SAND2020-3537C

Evaluating risk in an abnormal world: how arbitrary probability distributions affect false accept and reject evaluation

Speaker/Author: Dr. Collin J. Delker
Primary Standards Lab
Sandia National Laboratories
P.O. Box 5800, MS 0667, Albuquerque, New Mexico, 87185-0667 USA
Phone: 505-845-7431 Email: cjdelke@sandia.gov

Abstract

Risk mitigation strategies commonly use the Test Uncertainty Ratio (TUR) and End of Period Reliability (EOPR) to ensure a measurement is adequate for making acceptance decisions. In some cases, TUR and EOPR are used to determine an appropriate guardband factor, which is then used to reduce the risk of an incorrect decision. Unfortunately, the common guidance of maintaining a TUR of at least 4:1 was developed to simplify the underlying calculus in an era predating modern computing. As such, using a TUR to determine the adequacy of a measurement and set guardband limits assumes that the probability distributions describing the product and the measurement uncertainty are unbiased and normally distributed. The Guide to the Expression of Uncertainty in Measurement (GUM) and its supplements describe several situations where uncertainty in the measurement will not follow a normal distribution. Additionally, it is frequently assumed that the prior distribution of products being measured is normal and unbiased, often with little or no evidence. Sometimes a 95% EOPR is even assigned with no justification. Despite the evidence of non-normal behavior in measurements and products, risk evaluations typically assume normality in both distributions. While evaluating the Probability of False Accept (PFA) and the Probability of False Reject (PFR) is more challenging when the probability distributions are non-normal, the calculus is straightforward using either numerical integration or Monte Carlo techniques. In addition to covering methods for evaluating the actual PFA and PFR without relying on the archaic TUR metric, this work considers several case studies of risk evaluation, including both global and specific risk, when the product or the test measurement uncertainty do not follow normal distributions. Neglecting non-normal behavior can greatly affect PFA and PFR by either over- or underestimating the probabilities depending on the parameters of the distributions. A good prior knowledge of the product being measured is required for a meaningful global risk analysis.

1. Introduction

Measurements are used to make decisions, whether for acceptance or rejection of a product, or for a pass or fail determination on a calibration certificate. Because all measurements have some inherent uncertainty, there is always some probability of making an incorrect pass or fail

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

determination. Risk is calculated from a combination of test measurement uncertainty and the known distribution of products being tested [1], [2].

The Guide to the Expression of Uncertainty in Measurements (GUM) [3] recommends ways to evaluate the test measurement uncertainty distribution. It is limited to cases where the measurement equation can be modeled by a linear expansion and it converts all the input uncertainties into normal distributions. These simplifications result in a statement of total uncertainty that is always described by a normal distribution. As GUM Supplement 1 [4] describes in detail, a Monte Carlo (MC) method can be used to propagate uncertainty without these limitations. An MC uncertainty calculation results in a histogram representation of the uncertainty distribution. Depending on the measurement equation and inputs, this histogram may be highly non-normal.

The product distribution requires prior knowledge of the products being measured. While this information is sometimes scarce, it is essential for an accurate determination of risk. Non-normal distributions in the product can arise from physical limitations, non-normally distributed components used in manufacturing, or other factors in the process. For example, if a normal distribution is used to model the prior knowledge of a component's mass, it would predict some small but finite probability of the component having negative mass, which is clearly impossible. A bounded distribution, such as a gamma distribution, may be more appropriate in such cases.

Traditional risk analysis, which calculates the Probability of False Accept (PFA) and the Probability of False Reject (PFR), typically considers the probability distributions involved to be normal. The Test Uncertainty Ratio (TUR) is often used as a simple metric for understanding and controlling risk, but it is also based on normal distributions and in fact was introduced to simplify calculations before modern computers could readily solve integration problems [5].

The assumption of normality is clearly not always justified. Compliance with ANSI/NCSL Z504.3's requirement to maintain less than 2% PFA [6] cannot be ensured if non-normal behavior is neglected. Risk in non-normal distributions was considered in [7] by using a series approximation for the distribution, however its accuracy depends on the suitability of the series approximation, and due to computational limitations of the time it did not consider asymmetric or biased distributions. NCSLI's Recommended Practice 18 [8] sets up the calculus for risk analysis using any probability distribution yet makes a normal assumption for most further analysis. This work seeks to understand the potential effects of non-normal behavior – including asymmetry, bias, and "peakedness" – in distributions used for risk analysis and to offer solutions accounting for these effects where the only approximation is due to numerical integration techniques.

2. Calculating Risk

In terms of metrology, risk is the probability of making an incorrect decision based on a measurement result. Two types of risk are considered: specific risk and global risk.

