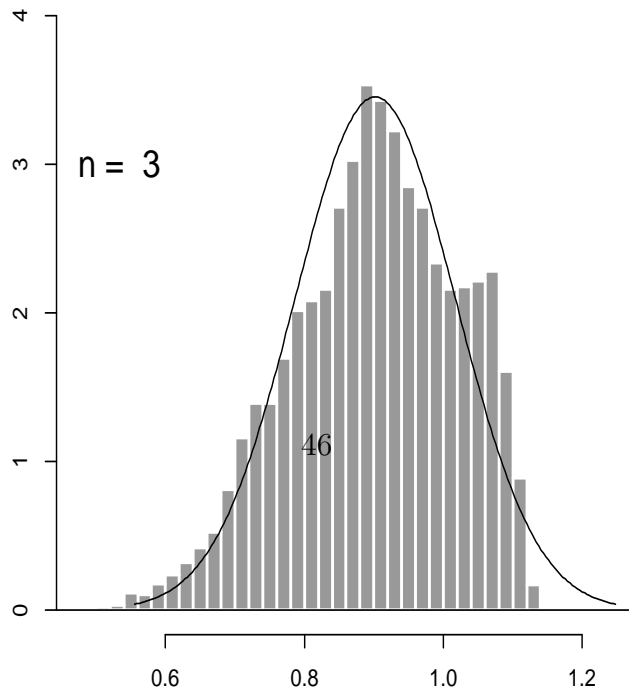
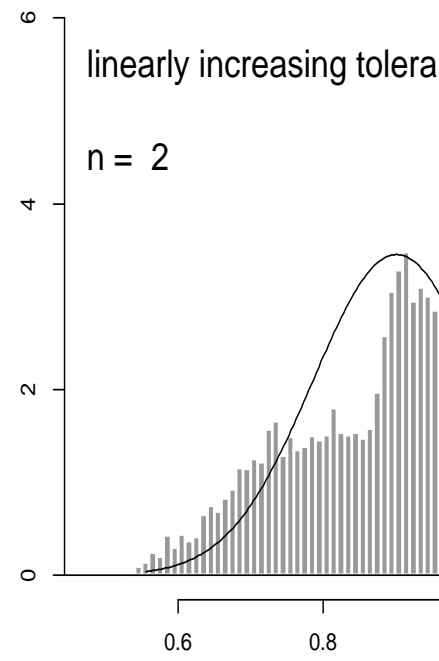
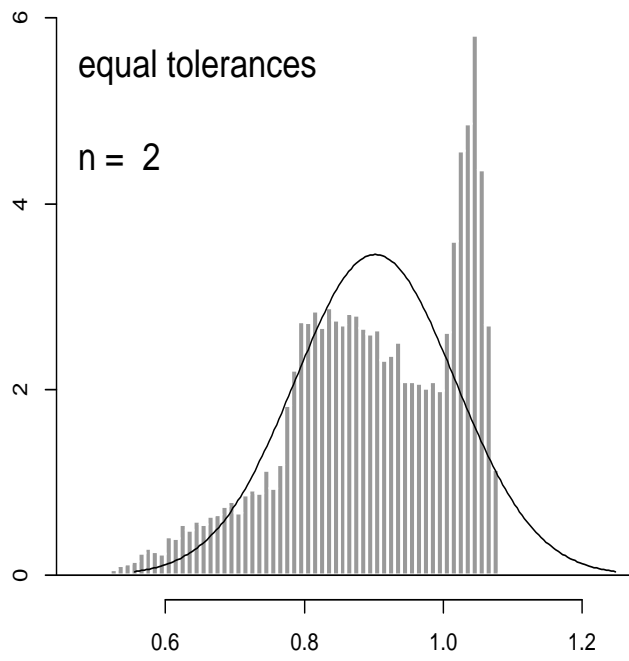


