RERTR 2019 – 40TH INTERNATIONAL MEETING ON
REDUCED ENRICHMENT FOR RESEARCH AND TEST REACTORS

OCTOBER 6-9, 2019
THE WESTIN ZAGREB
ZAGREB, CROATIA

Non-parametric Statistical Safety Analysis Tools to Support ATR Conversion to LEU

B. Hallee
MPR Associates, Inc.
320 King St., Ste. 400, Alexandria, VA 22314 – U.S.A.

ABSTRACT

To better support the Advanced Test Reactor (ATR) Low-enriched Uranium (LEU) conversion, a new ATR design basis accident safety analysis approach has been proposed which utilizes the Best-Estimate Plus Uncertainty (BEPU) code SASQUATCH. SASQUATCH employs a non-parametric statistical method which allows analysts to calculate a pre-defined number of runs required to reach a specified probability and confidence threshold for a distributionless output, and grants the ability to control the sampling error using order statistics. This presentation discusses the modernization of the ATR design-basis tools and evaluates the efficacy of the non-parametric statistical approach proposed for the LEU safety basis relative to the current parametric approach outlined in the ATR Safety Analysis Report (SAR).

1 Introduction

Best estimate codes are used to provide realistic versus conservative predictions of reactor response to initiating events. In using best-estimate approaches, uncertainty in code predictions must still be accounted for. Best-estimate plus uncertainty (BEPU) approaches are used to account for several sources of uncertainty, such as code models as well as uncertainties of plant and fuel parameters.

Historically, the Advanced Test Reactor (ATR) has used the SINDA-SAMPLE code, a 1-D thermal/hydraulic (T/H) model of an ATR fuel plate, in evaluating T/H safety margins in the ATR Safety Analysis Report (SAR) (Reference 1). As part of the safety analysis for incorporating U-10Mo low-enriched uranium (LEU) fuel into the ATR, new statistical methods are being developed for safety analysis. To support the implementation of these new methods, the Stochastic Analysis with SINDA for Quantification of Uncertainties in ATR Thermal Core Hydraulics (SASQUATCH) code is being developed at Idaho National Laboratory (INL) which includes a 3-D T/H model solver (Reference 2).

The legacy SINDA-SAMPLE code utilizes a parametric (Gaussian) statistical method used to quantify uncertainties in code output. The current safety basis requires SINDA-SAMPLE to be executed 1,200 times to generate the Gaussian response functions of output margins with a high degree of fidelity. The SASQUATCH code has been developed to include a more generalized, non-parametric statistical approach in order to justify fewer code runs, due to the increased computational cost of the 3-D solver, as well as to allow the output figure of merits (FOMs) to take any distribution form.

Many BEPU approaches developed by the nuclear industry and accepted by the Nuclear Regulatory Commission (NRC) rely on the propagation of input uncertainties and make use of the Wilks' or other ordered statistical method to determine the number of code calculations required to satisfy an established tolerance level such as 95% probability with 95% confidence (95/95) (References 3, 4). Accordingly, the calculated FOM that is compared with the corresponding acceptance criterion is often an upper or lower tolerance limit instead of the probability distribution of the output population.

2 Introduction

The one-sided Wilks' non-parametric formula is used to predict the number of samples needed to bound a percentile of the output distribution (α) with a desired confidence (β). The user specifies the desired order (p) of the Wilks' equation, as specified in Equation (1) below (Reference 5):

$$\sum_{\mathrm{j}=0}^{N-p} \frac{N!}{(N-\mathrm{j})!\,\mathrm{j}!} \alpha^{\mathrm{j}} (1-\alpha)^{\mathrm{N-j}} \geq \beta \qquad (1)$$

where:

α = The percentile of the output distribution
β = The confidence level of the output distribution
N = The minimum number of code runs
p = The order of the output population

The order states the number of code runs expected to lie in the "tail" of the distribution (i.e. beyond the α percentile value chosen by the user), and therefore the p^{th} largest, or smallest, output value is chosen as the limiting value. This is intended to reduce the sampling error inherent in first-order Wilks predictions (i.e., where $p = 1$).

The purpose of this work is to evaluate the predicted amount of safety margin and repeatability of the results between the Wilks non-parametric and parametric approaches, with the goal of understanding the impact of changing statistical approaches on the safety basis of the ATR.