Specific risk is the probability that a *given* test measurement result indicates a product is within specifications when it is actually outside specifications. This risk is due to the inherent uncertainty in the test measurement. It can be calculated by integrating the Probability Density Function (PDF) of the test measurement uncertainty outside of the specification limits.

$$p_{FA}|_y = \int_{-\infty}^{LL} p_{test}(t-y)\,dt + \int_{UL}^{\infty} p_{test}(t-y)\,dt$$

Here, $p_{FA}|_y$ is the PFA *given* the measurement result y. The function $p_{test}(t-y)$ is the PDF of the test measurement with measured value y. The upper and lower specification limits are given by UL and LL, respectively. The specific risk is illustrated in Figure 1, where the upper tail of the probability curve falls above the UL specification limit. The area under this curve represents the probability (specific risk) of incorrectly accepting a product given this measurement result.

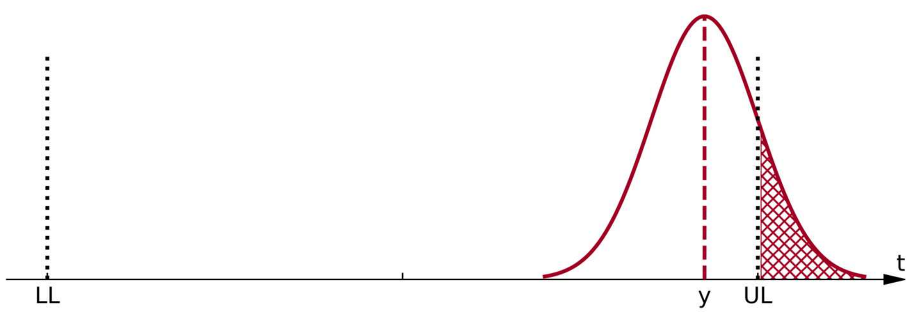

Figure 1. Specific risk given a measurement result of *y*.

Global risk is the probability that a test measurement on *any* product results in an incorrect pass determination. It is also called the PFA or *consumer's risk.* In addition to the test measurement probability distribution, it must account for the probability of encountering a product at that measured value. It is computed by integrating the joint PDF of the test measurement, $p_{test}(t-y)$ and the PDF of the unit under test $p_{uut}(t)$. Global risk cannot be calculated without specifying a distribution of values for the product, which is obtained from prior knowledge and measurements of the product.

$$PFA = \int_{LL}^{UL} \int_{-\infty}^{LL} p_{test}(t-y)p_{uut}(t)\, dt\, dy + \int_{LL}^{UL} \int_{UL}^{\infty} p_{test}(t-y)p_{uut}(t)\, dt\, dy$$

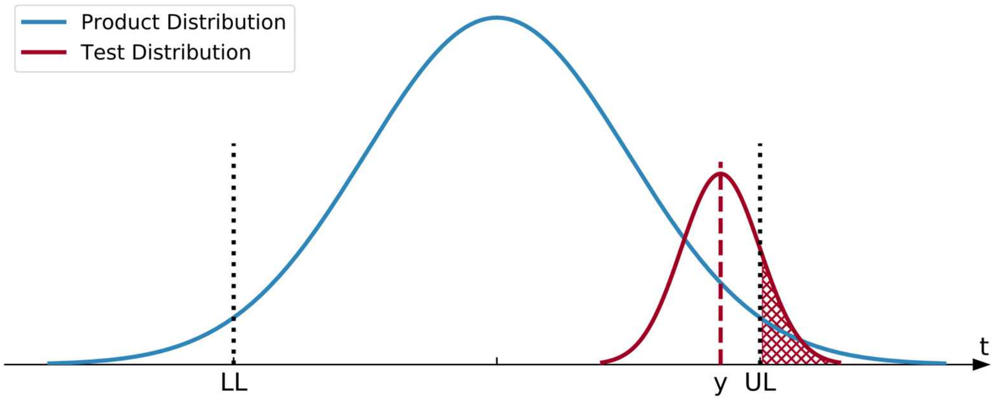

Figure 2. Global risk (PFA) combines the specific risk with the probability of encountering a product with value *y*.

The joint probability accounts for the assumption that it is less likely to encounter a product near the specification limit than a product near the center of the limits.

Similarly, the PFR, also called *producer's risk*, is the probability that a measurement result indicates a product is outside the limits when its true value is actually within the limits.

$$PFR = \int_{-\infty}^{LL} \int_{LL}^{UL} p_{test}(t-y)p_{uut}(t)\, dt\, dy + \int_{UL}^{\infty} \int_{LL}^{UL} p_{test}(t-y)p_{uut}(t)\, dt\, dy$$

Determining $p_{uut}(t)$ can be challenging without sufficient data about the product being measured. Frequently, the End of Period Reliability (EOPR), also called in-tolerance probability (ITP), is used to estimate the standard deviation of $p_{uut}(t)$ [8]; however, this approach also assumes a normal and unbiased distribution of products. When bias is present, the mean of the product distribution $p_{uut}(t)$ does not fall halfway between LL and UL.