Figure 6: Distribution of RSS Factors, $n = 5, 6, 7$

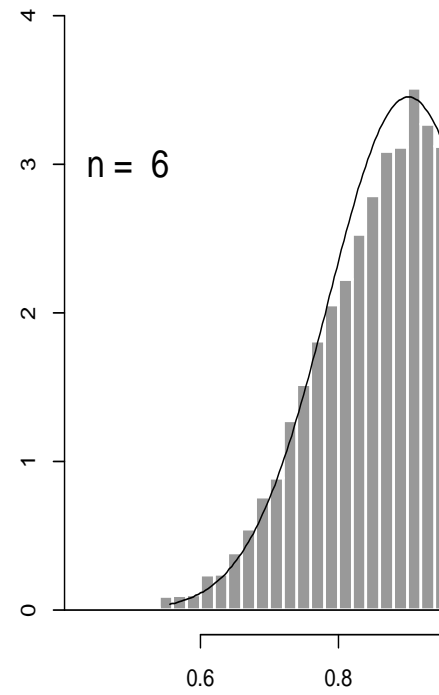
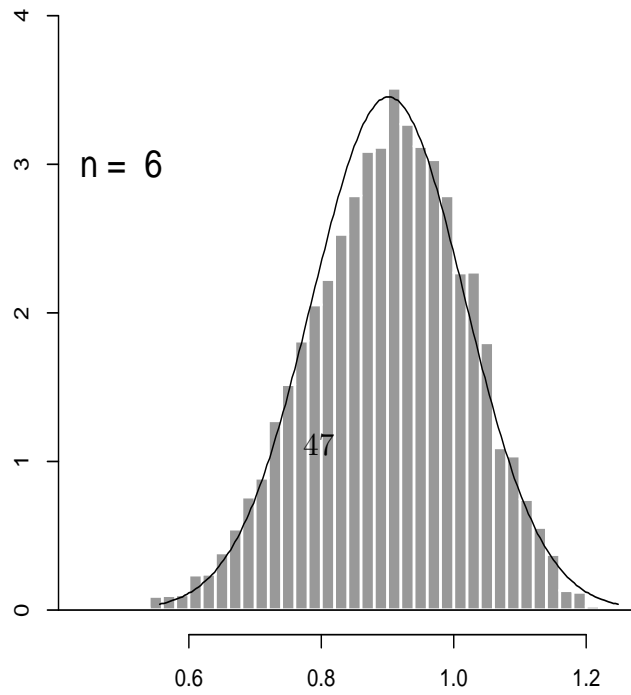
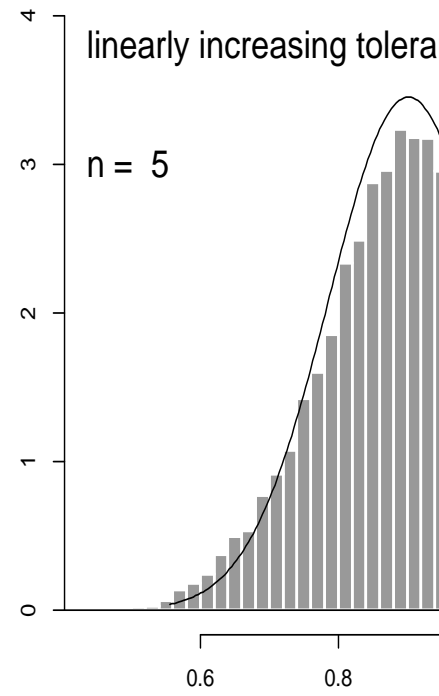
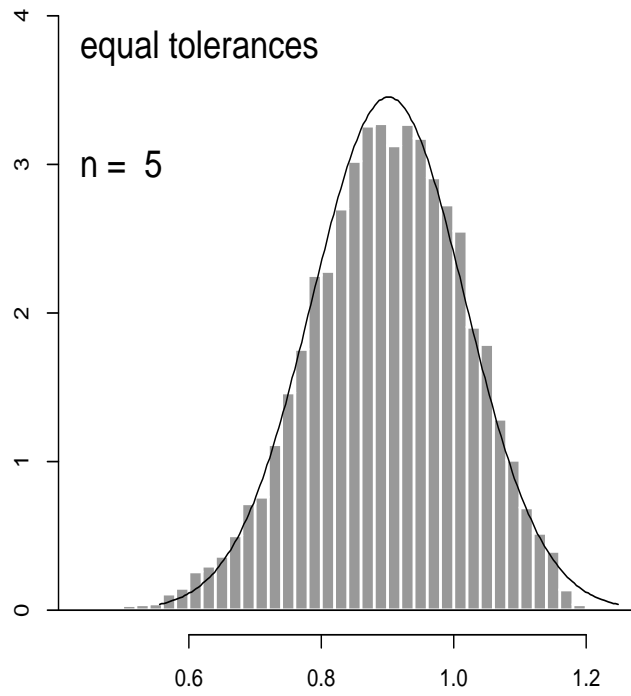


