Reliability Engineering and System Safety 149 (2016) 187–203

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Advantages of variance reduction techniques in establishing confidence intervals for quantiles

Dave Grabaskas [a,1,*], Marvin K. Nakayama [b,2], Richard Denning [c,3], Tunc Aldemir [c,4]

[a] Ohio State University, 201 W 19th Ave, Columbus, OH 43210, USA
[b] New Jersey Institute of Technology, Computer Science Department, Newark, NJ 07102, USA
[c] Ohio State University, USA

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form
18 December 2015
Accepted 19 December 2015
Available online 4 January 2016

Keywords:
Confidence intervals
Quantiles
Nuclear regulation
Best-estimate
Uncertainty

ABSTRACT

Over the past two decades, U.S. nuclear power plant regulation has increasingly depended on best-estimate plus uncertainty safety analyses. As a result of the shift to best-estimate analyses, the distribution of the output metric must be compared against a regulatory goal, rather than a single, conservative value. This comparison has historically been conducted using a 95% one-sided confidence interval for the 0.95-quantile of the output distribution, which is usually found following the technique of simple random sampling using order statistics (SRS-OS). While SRS-OS has certain statistical advantages, there are drawbacks related to the available sampling schemes and the accuracy and precision of the resulting value. Recent work has shown that it is possible to establish asymptotically valid confidence intervals for a quantile of the output of a model simulated using variance reduction techniques (VRTs). These VRTs can provide more informative results than SRS-OS. This work compares SRS-OS and the VRTs of antithetic variates and Latin hypercube sampling through several experiments, designed to replicate conditions found in nuclear safety analyses. This work is designed as an initial investigation into the use of VRTs as a tool to satisfy nuclear regulatory requirements, with hope of expanded analyses of VRTs in the future.

1. Introduction and background

In its efforts to transition from conservative regulatory models to best-estimate plus uncertainty analyses, the Nuclear Regulatory Commission (NRC) has developed safety analysis guidelines that require quantification of the impact of uncertainties on the output of accident simulations [1]. While different methods to meet this quantitative requirement have been discussed [2], the most common approach is to calculate a confidence interval for a quantile of the output distribution. An NRC-approved method of accomplishing this task has been the technique of simple random sampling using order statistics (SRS-OS) [2]. While this method has many positive aspects for nuclear safety analysts, it has certain drawbacks related to the available sampling schemes and the accuracy and precision of the resulting value. Recent work has shown that it is possible to establish asymptotically valid confidence intervals for a quantile of the output of a model simulated using variance reduction techniques (VRTs), such as Latin hypercube sampling (LHS) [3]. These VRTs can possibly provide more informative results than SRS-OS. The current work compares SRS-OS and the VRTs of antithetic variates and LHS through several experiments, designed to replicate conditions found in nuclear safety analyses. These tests include a simple nonlinear equation system, a design-basis accident analysis of a nuclear power plant using a response surface surrogate for the thermal-hydraulic code RELAP5 [4], and a beyond-design-basis accident analysis conducted using the severe-accident analysis computer code MELCOR [5]. This work was designed as an initial investigation into the use of VRTs as a tool to satisfy nuclear regulatory requirements, with the hope of expanded analyses of VRTs in the future.

Section 1 begins with an overview of regulatory history and a quick description of hypothesis testing, which is used to frame the issue of regulatory compliance in a more-rigorous fashion. This is followed in Section 2 by a review of the statistical methods that are later compared through a series of example problems in Section 3. The conclusions are reviewed in Section 4.

* Corresponding author.
E-mail addresses: Grabaskas.2@osu.edu (D. Grabaskas), Marvin@njit.edu (M.K. Nakayama), Denning.8@osu.edu (R. Denning), Aldemir.1@osu.edu (T. Aldemir).
[1] Tel.: +1 630 252 2195.
[2] Tel.: +1 973 596 3398.
[3] Tel.: +1 614 292 2544.
[4] Tel.: +1 614 292 4627.

http://dx.doi.org/10.1016/j.ress.2015.12.015
0951-8320/

Nomenclature

95/95	95% One-sided Confidence Interval for 0.95-quantile
AV	Antithetic Variates
BFD	Backward Finite-Difference
CDF	Cumulative Distribution Function
CFD	Central Finite-Difference
CI	Confidence Interval
CLT	Central Limit Theorem
LHS	Latin Hypercube Sampling
LOCA	Loss of Coolant Accident
NRC	Nuclear Regulatory Commission
rLHS	Replicated Latin Hypercube Sampling
SRS	Simple Random Sampling
SRS-OS	Simple Random Sampling Order Statistics
VRT	Variance Reduction Technique
Run	A single execution of a computer code
Case	A collection of runs for the rLHS method
Trial	A complete experiment that an analyst would conduct; for the rLHS method, it would consist of multiple cases.
Accuracy	A measure of the expected distance between the correct quantile and the upper endpoint of a one-sided CI that arises in the 95/95 analysis
Precision	The variance of possible upper endpoints of one-sided CIs