Previous studies analyzed these risk integrals under the assumption that the probability distributions for both the measurement process and the product were normal. While this may be true in many cases, the GUM and its supplements provide several examples of uncertainty distributions not resulting in normality. For example, the test measurement distribution can be non-normal if significant nonlinearities exist in the measurement model or if the distributions of the input variables are non-normal. The product distribution may not be normal if the product is manufactured from several non-normal components or if there is other bias or nonlinearity in the manufacturing process or stability of the product.

3. Computing Risk with Non-Normal Distributions

Except for a few simple PDFs for $p_{uut}(t)$ and $p_{test}(t-y)$, the PFA and PFR equations do not have closed-form solutions. However, it is not difficult to evaluate the integrals numerically. The simplest approach is to define a sufficient range of t, compute the PDFs over this range of t, and use trapezoidal integration approximation. Many statistical computing libraries have built-in numerical integration functions that take the guesswork out of determining the appropriate range of t, step sizes, etc.

A second approach requires no calculus. If random samples can be generated from the two probability distributions, an MC method can be used. First, draw a random sample from the product distribution $p_{uut}(t)$. Then use that sample as the y value for drawing a random sample from $p_{test}(t-y)$. After generating a few million samples, finding PFA and PFR simply becomes a matter of counting the number of samples that fall in the false accept or false reject region compared to the total number of samples drawn. Figure 3 illustrates an MC risk calculation on a measurement where both PDFs are normal, with a TUR of 4, $LL = -1$, and $UL = 1$.

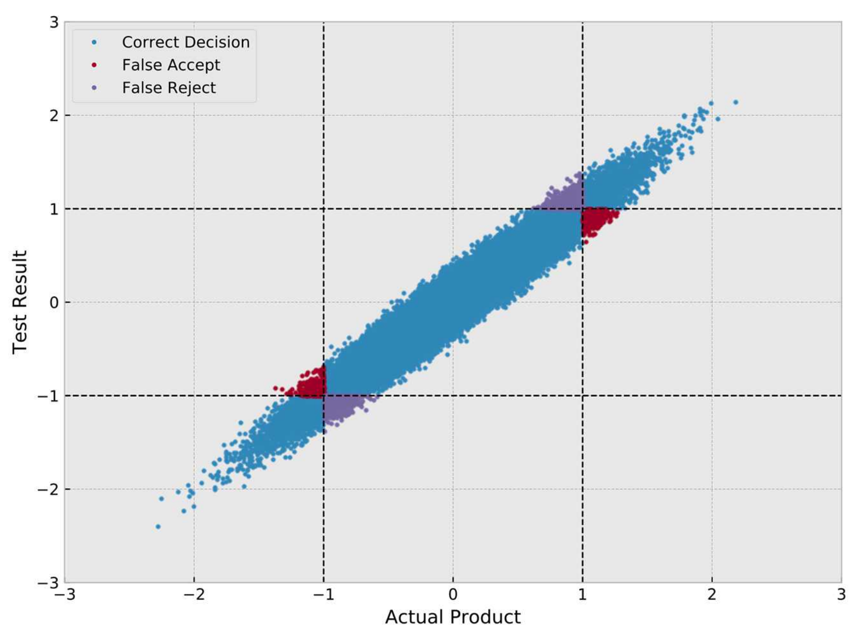

Figure 3. Monte Carlo risk evaluation.

Both approaches lend themselves to computing risk based on non-normal distributions. With numerical integration, all that is required is a PDF for each distribution. The MC method requires a random number generator that can draw samples from each distribution, already available in most statistical software for many common distribution types or even histogram-based distributions.

An uncertainty calculation using MC methods produces a histogram of the measurement's possible true values. This histogram needs to be converted into a PDF to define either $p_{test}(t-y)$ or $p_{uut}(t)$. The PDF can be approximated by either a piecewise function connecting the top of each histogram bar, or finding the best fit of the histogram to a known distribution, such as a skew-normal distribution. To avoid the complexities and potential errors arising from a poor choice of distribution to fit, piecewise PDF approximations are used for the analysis here. The example risk calculations and plots below were made using open-source software developed by the Primary Standards Lab at Sandia National Laboratories [9] to compute risk and other uncertainty-related calculations.

4. Case Studies

To illustrate the effects of non-normal distributions in risk analysis, three case studies are provided:

- Measurement of a step height standard with non-normal measurement uncertainty due to nonlinearity in the measurement model
- Measurement of the step height standard with non-normal inputs to the measurement model
- Measurement of the time constant of a Resistor-Capacitor (RC) circuit with uniformly distributed components

Non-normal test distribution due to nonlinearity: step height standard

A step-height check standard is calibrated using a ratio method by comparison against a calibrated reference. The measurement equation is

$$Y_{corr} = \frac{X_{cal}}{X_{meas}} Y_{meas}$$

Where Y_{corr} is the corrected measurement of the check standard, Y_{meas} is the uncorrected measurement, X_{meas} is a measurement on a reference standard using the same measurement equipment, and X_{cal} is the calibrated value of the reference standard. With values and (k = 1) uncertainties of Y_{meas}= 698 nm ± 20 nm, X_{meas}= 180 nm ± 20 nm, and X_{cal}= 182 nm ± 5 nm, the uncertainty in Y_{corr} can be found using the GUM and MC (GUM-S1) methods. The probability density for Y_{corr} is shown in Figure 4.