Figure 7: Distribution of RSS Factors, $n = 8, 9, 10$

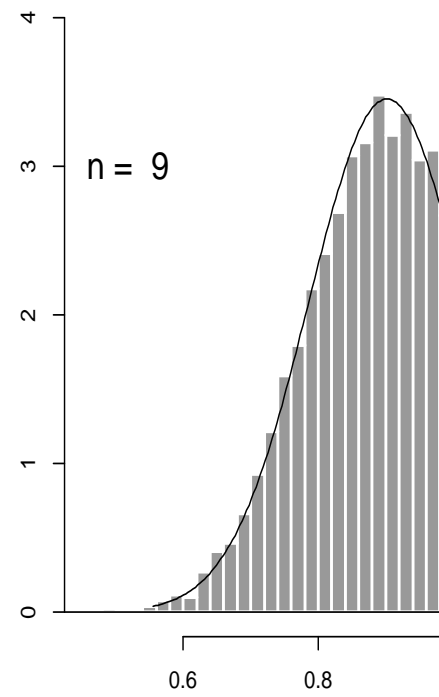
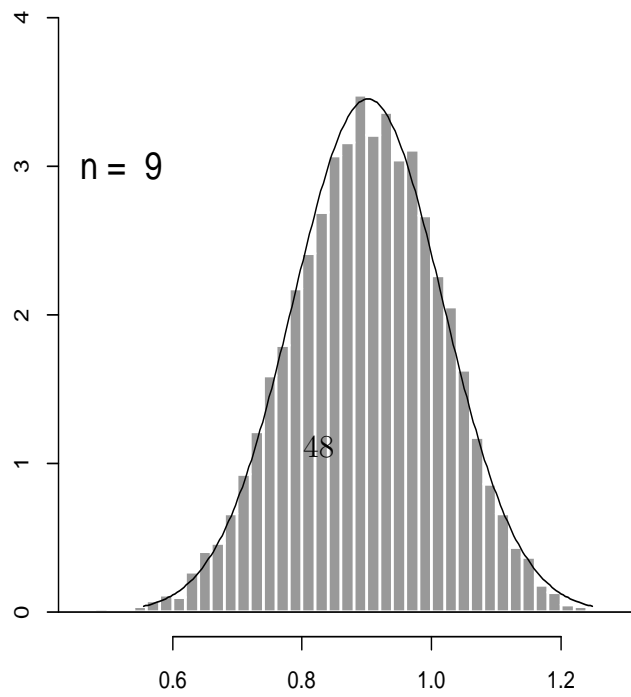
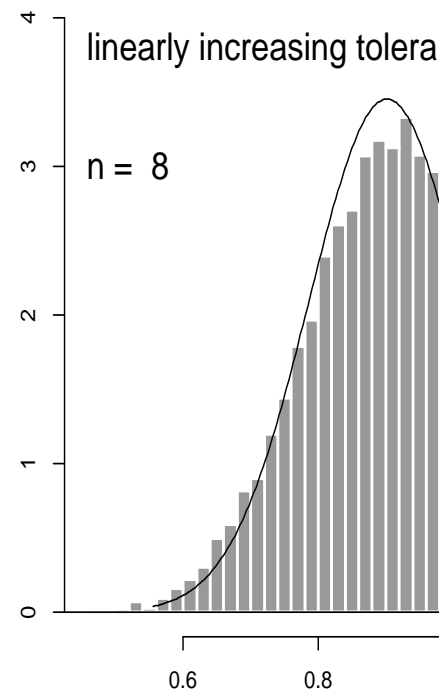
