Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

VARIANCE REDUCTION FOR ESTIMATING A FAILURE PROBABILITY WITH MULTIPLE CRITERIA

Andres Alban

Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, NJ 07102, USA

Hardik A. Darji

Department of Mathematical Sciences
Department of Mechanical Engineering
New Jersey Institute of Technology
Newark, NJ 07102, USA

Atsuki Imamura

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA

Marvin K. Nakayama

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA

ABSTRACT

We consider a system subjected to multiple loads with corresponding capacities to withstand the loads, where both loads and capacities are random. The system fails when any load exceeds its capacity, and the goal is to apply Monte Carlo methods to estimate the failure probability. We consider various combinations of variance-reduction techniques, including stratified sampling, conditional Monte Carlo, and Latin hypercube sampling. Numerical results are presented for an artificial safety analysis of a nuclear power plant, which illustrate that the combination of all three methods can greatly increase statistical efficiency.

1 INTRODUCTION

Consider a system with several random loads and corresponding random capacities, and the system fails when any load exceeds its capacity. The problem is motivated by nuclear power plant (NPP) safety analyses, and our goal is to estimate the failure probability using Monte Carlo simulation. Current rules of the U.S. Nuclear Regulatory Commission (2010) (NRC; paragraph 50.46(b)) specify that during a hypothesized loss-of-coolant accident (LOCA), the peak cladding temperature (PCT) must lie below 2200°F with at least 0.95 probability. To study a postulated LOCA, nuclear engineers have developed deterministic computer codes (Hess et al. 2009), which take as input random variables having a given joint distribution and output a PCT during the LOCA. The input random variables may specify the timing and location of events during the LOCA, e.g., when and where a pipe break occurs. Each code run is computationally expensive as it numerically solves systems of differential equations, limiting the number of runs that can be done. To account for the statistical variability of the Monte Carlo estimates, the NRC also requires that the analysis is carried out with 95% confidence. Nuclear engineers currently perform such a safety analysis by comparing a 95% upper confidence bound (UCB) for the 0.95-quantile of the PCT to the fixed capacity 2200°F. This is known as a *95/95 analysis*; e.g., see Section 24.9 of U.S. Nuclear Regulatory Commission (2011). The difference between the fixed capacity and the UCB provides a type of *safety margin*. In addition to the PCT, the NRC further indicates limits on the core-wide oxidation (CWO $< 1\%$) and the maximum local oxidation (MLO $< 17\%$). Thus, NPP safety analyses currently consider $q = 3$ criteria (PCT, CWO, and MLO) with their corresponding (random) loads and (fixed) capacities.

Important recent changes in NPPs have prompted exploring alternative approaches to assess risk. Many NPPs are now aging past their original 40-year operating licences, with resulting component wear, and licensees are applying for extensions. Also, plants are sometimes being run at higher output levels, known as *power uprates* (Dube et al. 2014), to increase their economic viability. These developments can lead to degradations in safety margins that were previously deemed acceptable. To better understand their impacts, the Nuclear Energy Agency Committee on the Safety of Nuclear Installations (2007) developed a framework called *risk-informed safety-margin characterization* (RISMC). The approach differs from the previous 95/95 analysis in several key aspects. First, rather than assuming fixed capacities, RISMC allows them to be random (with specified distributions). Moreover, instead of focusing on quantiles, RISMC considers the failure probability θ that at least one load exceeds its capacity. RISMC also partitions the sample space of a hypothesized accident into a finite collection of *scenarios* based on an *event tree*, where each scenario has a known probability of occurring but the failure probability within each scenario still needs to be estimated. The sample-space decomposition leads to applying *stratified sampling* (SS) to estimate θ. To account for the sampling error of the Monte Carlo results, we further want a UCB for θ.

The current paper combines SS with other variance-reduction techniques (VRTs)—Latin hypercube sampling (LHS) and conditional Monte Carlo (CMC)—with the goal of estimating the failure probability with $q \geq 1$ criteria. (See Chapter 4 of Glasserman 2004 and Chapter V of Asmussen and Glynn 2007 for overviews of these and other VRTs.) We provide asymptotically valid (as the total sample size grows large) UCBs for θ when applying replicated LHS (Iman 1981). An important modeling issue when $q > 1$ is how to incorporate dependencies among the criteria, especially for the capacities. In actual NPP safety analyses, detailed computed codes (Hess et al. 2009) simultaneously generate the multiple criteria's *loads*, thereby inducing particular dependencies. But there have not been any previous multi-criteria RISMC studies, so the issue of how to specify dependence among the *capacities* needs to be addressed. In our numerical experiments we apply Gaussian *copulas* (Nelsen 2006) to model the dependence structure, but this is a topic that deserves further study. Our numerical experiments show that the combination of the three VRTs can work synergistically to greatly reduce variance compared to applying smaller combinations of VRTs.