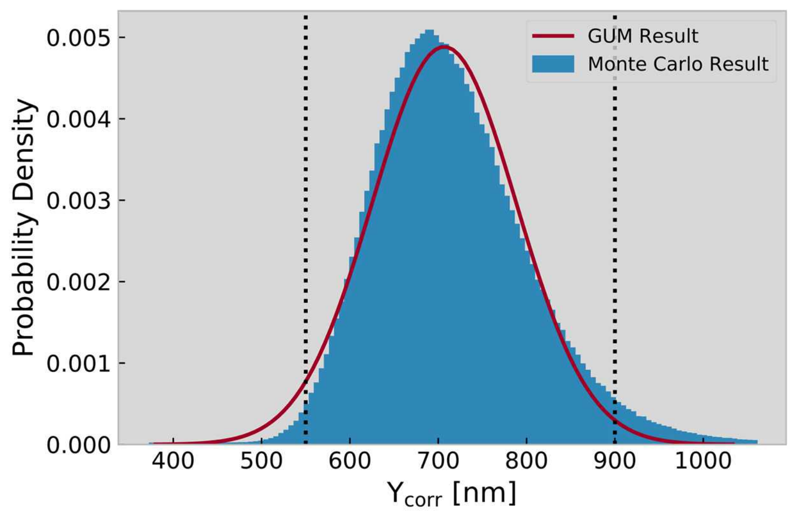

Figure 4. Uncertainty evaluation of step height standard.

Note the slight difference between the GUM and MC results. Using the GUM method, Y_{corr} = 706 nm ± 82 nm (k = 1) while the MC method results in an asymmetric 95% coverage range of 571 nm to 913 nm. In an uncertainty evaluation, the difference is frequently neglected, and a standard deviation is taken of the MC samples resulting in an uncertainty with a normal distribution.

Next, consider using these distributions in a risk analysis. Suppose there is a requirement that the step height be less than 900 nm. It should be apparent from the plot that the GUM's normal approximation and MC's more accurate histogram-based PDF will result in different specific risk values.

Table 1 shows a significant difference in specific risk probability when each measurement result above is used to check whether the standard is less than 900 nm or, alternatively, greater than 550 nm.

Table 1. Specific risk of step height standard calculated using a normal PDF from the GUM approximation and a histogram PDF from MC uncertainty analysis.

	Specific Risk Calculated Using:	
Limit	Normal PDF (GUM)	Histogram PDF (MC)
< 900 nm	0.90%	2.97%
> 550 nm	2.75%	0.80%

Now consider global risk. Assume the standard has lower and upper specification limits of 450 nm and 1050 nm, respectively. Global risk requires knowledge of a product distribution: here, assume a normal product distribution with standard deviation of 200 nm with no bias (mean of 750 nm).

Because bias in the distributions has a large effect on risk [10], a negatively biased product distribution (mean of 500 nm – i.e. not centered within the specification limits) is also considered. Table 2 shows the global false accept risk calculated using both GUM and MC results.

Table 2. Global risk of step height standard calculated using a normal PDF from the GUM approximation and a histogram PDF from MC uncertainty analysis.

	Global Risk (PFA) Calculated Using:	
Bias	Normal PDF (GUM)	Histogram PDF (MC)
None	2.81%	2.82%
Negative	5.54%	6.18%

There is only a modest difference in PFA for the non-biased case but a larger variation in PFA for the biased case. Without bias, the reduction in PFA due to false accept *below* the lower limit is made up for with additional PFA *above* the upper limit, resulting in the insignificant difference due to the numerical nature of the histogram and integration method. When the product distribution is biased, this tradeoff is not symmetrical, leading to a difference in total PFA.

Non-normal test distribution due to uniformly distributed inputs: step height standard

Next, consider the same step-height case study, but with the input uncertainties for each variable as uniform distributions. Uniform distributions are commonly used to specify uncertainty due to manufacturer's specifications. Both GUM and MC uncertainty calculations are shown in Figure 5, and result in approximately the same standard uncertainty (and therefore same TUR value) as a traditional risk assessment, but the two results have different PDFs.

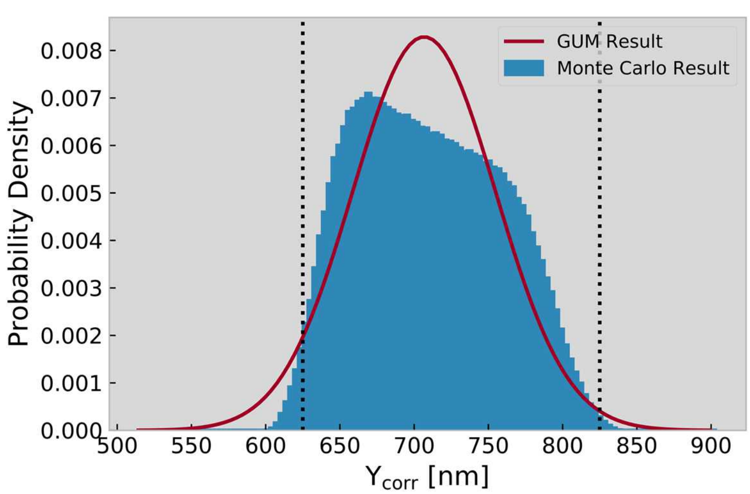

Figure 5. Uncertainty evaluation of step height standard with uniform inputs.

