The Gaussian Assumption in Enventive Tolerance Analysis

Dr. James Hurt

December 28, 2007

Abstract

This paper discusses the impact and implications of assuming that a contributor to a tolerance stack up or the analyzed variable in a tolerance analysis has a Gaussian or Normal probability distribution.

1 Introduction

This paper discusses the impact and implications of assuming that a contributor to a tolerance stack up or the analyzed variable in a tolerance analysis has a Gaussian or Normal probability distribution.

The tolerance analysis system in Enventive depends upon linearization and three basic theorems from probability theory. When the user performs a tolerance analysis on a derived variable, that derived variable is considered as a function of the residuals of all the constraints and dimension constraints which determine the value of the derived variable.

2 Linearization

Let r_i be the residual of the i-th equation, and let d be the derived variable. Then d is some implied function of the residuals:

$$d = f(r_0, r_1, \cdots, r_{n-1}) \tag{1}$$

This function $f(\cdots)$ is not explicitly constructed in any manner or form, but it is "computed" when all the residuals r_i are equal to zero by solving all the constraints. This function is linearized by expanding it into its multi-dimensional Taylor's series [4]:

$$d = d_N + \sum_{i=0}^{i<n} \left(\frac{\partial f}{\partial r_i}\right) r_i + \text{HOT} \tag{2}$$

where:

- d_N is the nominal value for the derived variable, which is the value of the function $f(\cdots)$ when all the residuals r_i are equal to zero,
- $\frac{\partial f}{\partial r_i}$ is the partial derivative of the function $f(\cdots)$ evaluated at the nominal geometry when all the residuals r_i are equal to zero, often called the sensitivity coefficient, and

- HOT stands for Higher Order Terms.

By ignoring the higher order terms, we obtain a linearized approximation to how the derived variable will respond to changes in the values of the residuals. The tolerance analysis system of Enventive computes and uses this linearized approximation.

3 Relationship of the Means

The first of the basic theorems from probability theory is that the mean of a linear combination of random variables is the same linear combination of the means [5]. Assume that each residual r_i is a random variable that has a mean μ_i, then the linearized approximation, equation (2), to the derived variable d is a random variable which has a mean[1] and this mean is μ_d:

$$\mu_d = d_N + \sum_{i=0}^{i<n} \left(\frac{\partial f}{\partial r_i} \right) \mu_i \tag{3}$$

Note that each of the random variables r_i can have a different probability distribution: one can have a uniform distribution, another can have a beta distribution, etc. For computing the mean μ_d, the only assumption we had to make was that each of the contributor random variables has a mean μ_i.

4 Relationship of the Variances

The second of the basic theorems from probability theory is that the variance of a linear combination of independent random variables is a different linear combination of the variances [5]. The variance of a random variable is simply the square of the standard deviation of that random variable. Assume that each residual r_i is a random variable that has a variance σ_i^2, then the linearized approximation, equation (2), to the derived variable d is a random variable which has a variance[2] and this variance is σ_d^2:

$$\sigma_d^2 = \sum_{i=0}^{i<n} \left(\frac{\partial f}{\partial r_i} \right)^2 \sigma_i^2 \tag{4}$$

Note that each of the random variables r_i can have a different probability distribution: one can have a uniform distribution, another can have a beta distribution, etc. For computing the standard deviation σ_d the only two assumptions we had to make were that each of the contributor random variables has a standard deviation σ_i, and that each of these random variables is independent of all the others.

For equation (4) we have assumed that the various residuals are statistically independent. This assumption is not always correct, for example when the contributor is a true position tolerance. When this assumption is not correct we have to add numerous co-variance terms to this equation.

[1] Not all random variables have a mean.

[2] Not all random variables that have a mean also have a variance, but all random variables that have a variance also have a mean.

5 The Central Limit Theorem

In equations (3) and (4), if each of the random variables r_i has a Gaussian distribution, then the linearized approximation to the analyzed variable also has a Gaussian distribution. But what happens when one or more of the random variables r_i do not have a Gaussian distribution?