Our work builds and extends results of several previous papers. The initial studies with the RISMC approach in Sherry, Gabor, and Hess (2013) and Dube et al. (2014) consider only a single criterion (PCT, so $q = 1$). They apply a combination of SS and LHS, omitting CMC. Moreover, they do not present UCBs to account for statistical error. Nakayama (2015) considers the combination of all three VRTs in the current paper and provides an asymptotically valid UCB, but for only a single criterion (PCT, so $q = 1$). Avramidis and Wilson (1996) study the combination of LHS and CMC but omit SS, which plays a critical role in RISMC; also, the authors do not provide UCBs.

The rest of the paper unfolds as follows. Section 2 lays out the mathematical framework of the problem. Section 3 reviews the use of simple random sampling to estimate θ. Sections 4, 5, and 6 develop the VRTs SS, LHS, and CMC, respectively, in combinations. Numerical results for a synthetic model of a RISMC analysis are presented in Section 7. In Section 8, we provide concluding remarks. The proofs of the theorems and additional numerical results appear in a follow-up paper (Alban et al. 2016).

2 MATHEMATICAL FRAMEWORK

For a fixed number $q \geq 1$ of criteria, let $L = (L^{[1]}, L^{[2]}, \ldots, L^{[q]})$ (resp., $C = (C^{[1]}, C^{[2]}, \ldots, C^{[q]})$) be a random vector of q loads (resp., capacities). (Throughout the paper, all vectors are column vectors.) For a nuclear safety analysis, we have $q = 3$ criteria, with $(L^{[1]}, C^{[1]})$, $(L^{[2]}, C^{[2]})$, and $(L^{[3]}, C^{[3]})$ representing load-capacity pairs for the PCT, CWO, and MLO criteria, respectively. Let H be the (unknown) joint cumulative distribution function (CDF) of (L, C), and let F (resp., G) denote the marginal CDF of the vector L (resp., C). Let P (resp., E) be the probability (resp., expectation) operator, and let I be the indicator function, which returns 1 (resp., 0) for a true (resp., false) argument.

We define θ as the probability that any load exceeds its capacity:

$$\theta = P\left(\cup_{k=1}^{q}\left\{L^{[k]} \geq C^{[k]}\right\}\right) = E\left[I\left(\cup_{k=1}^{q}\left\{L^{[k]} \geq C^{[k]}\right\}\right)\right]. \tag{1}$$

When estimating an expectation, we call the quantity whose expectation is desired a *response function*, which in (1) is an indicator function. Our goal is to determine if the failure probability is acceptably small, i.e., if $\theta < \theta_0$ for some given $0 < \theta_0 < 1$. We further require establishing this with a given confidence level $0 < \gamma < 1$. Thus, we summarize the requirements for an acceptably safe system as follows:

$$\text{given constants } 0 < \theta_0 < 1 \text{ and } 0 < \gamma < 1, \text{ determine with confidence level } \gamma \text{ if } \theta < \theta_0. \tag{2}$$

The values of θ_0 and γ may be specified by a regulator, e.g., $\theta_0 = 0.05$ and $\gamma = 0.95$.

We can use a hypothesis test to satisfy requirement (2). Let $\mathscr{H}_0 : \theta \geq \theta_0$ be the null hypothesis and $\mathscr{H}_1 : \theta < \theta_0$ be the alternative. We perform the hypothesis test at a significance level $\alpha = 1-\gamma$ by carrying out a total of n simulation runs. From the output of the n runs, we construct a point estimator $\hat{\theta}(n)$ of θ, along with a γ-level *upper confidence bound* $B(n)$, i.e., $P(\theta \leq B(n)) = \gamma$. Our methods to build UCBs are derived from central limit theorems (CLTs), so that $B(n)$ is instead an *asymptotic* γ-level UCB:

$$\lim_{n\to\infty} P(\theta \leq B(n)) = \gamma. \tag{3}$$

Therefore, we can asymptotically satisfy requirement (2) with the following *decision rule*:

$$\text{reject } \mathscr{H}_0 \text{ if and only if } B(n) < \theta_0, \tag{4}$$