Choosing limits of less than 825 nm or greater than 625 nm, the specific risk for this measurement result is shown in Table 3. Again, there is major discrepancy in risk values depending on the choice of distribution. Applying the GUM approach produces an error of about 3% PFA at the lower limit.

Table 3. Specific risk of step height standard calculated using a normal PDF from the GUM approximation and a histogram PDF from Monte Carlo uncertainty analysis.

	Specific Risk Calculated Using:	
Limit	Normal PDF (GUM)	Histogram PDF (MC)
< 825 nm	0.70%	0.14%
> 625 nm	4.48%	1.54%

Using the specification limits (450 nm, 1050 nm) and product distributions with and without bias from the previous example, the global risk can also be calculated, as shown in Table 4. In this case, with no bias, the assumption of a normal distribution would pass Z540.3's 2% PFA requirement but the more accurate MC result would fail by a slight amount. Both methods fail Z540.3's 2% PFA if the negative bias in the product identified from past experience is included in the global risk computations.

Table 4. Global risk of step height standard calculated using a normal PDF from the GUM approximation and a histogram PDF from MC analysis.

	Global risk (PFA) calculated Using	
Bias	Normal PDF (GUM)	Histogram PDF (MC)
None	1.90%	2.06%
Negative	3.24%	3.70%

Non-normal product distribution: RC Circuit

The previous two examples considered non-normality in test measurement uncertainty. This example considers the effects of a non-normal product distribution.

An RC circuit is constructed from off-the-shelf components. The components are specified with tolerances, assumed to be uniform distributions for use in an uncertainty analysis. The time constant of the circuit is given by

$$\tau = RC.$$

Using $R = 32$ kΩ ± 5% and $C = 1$ μF ± 10%, the GUM and MC uncertainty evaluations lead to the probability distributions shown in Figure 6, both methods resulting in $\tau = 32.0$ ms ± 2.1 ms. While the same standard uncertainty results from both methods, the probability distributions differ.

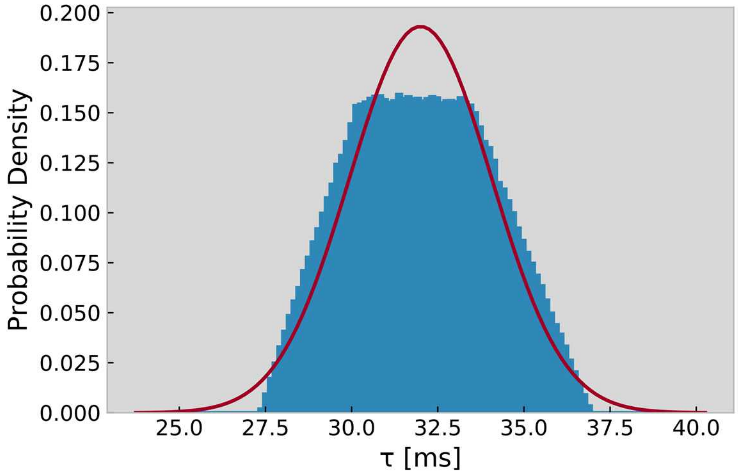

Figure 6. Uncertainty evaluation of RC circuit time constant.

Consider lower and upper specification limits of 28.5 ms and 35.5 ms. Suppose that each RC circuit is measured using an oscilloscope with 2% uncertainty in the time measurement. The global false accept risk resulting from both GUM and MC uncertainty evaluations can be computed and is shown in Table 5. Using a normal approximation from the GUM method passes Z540.3's 2% PFA requirement, but the more appropriate MC result fails slightly.

Table 5. Global risk of RC circuit calculated using a normal PDF from the GUM approximation and a histogram PDF from MC uncertainty analysis.

	Global Risk Calculated Using:	
Risk	Normal PDF (GUM)	Histogram PDF (MC)
False Accept	1.59%	2.10%
False Reject	2.99%	3.93%

5. Generalization

The previous examples demonstrate how risk can be affected by non-normal behavior and assumptions about the product. Unfortunately, it is difficult to know beforehand how great the effect will be without performing a full risk calculation using both GUM and MC results. To generalize the effects, two probability distribution parameters are considered: skewness and kurtosis.

Skewness is a measure of symmetry in a probability distribution [11]. A normal distribution is symmetric and thus has skewness of zero. Figure 7 (left) illustrates skewed distributions with positive, negative, and zero skew while maintaining a constant standard deviation and median.