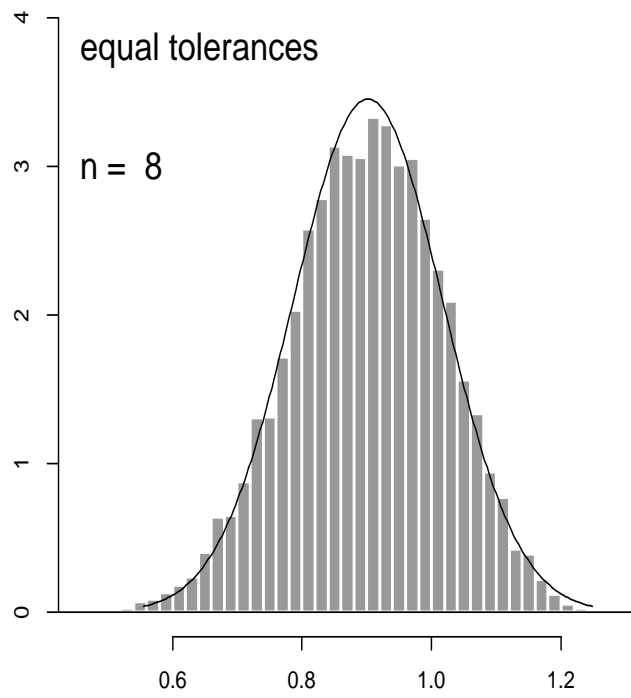
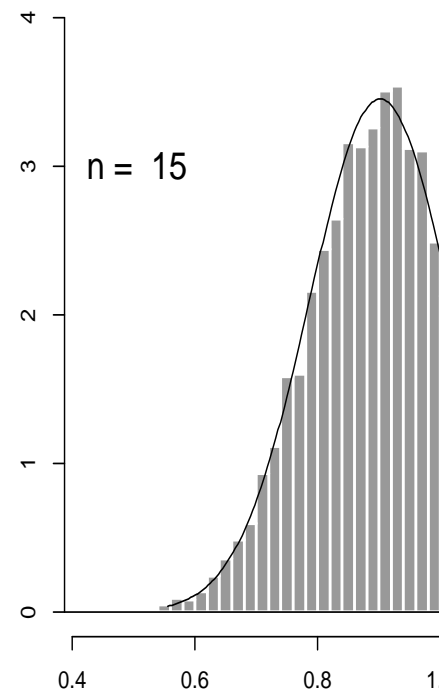
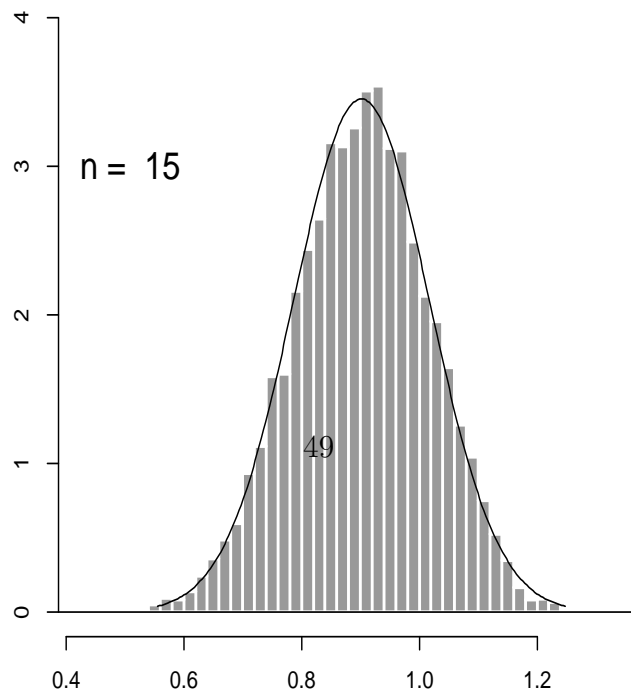
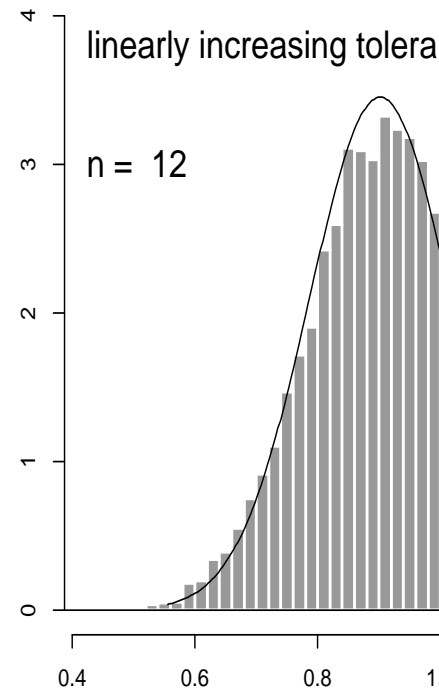
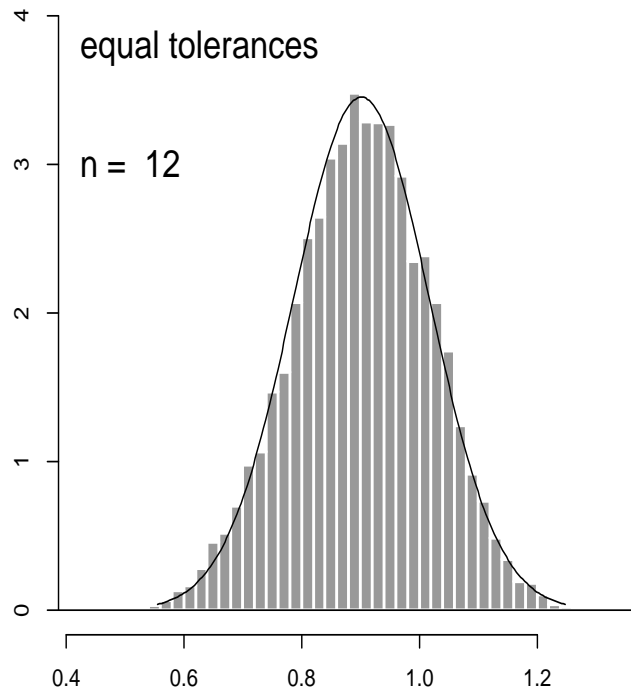