3 SIMPLE RANDOM SAMPLING

We first consider estimating θ with *simple random sampling* (SRS), which is also known as *naive simulation*, *standard simulation*, or *crude Monte Carlo*. To do this, generate a sample of $n \geq 2$ independent and identically distributed (i.i.d.) copies (L_i, C_i), $i = 1,2,\ldots,n$, of $(L,C) \sim H$, where $L_i = (L_i^{[1]}, L_i^{[2]}, \ldots, L_i^{[q]})$, $C_i = (C_i^{[1]}, C_i^{[2]}, \ldots, C_i^{[q]})$, and $\sim$ means "has distribution." Then the SRS point estimator of θ in (1) is $\hat{\theta}_{\mathrm{SRS}}(n) = (1/n)\sum_{i=1}^{n} I(\cup_{k=1}^{q}\{L_i^{[k]} \geq C_i^{[k]}\})$, which estimates the expectation in (1) by averaging i.i.d. copies of the response function. Because $I(\cup_{k=1}^{q}\{L_i^{[k]} \geq C_i^{[k]}\})$, $i = 1,2,\ldots,n$, are i.i.d. and bounded, the ordinary CLT (e.g., Theorem 27.1 of Billingsley 1995) implies $[\sqrt{n}/\hat{\sigma}_{\mathrm{SRS}}(n)]\left[\hat{\theta}_{\mathrm{SRS}}(n) - \theta\right] \Rightarrow N(0,1)$ as $n \to \infty$, where $\Rightarrow$ denotes weak convergence (e.g., see Chapter 5 of Billingsley 1995), $N(a,b^2)$ is a normal random variable with mean a and variance b^2, and $\hat{\sigma}^2_{\mathrm{SRS}}(n) = \hat{\theta}_{\mathrm{SRS}}(n)[1-\hat{\theta}_{\mathrm{SRS}}(n)]$ *consistently* estimates $\sigma^2_{\mathrm{SRS}} = \theta[1-\theta]$; i.e, $\hat{\sigma}^2_{\mathrm{SRS}}(n) \Rightarrow \sigma^2_{\mathrm{SRS}}$ as $n \to \infty$. Let z_γ be the γ-level upper one-sided critical point of $N(0,1)$, which satisfies $P(N(0,1) \leq z_\gamma) = \gamma$, e.g., $z_{0.95} = 1.645$. Then $B_{\mathrm{SRS}}(n) \equiv \hat{\theta}_{\mathrm{SRS}}(n) + z_\gamma \hat{\sigma}_{\mathrm{SRS}}(n)/\sqrt{n}$ is an asymptotic γ-level UCB satisfying (3). A simple modification of the well-known sign test (e.g., Example 3.2.4 of Lehmann 1999), this approach has been proposed for multi-criteria nuclear safety analyses with *fixed* capacities in Section 4.2 of Pál and Makai (2013), while our setup allows for random C.

4 STRATIFIED SAMPLING

One key aspect of RISMC (see Section 1) is decomposing the sample space into a finite number of scenarios through an *event tree*. Figure 1 portrays an example of an event tree, originally from Dube et al. (2014), with $t = 4$ *scenarios* for a hypothesized station blackout (SBO). The event tree has 3 intermediate events E_1, E_2, E_3, which determine how the SBO progresses. For example, the lower (resp., upper) branch of E_2 denotes that a safety relief valve is stuck open (resp., closes properly), the number in each case indicating its probability of occurrence, which is assumed known. Thus, the $t = 4$ scenarios are determined by following paths from left to right, and the probability of a scenario is computed as the product of the branches taken

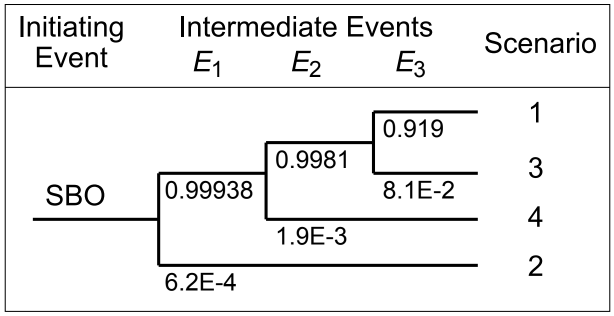

Figure 1: An event tree for a hypothesized station blackout.

for the intermediate events along the path. For example, scenario 4 has probability $0.99938 \times 1.9\text{E-3}$. But each scenario's failure probability is unknown and is estimated via some form of Monte Carlo.

The event-tree framework is well suited for applying stratified sampling. In general, SS partitions the sample space into $t \geq 1$ *strata*, where each stratum has a known probability of occurring; see Section 4.3 of Glasserman (2004) for an overview of SS. To apply SS, define a *stratification variable* A, which is generated along with (L, C) during the simulation, and partition the support R of A into $t \geq 1$ subsets $R_1, R_2, \ldots, R_t$, with $R = \cup_{s=1}^{t} R_s$ and $R_s \cap R_{s'} = \emptyset$ for $s \neq s'$. Let $\lambda = (\lambda_s : s = 1, 2, \ldots, t)$, where each $\lambda_s = P(A \in R_s)$ is assumed known. We call each R_s a *stratum*, and its index s a *scenario*. In the case of an event tree, we can take A to be the randomly chosen scenario, so $R_s = \{s\}$ for $s = 1, 2, \ldots, t$, with, e.g., $\lambda_4 = 0.99938 \times 1.9\text{E-3}$ in Figure 1. We assume that for each R_s, we can sample a load and capacity pair from its conditional distribution given that $A \in R_s$. Let $(L_{(s)}, C_{(s)})$ be a random vector having the conditional distribution of (L, C) given $A \in R_s$, where $L_{(s)} = (L_{(s)}^{[1]}, L_{(s)}^{[2]}, \ldots, L_{(s)}^{[q]})$ and $C_{(s)} = (C_{(s)}^{[1]}, C_{(s)}^{[2]}, \ldots, C_{(s)}^{[q]})$. Let $H_{(s)}$ denote the joint CDF of $(L_{(s)}, C_{(s)})$ in scenario s, and let $F_{(s)}$ (resp., $G_{(s)}$) be the marginal CDF of the vector of loads $L_{(s)}$ (resp., capacities $C_{(s)}$). Then we can express the overall failure probability in (1) as