Kurtosis is a statistic for comparing the strength of a distribution's tails to its peak [11]. By definition, a normal distribution has kurtosis of 3. Frequently, "excess kurtosis" is used by subtracting 3 so that a normal distribution has excess kurtosis of 0. Then distributions with positive excess kurtosis have stronger tails than a normal distribution, and distributions with negative excess kurtosis have weaker tails. Figure 7 (right) illustrates distributions with varying kurtosis that maintain a constant median and standard deviation.

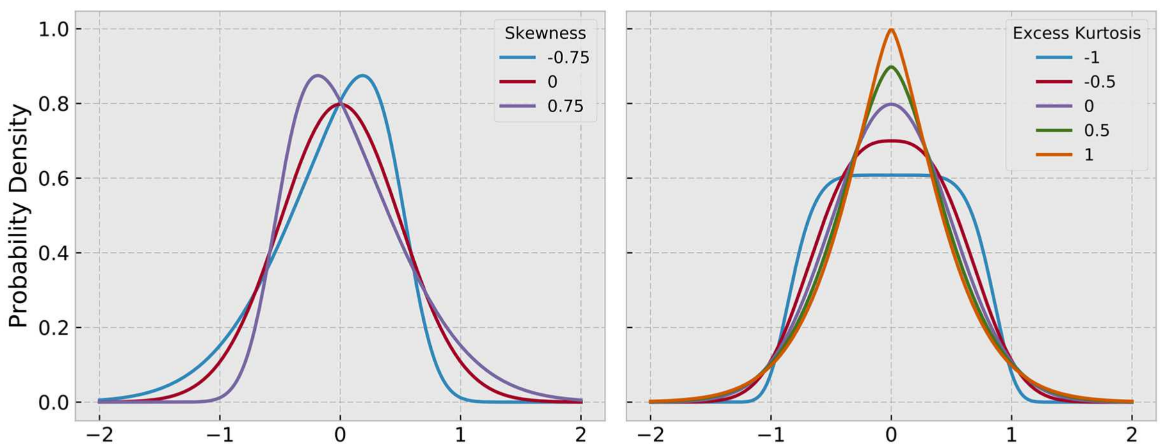

Figure 7. Distributions with varying skewness and kurtosis that maintain constant standard deviation and median.

Specific risk with skew and kurtosis in test measurement

First, consider specific risk with various skew parameters. In Figure 8, the x-axis indicates a measured value normalized in terms of the test measurement's standard deviations, such that a value of 0 indicates a measurement at the specification limit, and a value of -1 indicates a measurement 1 standard deviation below the limit. To increase the clarity of these figures, the lower limit is not shown (consider it to be many standard deviations below the 0 limit). A measured value at the limit results in the familiar 50% probability of an incorrect decision given this measurement result. Values of skewness greater than zero show higher specific risk compared to zero skewness, but values less than zero decrease the specific risk. The effect would be flipped at the lower specification limit, with positive skewness decreasing specific risk. The change in risk due to excess kurtosis depends on how close the measured value is to the specification limit. Because of its symmetry, kurtosis effects would be identical at the upper and lower limits.

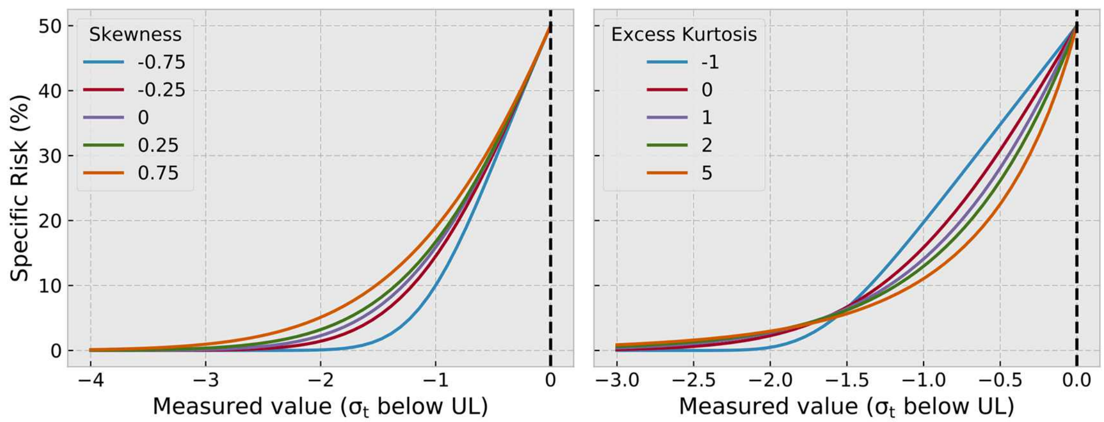

Figure 8. Specific risk with various skewness and kurtosis parameters.