Figure 8: Distribution of RSS Factors, $n = 12, 15, 20$



$\sqrt{6/[(n(n+1)(2n+1))]}$. The histograms of $F_U(\mathbf{Y})$ are displayed side by side for both sets of simulations in Figures 5-8.

Also shown are superimposed normal density curves, which seem to fit reasonably well for $n \geq 5$. These normal distributions arise from applying the CLT to

$$\eta_0 \sum_{i=1}^n w_i Y_i ,$$

i.e., treating it as normal with mean zero (since $E(Y_i) = 0$) and variance

$$\text{var} \left(\eta_0 \sum_{i=1}^n w_i Y_i \right) = \eta_0^2 \sum_{i=1}^n w_i^2 \sigma_Y^2 = \eta_0^2 \sigma_Y^2 ,$$

independent of the weights w_i . With Y being uniform on $[-1, 1]$, we have $\sigma_Y^2 = 4/12 = 1/3$. Furthermore, we treat the other term in $F_U(\mathbf{Y})$, i.e.,

$$\sqrt{\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_0)^2} ,$$

as an approximate constant, since the term under the square root is approximately constant, the extent of constancy depending somewhat on the weights w_i . However, the constant itself is independent of the weights. To see this note that

$$\begin{aligned} E \left(\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_0)^2 \right) &= \left(\sum_{i=1}^n w_i^2 \right) E \left((1 - |Y_1| \eta_0)^2 \right) \\ &= E \left(1 - 2|Y_1| \eta_0 + Y_1^2 \eta_0^2 \right) \\ &= 1 - \eta_0 + \frac{\eta_0^2}{3} , \end{aligned}$$