$$\theta = \sum_{s=1}^{t} P(A \in R_s) P\left(\cup_{k=1}^{q} \left\{ L^{[k]} \geq C^{[k]} \right\} \,\middle|\, A \in R_s \right) = \sum_{s=1}^{t} \lambda_s \theta_{(s)} \tag{5}$$

by the law of total probability, where each λ_s is known, but each

$$\theta_{(s)} = P\left(\cup_{k=1}^{q} \left\{ L_{(s)}^{[k]} \geq C_{(s)}^{[k]} \right\} \right) = E\left[I\left(\cup_{k=1}^{q} \left\{ L_{(s)}^{[k]} \geq C_{(s)}^{[k]} \right\} \right) \right] \tag{6}$$

is not. We then use some form of simulation to estimate each $\theta_{(s)}$, which we combine as in (5) to obtain an estimator of θ. (A RISMC study may also want to identify unsafe scenarios s for which $\theta_{(s)}$ is large.)

Specifically, we implement SS with overall sample size n by letting $n_s = \eta_s n$ be the sample size allocated to scenario s, where η_s, $s = 1, 2, \ldots, t$, are user-specified positive constants summing to 1. (As we will require each $n_s \to \infty$ for our asymptotic theory, the number t of strata cannot be too large in practice, limiting the number of intermediate events that can be considered in an event tree.) We call $\eta = (\eta_s : s = 1, 2, \ldots, t)$ the SS *allocation*. One possibility is $\eta = \lambda$, but we allow other choices. For simplicity, assume that n_s is an integer; otherwise, let $n_s = \lfloor \eta_s n \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function. For each scenario $s = 1, 2, \ldots, t$, let $(L_{(s),i}, C_{(s),i})$, $i = 1, 2, \ldots, n_s$, be a sample of n_s i.i.d. copies of $(L_{(s)}, C_{(s)})$, where $L_{(s),i} = (L_{(s),i}^{[1]}, L_{(s),i}^{[2]}, \ldots, L_{(s),i}^{[q]})$ (resp., $C_{(s),i} = (C_{(s),i}^{[1]}, C_{(s),i}^{[2]}, \ldots, C_{(s),i}^{[q]})$) is an observation of the q criteria's loads $(L^{[1]}, L^{[2]}, \ldots, L^{[q]})$ (resp., capacities $(C^{[1]}, C^{[2]}, \ldots, C^{[q]})$) given $A \in R_s$. The response function in (6) is an indicator, and we estimate its expectation $\theta_{(s)}$ by

$$\hat{\theta}_{(s),\mathrm{SS},\eta}(n) = \frac{1}{n_s} \sum_{i=1}^{n_s} I\left(\cup_{k=1}^{q} \left\{ L_{(s),i}^{[k]} \geq C_{(s),i}^{[k]} \right\} \right), \tag{7}$$

where the subscript SS denotes stratified sampling with simple random sampling applied within each stratum. The SS estimator of $\theta = \sum_{s=1}^{t} \lambda_s \theta_{(s)}$ in (5) is then

$$\hat{\theta}_{\mathrm{SS},\eta}(n) = \sum_{s=1}^{t} \lambda_s \hat{\theta}_{(s),\mathrm{SS},\eta}(n). \tag{8}$$