Global Risk with skew and kurtosis in test measurement (unbiased product)

Global risk evaluated as a function of skew and kurtosis in the test measurement is shown in Figure 9, assuming a normal and unbiased product distribution with itp of 75%. The skew parameter has no effect on global risk because any additional risk of false decision below the limit is offset by reduced risk of false decision above the limit (or vice versa). Kurtosis affects global risk especially at a lower TUR due to changes in the strength of the tails compared to the center of the test measurement uncertainty.

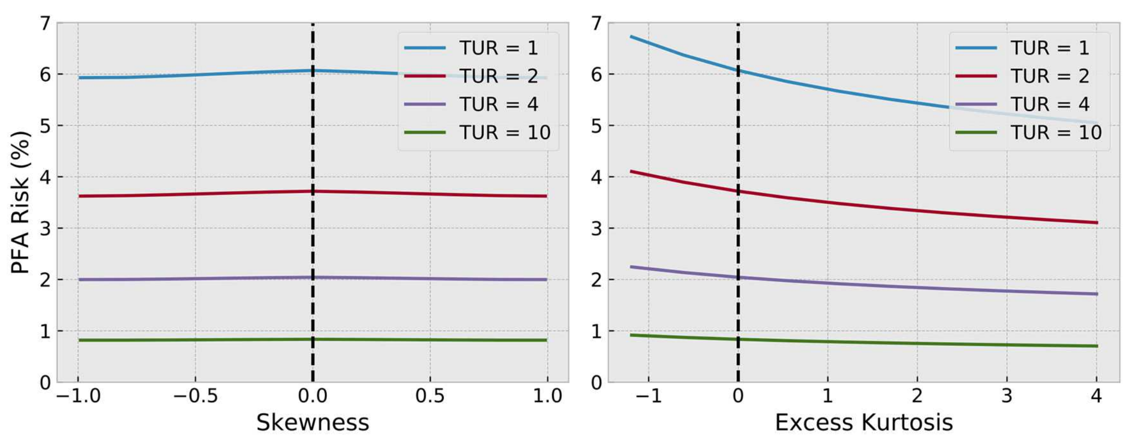

Figure 9. Global PFA risk as a function of skewness and kurtosis in the test measurement distribution[1]. The product distribution is normal and unbiased.

[1] The slight waviness in the skewness plots of Figure 9 and Figure 10 near skewness of zero, especially noticeable at lower TURs, stems from a small finite probability of values entering the upper left or lower right quadrant of the MC risk evaluation plot (Figure 3). While technically a correct pass/fail decision is made in these quadrants, the decision is made for the wrong reason. Having probability in these regions reduces the calculated PFA by an amount that varies with skewness.

Global risk with skew and kurtosis in test measurement (biased product)

When a bias is present in the product distribution (Figure 10), there is asymmetry, and the change in PFA below the limit is no longer equally offset by change in PFA above the limit.

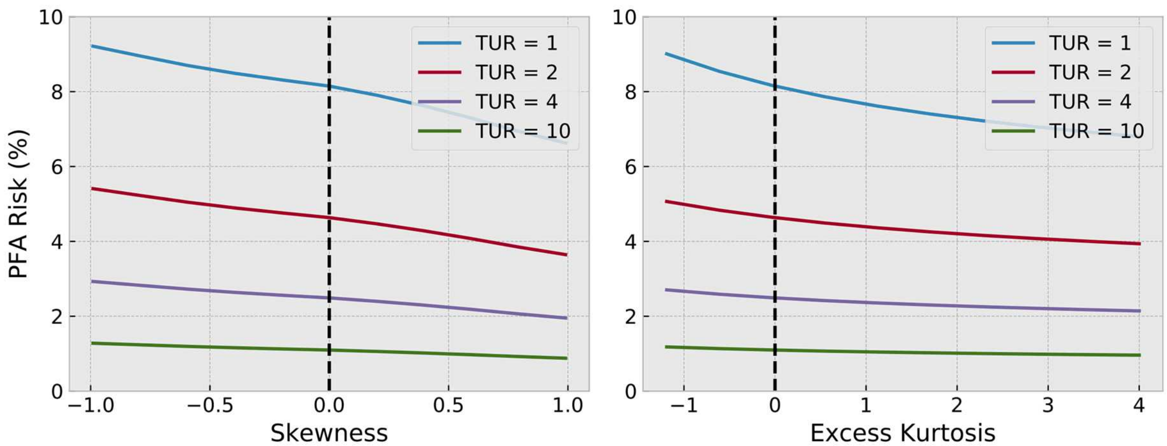

Figure 10. Global PFA risk as a function of skewness and kurtosis in the test measurement distribution, with a biased normal product distribution.

While the trends in these plots are specific to the parameters of the distributions used to calculate them, they can be generalized to other distributions. Skewness in the test measurement has little effect on global PFA, assuming the product distribution is unbiased and normal. Kurtosis has more influence, but as TUR is increased, the effects of both skewness and kurtosis are diminished. Skewness and kurtosis are just two parameters used to quantify non-normality; the full characteristics of the actual distribution should always be considered in a complete risk evaluation.