The third of the basic theorems from probability theory is called the *Central Limit Theorem* [2], and states that when there are several sensitivity coefficients that are large and approximately equal, the probability distribution for the derived variable d will be approximately Gaussian even when these random variables do not have Gaussian distributions.

Thus, when there are more than two contributors with large and approximately equal sensitivity coefficients, it is safe to assume that the derived variable has a Gaussian distribution. But, when there are only one or two contributors with large sensitivity coefficients, it is **not safe** to assume that the derived variable has a Gaussian distribution unless these contributors also have Gaussian distributions.

If there are only one or two significant contributors (random variables r_i) with large sensitivity coefficients which do not have Gaussian distributions, then the computations for the mean and standard deviation of the derived variable are still accurate, only the fraction of assemblies where the derived variable will be within its tolerance specification will be inaccurate. If the situation requires an accurate measure of the fraction which are within tolerance, then we suggest that you perform a Monte Carlo tolerance stack up analysis.

A simple illustration of the Central Limit Theorem is to compare the probability density functions for a derived variable which is the sum of one, two, three or more contributors each of which has a uniform distribution to the probability density function of a Gaussian random variable with the same mean and standard deviation.

5.1 One Uniform Distribution

A distribution whose probability density function is a constant between the two limits a and b and is zero outside these limits is called a *uniform distribution*. Such a distribution has a mean $\mu = (b+a)/2$ and a standard deviation $\sigma = (b-a)/\sqrt{12}$. Figure 1 on the following page shows the probability density function of a derived variable with only one contributor which has a uniform distribution. This is compared to the probability density function of a Gaussian distribution with the same mean and standard deviation. The black box is the probability density function for the derived variable, and the red curve is the bell shaped Gaussian probability density function.

Notice that if the contributor has a uniform distribution over its tolerance specification limits, then the values $\mu \pm 3\sigma$ are considerably outside the limit stack for the derived variable.

5.2 Sum of Two Uniform Random Variables

A derived variable which is the sum of two contributors, each with a uniform random distribution between the two limits a and b has a triangular shaped probability density function. Such a distribution has a mean $\mu = (b+a)$ and a standard deviation $\sigma = (b-a)/\sqrt{6}$. Figure 2 on the next page compares the probability density function of this derived variable with that of a Gaussian distribution with the same mean and standard deviation. The black triangle is the probability density function for the derived variable which is sum of two uniform random contributors, and the red curve is the bell shaped Gaussian probability density function.