For each scenario $s = 1,2,\ldots,t$, the estimator $\hat{\theta}_{(s),\mathrm{SS},\eta}(n)$ satisfies a CLT

$$\frac{\sqrt{n}}{\hat{\sigma}_{(s),\mathrm{SS},\eta}(n)/\sqrt{\eta_s}} \left[\hat{\theta}_{(s),\mathrm{SS},\eta}(n) - \theta_{(s)}\right] \Rightarrow N(0,1) \tag{9}$$

as $n \to \infty$, where $\hat{\sigma}^2_{(s),\mathrm{SS},\eta}(n) \equiv \hat{\theta}_{(s),\mathrm{SS},\eta}(n)[1-\hat{\theta}_{(s),\mathrm{SS},\eta}(n)]$ consistently estimates $\sigma^2_{(s),\mathrm{SS}} \equiv \theta_{(s)}[1-\theta_{(s)}]$. (The extra factor $\sqrt{\eta_s}$ appears on the left side of (9) because the scaling is $\sqrt{n}$ but the estimator $\hat{\theta}_{(s),\mathrm{SS},\eta}(n)$ in (7) is based on a sample size of $n_s = \eta_s n$.) Assuming that the t scenarios for SS are simulated independently, we then have that (9) jointly holds for $s = 1,2,\ldots,t$, by Problem 29.2 of Billingsley (1995), so the SS estimator $\hat{\theta}_{\mathrm{SS},\eta}(n)$ of the overall failure probability θ satisfies the CLT $[\sqrt{n}/\hat{\sigma}_{\mathrm{SS},\eta}(n)]\left[\hat{\theta}_{\mathrm{SS},\eta}(n) - \theta\right] \Rightarrow N(0,1)$ as $n \to \infty$, where $\hat{\sigma}^2_{\mathrm{SS},\eta}(n) \equiv \sum_{s=1}^{t} \lambda_s^2 \hat{\sigma}^2_{(s),\mathrm{SS},\eta}(n)/\eta_s$ consistently estimates $\sigma^2_{\mathrm{SS},\eta} \equiv \sum_{s=1}^{t} \lambda_s^2 \sigma^2_{(s),\mathrm{SS}}/\eta_s$; e.g., see p. 215 of Glasserman (2004). Finally, an asymptotic γ-level UCB for θ satisfying (3) when applying SS is

$$B_{\mathrm{SS},\eta}(n) = \hat{\theta}_{\mathrm{SS},\eta}(n) + z_\gamma \frac{\hat{\sigma}_{\mathrm{SS},\eta}(n)}{\sqrt{n}}. \tag{10}$$

5 COMBINED SS AND LATIN HYPERCUBE SAMPLING

LHS is one of the most popular VRTs applied in nuclear engineering; e.g., Helton and Davis (2003) cite over 300 references using LHS. Further incorporating LHS requires imposing additional problem structure. Let $U[0,1)$ represent a uniform random number on the unit interval, and we assume the following:

Assumption 1 For each scenario $s = 1,2,\ldots,t$, there is a deterministic function $w_{(s)} : \Re^{d_s} \to \Re^{2q}$ such that if $U_1, U_2, \ldots, U_{d_s}$ are d_s i.i.d. $U[0,1)$ random variables, then

$$(L^{[1]}_{(s)}, L^{[2]}_{(s)}, \ldots, L^{[q]}_{(s)}, C^{[1]}_{(s)}, C^{[2]}_{(s)}, \ldots, C^{[q]}_{(s)}) = w_{(s)}(U_1, U_2, \ldots, U_{d_s}) \sim H_{(s)}. \tag{11}$$

The function $w_{(s)}$ takes d_s i.i.d. $U[0,1)$ random numbers as inputs, and transforms them into an observation of the load and capacity vectors having the correct joint distribution $H_{(s)}$ for scenario s. In the context of nuclear safety, the function $w_{(s)}$ first converts the uniforms into a random vector X having a specified joint distribution, where the components of X may be dependent and may have different marginal distributions. Then $w_{(s)}$ feeds X into the nuclear-specific computer code to produce load and capacity vectors, typically by numerically solving a system of differential equations. The function $w_{(s)}$ is analytically intractable and computationally expensive to execute, motivating the use of Monte Carlo methods and VRTs.

Before explaining the implementation of LHS, we first describe how to employ SRS in the setting of Assumption 1 to obtain n_s i.i.d. outputs of load and capacity vectors for scenario s. Arrange i.i.d. $U[0,1)$ random numbers $U_{(s),i,j}$, $1 \le i \le n_s$, $1 \le j \le d_s$, into an $n_s \times d_s$ grid. Then applying $w_{(s)}$ to each grid row yields

$$\begin{array}{ccccccccc}
(L_{(s),1}, C_{(s),1}) & = & w_{(s)}(U_{(s),1,1}, & U_{(s),1,2}, & \ldots, & U_{(s),1,d_s}) & \sim & H_{(s)}, \\
(L_{(s),2}, C_{(s),2}) & = & w_{(s)}(U_{(s),2,1}, & U_{(s),2,2}, & \ldots, & U_{(s),2,d_s}) & \sim & H_{(s)}, \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(L_{(s),n_s}, C_{(s),n_s}) & = & w_{(s)}(U_{(s),n_s,1}, & U_{(s),n_s,2}, & \ldots, & U_{(s),n_s,d_s}) & \sim & H_{(s)}
\end{array} \tag{12}$$