1.1. Regulatory background

The initial approach to the treatment of modeling uncertainties in regulatory analysis was to use non-mechanistic, conservative models. In the implementation of the Part 50 Appendix K of the Code of Federal Regulations [6], which describes a prescription for the conservative treatment of uncertainties in the analysis of loss-of-coolant accidents (LOCAs), it became apparent that what was thought to be conservative might not be conservative in all cases, and that conservative regulatory models could be misleading with regard to the improvement of reactor safety. The transition to best-estimate plus uncertainty regulatory requirements began with an amendment to 10 CFR 50.46 [1] in 1988, which allowed for realistic modeling of LOCAs. While this rule-change signaled an advancement in regulatory safety analysis, the statistical requirements of the output result were vague, stating only that there should be a "high level of probability that the criteria would not be exceeded."

In 1989, the NRC issued RG1.157 [7], which helped clarify the procedure for performing a best-estimate calculation relating to the design bases for essential safety systems. It set the standard for the handling of computational uncertainty for nuclear safety applications by stating that a 95% probability level is considered acceptable to the NRC staff for comparison of best-estimate predictions to safety limits. However, the ambiguity of the term "95% probability level" remained an issue for the analyst.

The most obvious solution to the "95% probability" requirement was to estimate the 0.95-quantile of the output distribution. One method to do this was to perform a large number of simple random sampling (SRS) computer code runs using Monte Carlo sampling and simply order and count the results until 95% of the runs fell below that value. Then this point estimate of the 0.95-quantile would be compared to the safety limit. The large number of runs required by SRS to obtain sufficient accuracy represented a major problem for safety analysts, due to minimal computing power and extended code run times. There was also the question of just how many runs would be necessary for an analyst to be able to claim that the estimate of the 0.95-quantile was sufficiently accurate.

Response-surface methods [8] were initially proposed as a way of reducing runs and increasing knowledge of the overall behavior of the parameters of interest. An advantage of this approach is that it employs a fixed matrix of runs to be conducted to estimate the desired surface. This property not only gives the analyst a plan to provide to the regulator, but also produces a level of understanding about the impact of different input parameters. However, like the large-sample SRS case, run designs often needed to be very large to capture input interactions and nonlinearities, and the only way around this was to group input parameters based on the analyst's judgment [2]. In response to these considerations, methods were developed that required a smaller number of runs, but which could satisfy the regulatory guidelines.

Both AREVA [9] and Westinghouse [2] developed approaches for the use of simple random sampling using order statistics (SRS-OS) for their regulatory LOCA analyses. While the method of SRS-OS was first considered for use in the nuclear industry in the 1970's [10], it was not until the NRC published NUREG-1475 [11], a guide to applying statistics, in 1994 that the NRC provided a more comprehensive picture of its use for regulatory requirements. Gesellschaft fürAnlagen-und Reaktorsicherheit (GRS) helped bring SRS-OS to the thermal hydraulic and safety fields soon after [12]. Major steps forward occurred in 2003 and 2004 with publications by Guba, Pál, and Makai [13], and Nutt and Wallis [14]. These works not only expanded on how SRS-OS could be used in safety analyses, but also demonstrated how it could be applied to satisfy the 95% probability reporting requirement. The solution provided by Guba, Pál, and Makai [13] and Nutt and Wallis [14] to this question was to report a 95% upper one-sided confidence interval for the 0.95-quantile of the output distribution. Based on the works of Wilks [15] and Wald [16], this method simulates the model using SRS and specifies a particular order statistic as an endpoint of a 95% tolerance interval with 95% confidence. This method was considered acceptable by the NRC in regards to the 95% probability requirement [2], and is discussed in detail in Section 2.1.

While the acceptance of the 95% confidence interval for the 0.95-quantile has been adopted by the NRC for satisfying design-basis accident requirements, there are other safety applications for which less stringent requirements may be appropriate, such as for the analysis of beyond-design-basis events. For the analysis of these events, similar, but less stringent limits could be established, such as the use of the 95% confidence level with a lower quantile.

1.2. Comparison to safety requirement within the context of hypothesis testing

The process of using a confidence interval for a quantile to compare to a regulatory safety limit can be explained more rigorously using hypothesis testing. For example, assume there is a regulatory safety goal with value G, that represents a prescribed limit that the true 0.95-quantile ($\xi_{0.95}$) of the output of a safety analysis cannot exceed. In this case, we define a hypothesis test, with null hypothesis $H_0{:}\xi_{0.95} \geq G$ and alternative hypothesis $H_1{:}\xi_{0.95} < G$. This framework puts the burden of proof on H_1, which hypothesizes that the true 0.95-quantile value of the output falls below the prescribed limit. Hypothesis testing uses a statistic to make a decision about a parameter. Since the true 0.95-quantile $\xi_{0.95}$ of the system, a parameter, is unknown, it needs to be estimated. Define a 95/95 value to be the upper confidence limit of an