3 Description of Approach

A challenge with comparing these two approaches is that the parametric approach does not rely on making a statement about confidence and makes only an implicit statement about the probability level (i.e. 3σ represents the 99.7^{th} percentile of a normal distribution but represents different probability levels for non-normal distributions). In order to make an equal comparison between these two approaches, the following methodology is used (using the 95/95 level as an example):

Parameric Approach

1. Generate *N* samples from a standard normal distribution
2. Using the *N* samples, determine the sample variance as an estimate for the population variance
3. Determine the uncertainty in the sample variance based on the number of samples and the fact that the output distribution is normal (for a normal distribution, an analytical formula is available to estimate this uncertainty)
4. Determine an estimate of the 95^{th} percentile value of the distribution and its associated uncertainty to determine a 95/95 value.

The above method simulates the current parametric approach used in the ATR SAR for the simple case of a single, normally-distributed variable, and extends it by additionally calculating the 95% confidence level of the output.

Non-parametric approach

1. Determine number of samples (*N*) required to estimate 95/95 value using the Wilks' method.
2. Run 10,000 *N*-sample runs to provide 10,000 estimates of the 95/95 value
    a. While for the parametric approach an analytical solution is available to estimate the uncertainty in the 95^{th} percentile value (since it's based on the standard deviation), the non-parametric approach has no equivalent analytical solution so an outer loop of sampling is performed to determine the spread of 95/95 estimates provided.
3. Determine the expected value of the 95/95 from 10,000 trials

Once both of these evaluations have been performed, their results can be compared to determine the difference in the amount of safety margin that is predicted by each and the repeatability of the results. As shown in Fig. 1 below, using the parametric approach (left figure) the true value of the 95^{th} percentile is estimated, and repeating this multiple times will result in a spread of 95^{th} percentile estimates around the "true" value. The non-parametric results are shown to the right of Fig. 1, where a bound on the 95^{th} percentile value is estimated with 95% confidence. Therefore, as shown, when repeated many times, the non-parametric method will tend to over-estimate the 95^{th} percentile value relative to the "true" value.

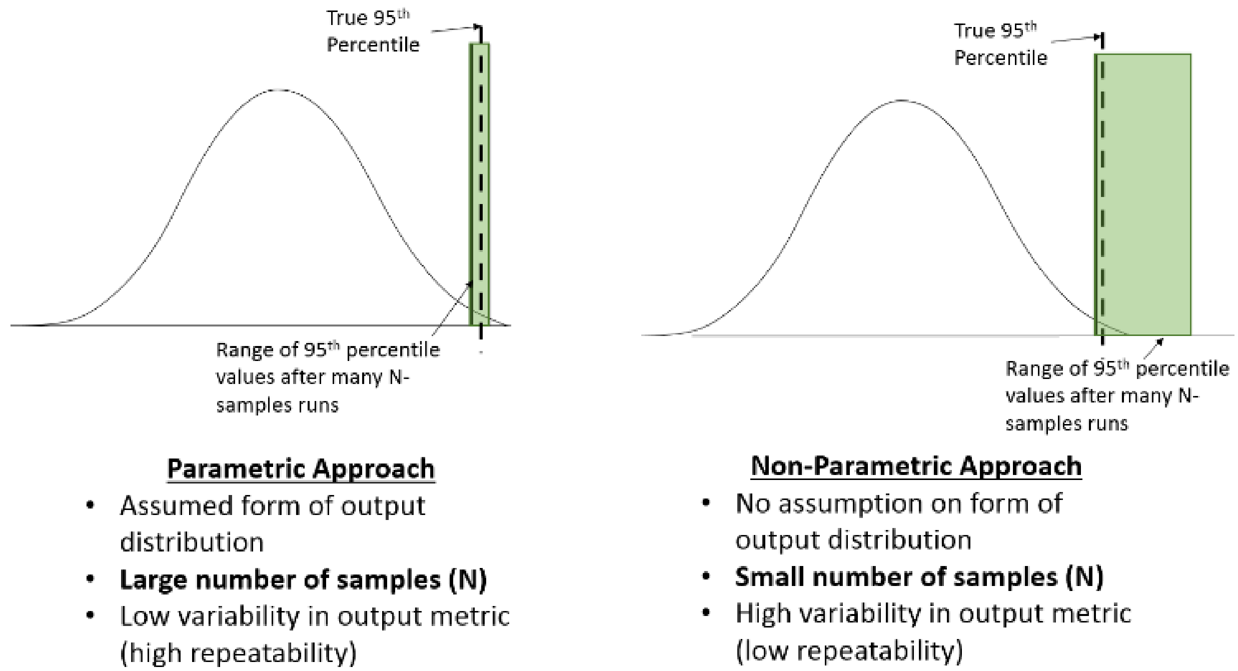

Fig. 1. Example 95**th** Percentile Values and Spreads using the LEFT: Current Parametric Approach and RIGHT: Proposed Non-parametric Approach.