by (11), where $L_{(s),i} = (L^{[1]}_{(s),i}, L^{[2]}_{(s),i}, \ldots, L^{[q]}_{(s),i})$ (resp., $C_{(s),i} = C^{[1]}_{(s),i}, C^{[2]}_{(s),i}, \ldots, C^{[q]}_{(s),i})$) is the vector of loads (resp., capacities) for the ith run, $i = 1,2,\ldots,n_s$. Also, $(L_{(s),i}, C_{(s),i})$, $i = 1,2,\ldots,n_s$, are n_s independent pairs by the independence of the rows of uniforms in (12), but $L_{(s),i}$ and $C_{(s),i}$ may be dependent. We can then use them to construct the estimator $\hat{\theta}_{(s),\mathrm{SS},\eta}(n)$ in (7). Moreover, assuming the uniform grids in (12) across scenarios $s = 1,2,\ldots,t$, are independent, we then apply (8) to obtain the asymptotic UCB in (10).

We now explain how to implement LHS to obtain a *dependent* sample of n_s pairs of load and capacity vectors for a scenario s. For each $s = 1,2,\ldots,t$, and each input dimension $j = 1,2,\ldots,d_s$, in (11), let $\pi_{(s),j} = (\pi_{(s),j}(1), \pi_{(s),j}(2), \ldots, \pi_{(s),j}(n_s))$ be a random permutation of $(1,2,\ldots,n_s)$; i.e., each of the $n_s!$ permutations of $(1,2,\ldots,n_s)$ is equally likely, and $\pi_{(s),j}(i)$ is the number to which i is mapped in permutation $\pi_{(s),j}$. Also, let $\pi_{(s),j}$, $j = 1,2,\ldots,d_s$, be independent random permutations, independent of the i.i.d. $U[0,1)$ random numbers $U_{(s),i,j}$, $1 \le i \le n_s$, $1 \le j \le d_s$. Next let $V_{(s),i,j} = [\pi_{(s),j}(i) - 1 + U_{(s),i,j}]/n_s$, which we arrange into an $n_s \times d_s$ grid. We then apply the function $w_{(s)}$ to each row of the grid to get

$$\begin{array}{ccccccc} (L_{(s),1}, C_{(s),1}) & = & w_{(s)}(V_{(s),1,1}, & V_{(s),1,2}, & \ldots, & V_{(s),1,d_s}), \\ (L_{(s),2}, C_{(s),2}) & = & w_{(s)}(V_{(s),2,1}, & V_{(s),2,2}, & \ldots, & V_{(s),2,d_s}), \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (L_{(s),n_s}, C_{(s),n_s}) & = & w_{(s)}(V_{(s),n_s,1}, & V_{(s),n_s,2}, & \ldots, & V_{(s),n_s,d_s}), \end{array} \tag{13}$$

where $L_{(s),i} = (L_{(s),i}^{[1]}, L_{(s),i}^{[2]}, \ldots, L_{(s),i}^{[q]})$ and $C_{(s),i} = (C_{(s),i}^{[1]}, C_{(s),i}^{[2]}, \ldots, C_{(s),i}^{[q]})$, $i = 1,2,\ldots,n_s$. It is easy to show that each $V_{(s),i,j} \sim U[0,1)$, with the d_s entries in each row i of (13) being independent. Thus, each $(L_{(s),i}, C_{(s),i}) \sim H_{(s)}$ by (11). But the n_s inputs in each column j of (13) are *dependent* as they share the same permutation $\pi_{(s),j}$, so $(L_{(s),i}, C_{(s),i})$, $i = 1,2,\ldots,n_s$, are *dependent*. Nevertheless, $\hat{\theta}_{(s),\mathrm{SS+LHS},\eta}(n) \equiv (1/n_s)\sum_{i=1}^{n_s} I(\cup_{k=1}^{q}\{L_{(s),i}^{[k]} \ge C_{(s),i}^{[k]}\})$ is an *unbiased* estimator of $\theta_{(s)}$ in (6); i.e., $E[\hat{\theta}_{(s),\mathrm{SS+LHS},\eta}(n)] = \theta_{(s)}$.