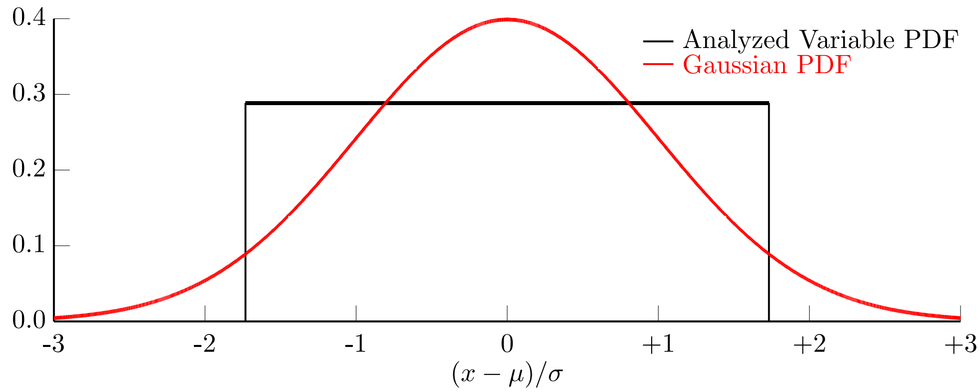

Figure 1: Derived Variable is Sum of One Uniform Contributor

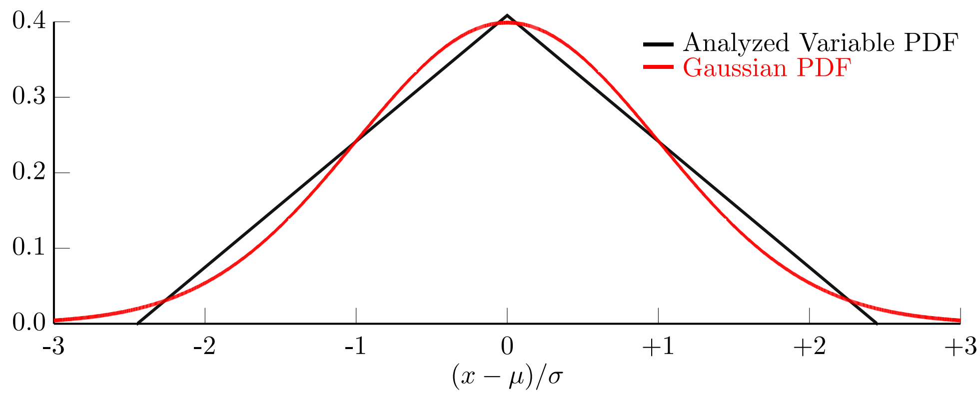

Figure 2: Derived Variable is Sum of Two Uniform Contributors

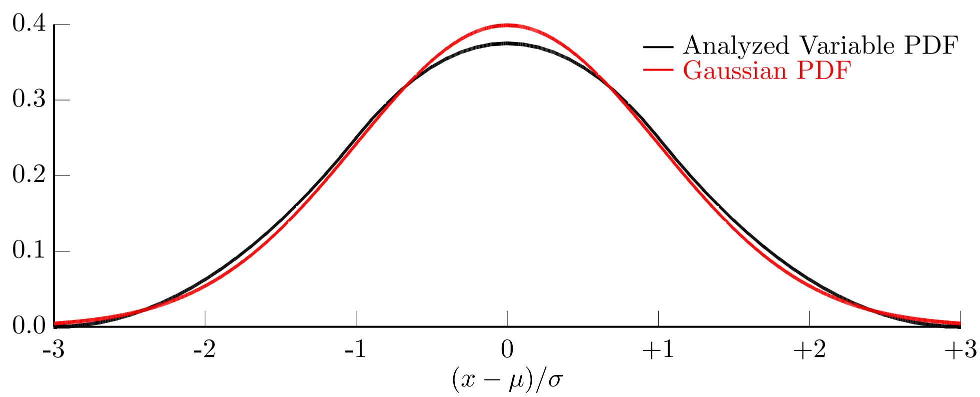

Figure 3: Derived Variable is Sum of Three Uniform Contributors

Notice that if the two contributors have uniform distributions over their tolerance specification limits, then the values $\mu \pm 3\sigma$ are outside the limit stack for the derived variable.

5.3 Sum of Three Uniform Random Variables

A derived variable which is the sum of three contributors each with a uniform random distribution between the two limits a and b has a bell shaped probability density function. Such a distribution has a mean $\mu = 3(b+a)/2$ and a standard deviation $\sigma = (b-a)/\sqrt{4}$. Figure 3 on the preceding page compares the probability density function of this derived variable with that of a Gaussian distribution with the same mean and standard deviation. The black curve is the bell shaped probability density function for the derived variable, and the red curve is the bell shaped Gaussian probability density function.

Notice that if the three contributors have uniform distributions over their tolerance specification limits, then the values $\mu \pm 3\sigma$ are equal to the limit stack for the derived variable.

5.4 More Than Three Uniform Random Variables

As more contributors are added to the sum, each with a uniform distribution, the probability density function for the derived variable quickly converges to that of a Gaussian random variable. If there are more than three contributors and they have uniform distributions over their tolerance specification limits, then the values $\mu \pm 3\sigma$ will be less than the limit stack for the derived variable.