where in the last line we again assumed for Y_1 a uniform distribution over the interval $[-1, 1]$. Furthermore

$$\begin{aligned} \text{var} \left(\sum_{i=1}^n w_i^2 (1 - |Y_i| \eta_0)^2 \right) &= \sum_{i=1}^n w_i^4 \text{var} \left((1 - |Y_i| \eta_0)^2 \right) \\ &= \left(\sum_{i=1}^n w_i^4 \right) \text{var} \left((1 - |Y_1| \eta_0)^2 \right) \end{aligned}$$

$$= \kappa_0^2 \left(\sum_{i=1}^n w_i^4 \right).$$

Assuming a uniform distribution over the interval $[-1, 1]$ we have

$$\begin{aligned} \kappa_0^2 = \text{var} \left((1 - |Y_1| \eta_0)^2 \right) &= E(1 - |Y_1| \eta_0)^4 - \left(E(1 - |Y_1| \eta_0)^2 \right)^2 \\ &= 1 - 4\eta_0 E(|Y_1|) + 6\eta_0^2 E(|Y_1|^2) - 4\eta_0^3 E(|Y_1|^3) \\ &\quad + \eta_0^4 E(|Y_1|^4) - \left(1 - \eta_0 + \frac{\eta_0^2}{3} \right)^2 \\ &= \frac{\eta_0^2}{3} \left(1 - \eta_0 + \frac{4}{15} \eta_0^2 \right). \end{aligned}$$

For $\eta_0 = .2$ we have $\kappa_0 = .104$. The sum $\sum_{i=1}^n w_i^4$ tends to be small as n gets large. For example, if $w_1 = \dots = w_n = 1/\sqrt{n}$, we get $\sum_{i=1}^n w_i^4 = 1/n \rightarrow 0$ as $n \rightarrow \infty$. Similarly, if

$$w_i = m i \quad \text{with} \quad m = \sqrt{\frac{6}{n(n+1)(2n+1)}}$$

we have

$$\begin{aligned} \sum_{i=1}^n w_i^4 &= m^4 \sum_{i=1}^n i^4 \\ &= \left(\frac{6}{n(n+1)(2n+1)} \right)^2 \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30} \\ &= \frac{6}{5} \frac{3n^2+3n-1}{n(n+1)(2n+1)} \\ &\approx \frac{9}{5n} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \end{aligned}$$

Using the above assumption of $Y_i \sim U([-1, 1])$ and $\eta_1 = \dots = \eta_n = \eta_0$ we can treat $F_U(\mathbf{Y})$ as approximately normal with mean and standard deviation

$$\mu_F = \sqrt{1 - \eta_0 + \frac{\eta_0^2}{3}} \quad \text{and} \quad \sigma_F = \frac{\eta_0}{\sqrt{3}},$$

respectively. As pointed out above, the distribution of $F_L(\mathbf{Y})$ is the same as that of $-F_U(\mathbf{Y})$, i.e., approximately normal with mean and standard deviation

$$-\mu_F = -\sqrt{1 - \eta_0 + \frac{\eta_0^2}{3}} \quad \text{and} \quad \sigma_F = \frac{\eta_0}{\sqrt{3}},$$

respectively.

When $\eta_0 = .2$, as in the simulations underlying Figure 5-8, we have $\mu_F = .90185$ and $\sigma_F = .1155$. The density corresponding to this normal distribution is superimposed on each of the histograms in Figure 5-8. The fit appears quite reasonable for $n \geq 5$. According to the above derivation and verified by the simulations, the distribution of $F_U(\mathbf{Y})$ settles down rapidly, i.e., it does not move with n as n gets larger. Only for $n \leq 4$ do we have larger discrepancies between the histograms and the approximating normal density. Note however, that for all $n \geq 2$ the right tail of the “approximating” normal distribution always extends farther out than that of the simulated histogram distribution. This is illustrated more explicitly in Figure 9, which shows the p -quantiles of the 10,000 simulated $F_U(\mathbf{Y})$ values plotted against n (for the case with equal tolerances) for high values of p , namely $p = .99, .995, .999, .9995$.