The fact that LHS outputs across rows are not independent complicates the form and estimation of the asymptotic variance of the estimator $\hat{\theta}_{(s),\mathrm{SS+LHS},\eta}(n)$ of $\theta_{(s)}$. To avoid this issue, we use *replicated LHS* (rLHS) (Iman 1981). Rather than creating one LHS grid of dependent uniforms with n_s rows for scenario s, rLHS generates $b \ge 2$ independent replications (e.g., $b = 10$) of LHS grids, each with $m_s = n_s/b$ rows. Thus, we obtain b independent samples, allowing us to estimate the failure probabilities, variance, and upper confidence bound as follows. For each replication $r = 1,2,\ldots,b$, let $\hat{\theta}_{(s)}^{\langle r\rangle}(n) = (1/m_s)\sum_{i=1}^{m_s} I(\cup_{k=1}^{q}\{L_{(s),i}^{\langle r\rangle[k]} \ge C_{(s),i}^{\langle r\rangle[k]}\})$, where $L_{(s),i}^{\langle r\rangle[k]}$ and $C_{(s),i}^{\langle r\rangle[k]}$ are the ith observation of the load and capacity of criterion $k = 1,2,\ldots,q$, for scenario $s = 1,2,\ldots,t$, in replication r. The estimator of the overall failure probability θ in (5) from replication r is $\hat{\theta}^{\langle r\rangle}(n) = \sum_{s=1}^{t} \lambda_s \hat{\theta}_{(s)}^{\langle r\rangle}(n)$. The final SS+rLHS estimator of θ from all b replications using SS allocation η across scenarios is $\hat{\theta}_{\mathrm{SS+rLHS},\eta,b}(n) = (1/b)\sum_{r=1}^{b} \hat{\theta}^{\langle r\rangle}(n)$.

To derive a UCB for θ when applying SS+rLHS with an overall sample size of n, let $S_b^2(n) = (1/(b-1))\sum_{r=1}^{b}\left[\hat{\theta}^{\langle r\rangle}(n) - \hat{\theta}_{\mathrm{SS+rLHS},\eta,b}(n)\right]^2$ be the sample variance of $\hat{\theta}^{\langle r\rangle}(n)$, $r = 1,2,\ldots,b$. Let $\tau_{b-1,\gamma}$ be the upper one-sided γ-level critical point of a Student-t random variable T_{b-1} with $b-1$ degrees of freedom; i.e., $\gamma = P(T_{b-1} \le \tau_{b-1,\gamma})$. A proof of the following result is in Alban et al. (2016).

Theorem 1 Under Assumption 1, an asymptotic γ-level UCB for θ when using SS+rLHS is $B_{\mathrm{SS+rLHS},\eta,b}(n) = \hat{\theta}_{\mathrm{SS+rLHS},\eta,b}(n) + \tau_{b-1,\gamma} S_b(n)/\sqrt{b}$; i.e., $\lim_{n\to\infty} P(\theta \le B_{\mathrm{SS+rLHS},\eta,b}(n)) = \gamma$ as in (3) for any fixed number $b \ge 2$ of replications for rLHS and any SS allocation η.

6 COMBINED SS, CONDITIONAL MONTE CARLO, AND LHS

Conditional Monte Carlo reduces variance by analytically integrating out some of the variability through a conditional expectation; see Section V.4 of Asmussen and Glynn (2007) for an overview of CMC. The method is based on the well-known (e.g., pp. 448 and 456 of Billingsley 1995) formulas

$$E[Y] = E[E[Y|Z]] \quad \text{and} \quad \mathrm{Var}[Y] = \mathrm{Var}[E[Y|Z]] + E[\mathrm{Var}[E|Z]] \ge \mathrm{Var}[E[Y|Z]]. \tag{14}$$

Applying the first relation in (14) to (6) with $Y = I(\cup_{k=1}^{q}\{L_{(s)}^{[k]} \geq C_{(s)}^{[k]}\})$ and $Z = L_{(s)}$ yields

$$\theta_{(s)} = E\left[E\left[I\left(\cup_{k=1}^{q}\left\{L_{(s)}^{[k]} \geq C_{(s)}^{[k]}\right\}\right)\middle| L_{(s)}\right]\right] = E\left[J_{(s)}(L_{(s)})\right], \tag{15}$$

where the SS+CMC response function $J_{(s)}(L_{(s)}) \equiv P(\cup_{k=1}^{q}\{C_{(s)}^{[k]} \leq L_{(s)}^{[k]}\}|L_{(s)})$ is a function of only the load vector $L_{(s)}$ as the capacity vector $C_{(s)}$ has been integrated out through the conditional expectation. Also, $J_{(s)}(L_{(s)})$ has no greater variance than $I(\cup_{k=1}^{q}\{L_{(s)}^{[k]} \geq C_{(s)}^{[k]}\})$ by the second relation in (14).

A key to being able to apply CMC in practice is the tractability of the conditional expectation $J_{(s)}(L_{(s)})$. To this end, we extend Assumption 1 through the following:

Assumption 2 For each scenario $s = 1, 2, \ldots, t$, there are deterministic functions $w_{(s),L} : \Re^{d_{s,L}} \to \Re^q$ and $w_{(s),C} : \Re^{d_{s,C}} \to \Re^q$ such that $d_{s,L} + d_{s,C} = d_s$ in Assumption 1, and the function $w_{(s)}$ in (11) satisfies

$$w_{(s)}(u_1, u_2, \ldots, u_{d_s}) = \left(w_{(s),L}(u_1, u_2, \ldots, u_{d_{s,L}}), w_{(s),C}(u_{d_{s,L}+1}, u_{d_{s,L}+2}, \ldots, u_{d_{s,L}+d_{s,C}})\right)$$

for every $(u_1, u_2, \ldots, u_{d_s}) \in [0,1)^{d_s}$.