6. Guardbanding

When an unacceptable level of risk is determined for a measurement process, guardbanding can mitigate the risk by adjusting the upper and lower specification limits. Typically, guardbanding has been based on the root-sum-square (RSS) method or other calculations involving the TUR [12]. Of course, guardbanding based on TUR again assumes normal distributions are defining the risk calculation. An alternative is to use the equation for global PFA with guardbanding:

$$PFA_{GB} = \int_{LL+GB}^{UL-GB} \int_{-\infty}^{LL} p_{test}(t-y) p_{uut}(t)\, dt\, dy + \int_{LL+GB}^{UL-GB} \int_{UL}^{\infty} p_{test}(t-y) p_{uut}(t)\, dt\, dy$$

and use numerical minimization to solve for the value of the guardband GB that produces an acceptable PFA. While not nearly as easy as an arithmetical calculation on TUR, this method is guaranteed to set the global PFA to an acceptable level. In theory, the equation could be further complicated by applying a different guardband factor to the upper and lower limits.

If guardbanding is calculated based on the GUM's normal approximation and a TUR value using the RSS method, the actual risk may be higher than anticipated. Consider the RC circuit measured previously. The PFA and PFR determined using the normal assumption with different guardbanding methods was calculated and compared to the actual PFA and PFR determined using the full histogram PDF. While the differences in percentages may seem small, even a few tenths of a percent increase in actual false accept rate can be significant when dealing with large numbers of products or high costs and consequences of false acceptance.

Table 6. RC circuit risk with guardbanding.

Guardband	PFA calculated using normal assumption	Actual PFA	PFR calculated using normal assumption	Actual PFR
None	1.59%	2.10%	2.99%	3.93%
RSS Method (TUR)	1.01%	1.35%	4.66%	6.05%
Target 0.8%	0.80%	1.07%	5.60%	7.25%

Where prior knowledge of the product is scarce, or where a manufacturing process is not in control and varies greatly over time, it may be beneficial to calculate a guardband based only on specific risk. Such a method, based on expected costs of false accept and false reject conditions, is described in Reference [13]. While this method is likely to produce a more conservative guardband limit, it makes no unjustified assumptions about the product.

7. Conclusions

Many uncertainty calculations lead to non-normal probability distributions. If non-normal behavior is not accounted for in a risk evaluation, the false accept and false reject probabilities may be under- or overestimated. It is recommended to evaluate uncertainty using both the GUM methodology as well as an MC analysis and compare the results to check for non-normal behavior in the test measurement. An accurate understanding of the product's prior distribution is also essential to accurately determine global risk. The TUR as a risk metric assumes normal distributions and may not always be appropriate; maintaining a TUR of 4 is not sufficient to guarantee PFA risk of less than 2% if non-normal behavior or bias is present in the underlying distributions. With modern computing power, a full risk evaluation including arbitrary probability distributions is not difficult and should be used when there is sufficient data to characterize the test measurement and product distributions.

8. References

[1] JCGM 106:2012, "Evaluation of measurement data - The role of measurement uncertainty in conformity assessment," BIPM, 2012.

[2] A. R. Eagle, "A Method for Handling Errors in Testing and Measuring," *Industrial Quality Control,* vol. 10, no. 3, pp. 10-15, 1954.

[3] JCGM 100:2008, "Guide to the Expression of Uncertainty in Measurement," BIPM, 2008.

[4] JCGM 101:2008, "Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurment" - Propagation of distributions using a Monte Carlo method," BIPM, 2008.

[5] J. Harben and P. Reese, "Implementing strategies for risk mitigation in the modern calibration laboratory," in *NCSLI Workshop and Symposium*, 2011.

[6] ANSI/NCSL Z540.3-2006, "Requirements for the calibration of measuring and test equipment," ANSI/NCSL, 2006.

[7] H. R. Singh, "Producer and Consumer Risks in Non-Normal Populations," *American Society for Quality,* vol. 8, no. 2, pp. 335-343, 1966.

[8] NCSL International, "Recommended Practice 18: Estimation and evaluation of measurement decision risk," 2014.

[9] S. M. Mimbs, "Using reliability to meet Z540.3's 2% rule," in *NCSLI Workshop and Symposium*, 2011.

[10] Primary Standards Lab, Sandia National Laboratories, "Sandia Uncertainty Calculator," [Online]. Available: https://sandiapsl.github.io.

[11] D. Deaver, "Managing calibration confidence in the real world," in *NCSLI Workshop and Symposium*, 1995.

[12] R. Shanmugam and R. Chattamvelli, Statistics for Scientists and Engineers, John Wiley & Sons, 2015.

[13] D. Deaver, "Guardbanding with confidence," in *NCSLI Workshop and Symposium*, 1994.

[14] R. G. Easterling, M. E. Johnson, T. R. Bement and C. J. Nachtsheim, "Statistical Tolerancing Based on Consumer's Risk Considerations," *Journal of Quality Technology,* vol. 23, no. 1, pp. 1-11, 1991.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.