5.5 Conclusions

A major contributor is one whose percent contribution is greater than half that of the contributor with the largest percent contribution. These major contributors help make the Central Limit Theorem converge to a Gaussian distribution. The remaining contributors contribute little to either the tolerance stackup or to the convergence to a Gaussian distribution.

If the major contributors are all Gaussian, then the analyzed variable will also have a Gaussian distribution no matter how many of them there are, especially if there are only one or two such contributors.

Without making any assumption about the probability distributions of the residuals, if there are three or more major contributors to the analyzed variable, we can safely assume that the analyzed variable has a Gaussian distribution. In particular, if there are three or more major contributors to the analyzed variable, we can assume that only about 0.27% of all assemblies will be outside the interval $\mu_d \pm 3\sigma_d$.

If there are only one or two major contributors which do not have a Gaussian distribution, then the Central Limit Theorem has not converged to the Gaussian distribution, and any results which rely on the assumption that the analyzed variable has a Gaussian distribution will be inaccurate. The only results which rely on the analyzed variable having a Gaussian distribution is the fraction of parts that will be in tolerance.

6 Tolerance Analysis in Enventive

In the Enventive system, the various GD&T tolerance specifications are represented by constraints (equations) whose residuals r_i are the contributors to the tolerance stack up. When all the residuals are equal to zero, we have the nominal geometry. Each tolerance specification implies limits on one or more of the residuals [6].

The tolerance analysis algorithm in Enventive performs the following steps:

1. The user picks one or more derived variables and asks for them to be analyzed. Usually these variables are the values of derived dimensions, but any of them can be variables determined by engineering equations.

2. The sensitivity coefficients $\left(\frac{\partial f}{\partial r_i}\right)$ used in equation (2) on page 1 are computed for each variable to be analyzed. These coefficients are computed using an Automatic Differentiation algorithm [1]. The function $f(\cdots)$ in equation (1) on page 1 is not explicitly formed.

3. The tolerance specifications are analyzed [6] and used to determine values for the mean μ_i and standard deviation σ_i of each of the residuals r_i. Since Enventive is a preliminary design tool, we do not ask the user for the value of each mean and standard deviation. Instead, we determine the mean and standard deviation by assuming that the process which controls each residual r_i is both centered and capable ($C_p = 1$) [3].

4. Equation (3) on page 2 is used to approximate the mean of each analyzed variable. Any error in this approximation is caused by the error in the linearized approximation, equation (2) on page 1.

5. Equation (4) on page 2 is used to approximate the variance of each analyzed variable. Any error in this approximation is caused by the error in the linearized approximation, equation (2) on page 1. Taking the square root of this variance produces the standard deviation of the analyzed variable.

6. The Central Limit Theorem is invoked to assume that each analyzed variable has a Gaussian distribution. This computed mean and standard deviation are used to approximate the probability that the analyzed variable satisfies its tolerance specification. If there are only one or two major contributors each of which do not have a Gaussian assumption, then this computed probability that the analyzed variable satisfies its tolerance specification will be highly inaccurate.

References

[1] Automatic differentiation. Wikipedia, 2007. This gives a good description of Automatic Differentation, but not of the specific algorithm used to compute sensitivity coefficients.

[2] Central limit theorem. Wikipedia, 2007. URL is http://en.wikipedia.org/wiki/Central_limit_theorem.

[3] Statistical process control. Wikipedia, 2007. URL is http://en.wikipedia.org/wiki/Statistical_process_control.

[4] Granino Arthur Korn and Theresa M. Korn. *Mathematical Handbook for Scientists and Engineers*. Dover, 2000. ISBN 0-486-41147-8.

[5] Ron Mittelhammer. *Mathematical Statistics for Economics and Business*. Springer, 1996. ISBN 0-387-94587-3.

[6] ANSI Committee Y14.5.1. *Mathematical Definition of Dimensioning and Tolerancing Principles*. ASME, 1994.