Parametric Approach – Uncertainty in the Sample Variance

If an analyst wishes to estimate the population variance of a probability distribution a typical approach is to generate *N* trials from that distribution and make an estimate of the population variance by computing the sample variance, namely:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (2)$$

Where s^2 is the sample variance, n is the number of trials, x_i is the value from the i^{th} trial and $\bar{x}$ is the mean value from all of the trials. For a finite number of samples, there will be some variability in the value of s^2. Namely, if another set of n trials is run, a different value of s^2 will be obtained due to the inherent variability in the underlying random variable.

The uncertainty in s^2 can be estimated based upon the number of trials. This can be used to determine how good of an estimate s^2 is for the population variance, σ^2. If the underlying distribution is normal, then the distribution of s^2 can be determined analytically.

In order to determine the uncertainty in the estimate of the population variance, first start with the distribution of the quantity $(\frac{(n-1)s^2}{\sigma^2})$ which follows a chi-square distribution (χ^2) with $n-1$ degrees of freedom (see Reference 6):

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \qquad (3)$$

Where n is the number of sample trials, s^2 is the sample variance, σ^2 is the population variance and χ^2 is the chi-squared distribution with $n-1$ degrees of freedom. What is actually sought is the distribution of the sample variance, s^2, which can be determined using the properties of the chi-squared distribution.

In order to determine the distribution of s^2, first consider a random variable X that follows a chi-square distribution and a constant $c > 0$, so that:

$$X \sim \chi^2(\nu) \qquad (4)$$

$$c > 0 \qquad (5)$$

Where ν is the number of degrees of freedom in the chi-square distribution. The distribution of the random variable cX is then (Reference 6):

$$cX \sim \Gamma(k = \frac{\nu}{2}, \theta = 2c) \qquad (6)$$

Where Γ represents the gamma distribution with parameters k and θ. Using this transformation property, the distribution of the sample variance can be determined:

$$s^2 \sim \Gamma\left(k = \frac{n-1}{2}, \theta = \frac{2\sigma^2}{n-1}\right) \qquad (7)$$

Hence, the sample variance follows a gamma distribution with parameter $k = \frac{n-1}{2}$ and $\theta = \frac{2\sigma^2}{n-1}$. Using this result, the properties of the sample variance as a function of the number of trials can be determined. For this investigation, an example case was used where the population variance was equal to 1 ($\sigma^2 = 1$). Using the R code, the distribution of the sample variance was determined for various trial sizes and the 95th percentile of the sample variance was determined for each trial size investigated. The results of this evaluation are plotted in Fig. 2.

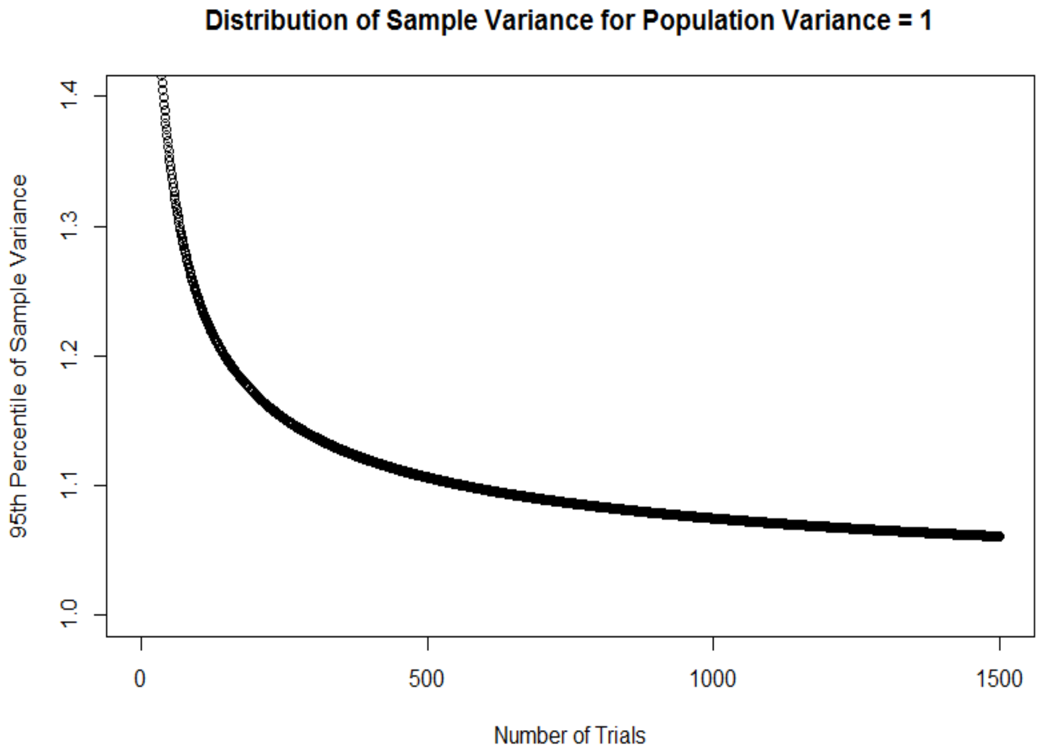

Fig. 2. Change in Sample Variance Distribution vs. Number of Trials

For n = 1200, s^2_{95} = 1.068 (the 95th percentile of the sample variance for N = 1200 samples). Note that 1200 samples corresponds to the default number of trials specified in the legacy SINDA-SAMPLE code for the parametric approach. This implies that $s_{95} \approx 1.033$ (the 95th percentile of the sample standard deviation for N=1200 samples)[a]. This implies that if 1200 trials are performed there is approximately a 3% uncertainty in the computed standard deviation at the 95% confidence level.

For a general situation where a set of N trials is performed and the sample standard variance is computed as a means to estimate the percentiles of an unknown output distribution, there is an unknown amount of confidence in the estimate of the percentile due to random variability in the variance estimate. However, as the number of trials increases, the error in the sample standard variance significantly decreases. Hence, while the confidence level is unknown this is of little practical significance since the error is expected to be small (only ~3% error in the sample standard deviation at the 95% level for 1200 trials).

Non-parametric (Wilks) Tests

Simple studies are performed and discussed in this section to better understand the implications of using Wilks' method as compared to the sample variance method described previously. The same example case was used, where an "inner" loop was used to generate N number of trials, and an "outer" loop was used to generate those N trials many times to understand the distribution of the tolerance values:

1. Determine number of samples (*N*) required to estimate an upper "α" percentile with "β" percent confidence using Equation (1). The number of samples that are expected to lie outside of the "α^{th}" percentile, and are therefore discarded, is controlled by setting the "order" value (p).
2. Use the R statistical package to monte-carlo sample results from a standard normal distribution ($\mu = 0, \sigma = 1$). As such, the 'theoretical' value for the α^{th} percentile is known exactly and can be compared to the Wilks estimated value.
3. Run 10,000 *N*-sample runs to provide 10,000 estimates of the α/β value
4. Determine the average of the α/β estimates from the 10,000 trials.
5. Calculate the % error of the average value from the "true" α^{th} percentile of a standard normal distribution $\left(\%\ error = \frac{\overline{\alpha/\beta_{wilks}} - \alpha_{theoretical}}{\alpha_{theoretical}} * 100\%\right)$.

Note that since the underlying distribution is known, we know what the true 95th percentile one-sided upper tolerance limit should be. Since the chosen distribution is a standard normal, the 95th percentile one-sided tolerance limit can be determined from Equation (8):

$$0.95 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \qquad (8)$$

Solving for *x* yields that the 95th percentile upper tolerance bound for a standard normal (p_{95}) is 1.645. Solving for *x* at the 99th percentile upper tolerance bound yields a value of 2.326.
A total of twelve Wilks sensitivity tests were executed to predict either the 95th or 99th percentile value of the normal distribution: 1.) 95/95, *p*=1 2.) 95/99, *p*=1 3.) 9 9/95, *p*=1 4.) 99/99, *p*=1 5.) 95/95, *p*=2 6.) 95/95, *p*=3 7.) 95/95, *p*=4 8.) 95/95, *p*=5 9.) 99/95, *p*=2 10.) 99/95, *p*=3 11.) 99/95, *p*=4 12.) 99/95, *p*=5. The number of runs required by these twelve sensitivity tests are calculated via Equation (1), and are summarized below in Table I. These test loops described above were automated using the R statistical package.

4 Results and Conclusions

Table I below summarizes the results of the twelve Wilks non-parametric tests. As shown, the error of the mean Wilks α/β estimate is always positive (i.e. the percentile is over-predicted), and tends to be larger at the first-order (where no results are discarded). The absolute error of the Wilks approach is shown to decrease monotonically over an increase in requested percentile or order, and increases monotonically with increased confidence. The results of the Wilks method at the first order (*p*=1) specifically are shown to be substantially higher than the errors predicted using higher order approaches. This study suggests that while the Wilks first-order 95/95 formula (a commonly used approach in the nuclear industry for BEPU methods) can be effective at quantifying uncertainty of results with as few as 59 code runs, the limiting values will tend to substantially over predict the 95th percentile, and may lead to overly-conservative results being generated.

Table I. Summary of Results – Wilks' Parametric Tests				
# of Trials	**Case Name**	**Theoretical Value**	***α/β* estimate [percentile]**	**% error (Wilks)**
59	(95/95) p=1	1.645	2.313 [99.97th]	+*40.6*
90	(95/99) p=1	1.645	2.471 [99.32th]	+*50.2*
299	(99/95) p=1	2.326	2.874 [99.80th]	+*23.6*
459	(99/99) p=1	2.326	3.011 [99.87th]	+*29.4*
93	(95/95) p=2	1.645	2.115 [98.27th]	+*28.6*
124	(95/95) p=3	1.645	2.034 [97.90th]	+*23.6*
153	(95/95) p=4	1.645	1.989 [97.67th]	+*20.9*
181	(95/95) p=5	1.645	1.954 [97.46th]	+*18.8*
473	(99/95) p=2	2.326	2.716 [99.67th]	+*16.8*
628	(99/95) p=3	2.326	2.643 [99.59th]	+*13.6*
773	(99/95) p=4	2.326	2.605 [99.54th]	+*12.0*
913	(99/95) p=5	2.326	2.576 [99.50th]	+*10.7*

The evaluation discussed here and summarized in Table I has demonstrated that Wilks' formula is a robust approach that is comparable to analytical methods for similar levels of probability/confidence. This work has also shown that both methods are conservative at predicting upper percentile values, and the conservatism is reduced with increased number of samples. The number of code runs is heavily dependent on the statistical probability requested of the output, while increased probability is shown to have only a moderate impact on reducing the relative error and variance of the results, especially at higher orders.

The order of Wilks' equation used to generate sample numbers is demonstrated to have the most impact on repeatability and accuracy of the results. A trade-off is made to set the order at a respectable level where diminishing returns are expected for higher orders. It is therefore recommended that the SASQUATCH code adopt the 4th order approach (95/95, *p*=4), which requires 153 samples (code runs) to meet the 95/95 probability/confidence threshold. This approach is recommended to optimize the number of code runs with a relatively accurate, generally conservative, and repeatable set of output. While the 95/95 parametric method used in SINDA-SAMPLE demonstrated an approximately +/-3% error in the computed output at 1,200 runs, the 95/95, *p*=4 Wilk's approach demonstrates a +21% error in the computed output. This is because, on average, the 4th order 95/95 approach will in fact estimate the 97.7th percentile. This is considered an acceptable increase in conservatism in exchange for the marked reduction in number of samples required.

It is therefore recommended that the 153 code runs be applied for safety basis analyses using SASQUATCH to generate safety margin data at the 95th percentile with 95% confidence.

5 Acknowledgements

This work was supported by the Idaho National Laboratory (INL). The author wishes to thank colleagues from the INL who provided insight and expertise that greatly assisted this work.

6 References

[1] A. Olson, S.A. Jonah, “MNSR Transient Analysis and Thermal-Hydraulic Safety Margins for HEU and LEU Cores Using PARET,” Proceedings of the XXIX International Meeting on Reduced Enrichment for Research and Test Reactors, Prague, Czech Republic, September, 23-27, 2007.

[2] K. H. LIEBELT, “User’s Manual for SINDA-SAMPLE Version A, Mod 1 A SINDA Thermal-Hydraulic/Statistical Model of the Hot Stripe in an ATR Fuel Plate,” PG-T-96-006, November 1996.

[3] L. ZOU, “User Manual for SASQUATCH,” GDE-793, Rev. 1, 2018.

[4] R. MARTIN, “Development Considerations of AREVA NP Inc.’s Realistic LBLOCA Analysis Methodology,” *Science and Technology of Nuclear Installations*, 2008, (2008).

[5] C. FREPOLI, “Realistic Large Break LOCA Analysis of AP1000 with ASTRUM,” *The 6th International Conference on Nuclear Thermal Hydraulics, Operations and Safety (NUTHOS ’04)*, Nara, Japan, October 4-8, 2004 (CD-ROM).

[6] G. B. WALLIS, “Uncertainties and Probabilities in Nuclear Reactor Regulation,” *Nuclear Engineering and Design*, 237, 1586-1592 (2007).

[7] G. J. Kerns, *Introduction to Probability and Statistics Using R*, p. 148, (2010).