Whereas Assumption 1 states that there is a single vector-valued function $w_{(s)}$ that generates both the loads and capacities for scenario s, Assumption 2 stipulates that $w_{(s)}$ splits into two functions operating on disjoint sets of inputs, where the first function $w_{(s),L}$ generates the loads and the second $w_{(s),C}$ generates the capacities. Hence, if $U_1, \ldots, U_{d_{s,L}}, U_{d_{s,L}+1}, \ldots, U_{d_{s,L}+d_{s,C}}$ are d_s i.i.d. $U[0,1)$ random variables, then

$$L_{(s)} = w_{(s),L}(U_1, U_2, \ldots, U_{d_{s,L}}) \sim F_{(s)} \quad \text{and} \quad C_{(s)} = w_{(s),C}(U_{d_{s,L}+1}, U_{d_{s,L}+2}, \ldots, U_{d_{s,L}+d_{s,C}}) \sim G_{(s)}, \tag{16}$$

where $L_{(s)} = (L_{(s)}^{[1]}, L_{(s)}^{[2]}, \ldots, L_{(s)}^{[q]})$ and $C_{(s)} = (C_{(s)}^{[1]}, C_{(s)}^{[2]}, \ldots, C_{(s)}^{[q]})$. Because $w_{(s),L}$ and $w_{(s),C}$ operate on disjoint sets of i.i.d. uniform inputs, we then see that $L_{(s)}$ and $C_{(s)}$ are *independent* under Assumption 2. For nuclear safety analyses, the independence is reasonable because the loads arise from how the hypothesized accident unfolds, whereas the capacities are determined by material properties and manufacturing variability.

We next give an expression to compute the response function $J_{(s)}(L_{(s)})$ in (15). For integers $1 \leq p \leq q$ and $1 \leq k_1 < k_2 < \cdots < k_p \leq q$, define $G_{(s)}^{[k_1,k_2,\ldots,k_p]}(x_1, x_2, \ldots, x_p) = P(\cap_{l=1}^{p}\{C_{(s)}^{[k_l]} \leq x_l\})$ for any real-valued constants $x_1, x_2, \ldots, x_p$, so $G_{(s)}^{[k_1,k_2,\ldots,k_p]}$ is the marginal joint CDF of capacities $C_{(s)}^{[k_l]}$, $l = 1, 2, \ldots, p$. We assume each $G_{(s)}^{[k_1,k_2,\ldots,k_p]}$ can be computed analytically or numerically. The independence of $L_{(s)}$ and $C_{(s)}$ implied by Assumption 2 ensures $P(\cap_{l=1}^{p}\{C_{(s)}^{[k_l]} \leq L_{(s)}^{[k_l]}\}|L_{(s)}) = G_{(s)}^{[k_1,k_2,\ldots,k_p]}(L_{(s)}^{[k_1]}, L_{(s)}^{[k_2]}, \ldots, L_{(s)}^{[k_p]})$. Thus, the inclusion-exclusion principle permits us to write the response function as

$$J_{(s)}(L_{(s)}) = \sum_{p=1}^{q}(-1)^{p+1}\sum_{1 \leq k_1 < k_2 < \cdots < k_p \leq q} G_{(s)}^{[k_1,k_2,\ldots,k_p]}(L_{(s)}^{[k_1]}, L_{(s)}^{[k_2]}, \ldots, L_{(s)}^{[k_p]}), \tag{17}$$

whose number of terms grows exponentially in the number q of criteria, but is manageable for small q.

To estimate $\theta_{(s)}$ via (15) and (17), we only need to generate the loads as the capacities have been analytically integrated out from $J_{(s)}(L_{(s)})$ and replaced by their marginal CDFs in (17). Thus, when employing SS, CMC, and LHS under Assumption 2, we generate an LHS grid of dependent uniforms $V_{(s),i,j}$ as in (13) but with only $d_{s,L}$ columns instead of d_s. We then apply the function $w_{(s),L}$ to each row to get

$$\begin{array}{lcllllll}
L_{(s),1} & = & w_{(s),L}(V_{(s),1,1}, & V_{(s),1,2}, & \ldots, & V_{(s),1,d_{s,L}}) & \sim & F_{(s)}, \\
L_{(s),2} & = & w_{(s),L}(V_{(s),2,1}, & V_{(s),2,2}, & \ldots, & V_{(s),2,d_{s,L}}) & \sim & F_{(s)}, \\
\vdots & \vdots & \quad\vdots & \quad\vdots & \ddots & \quad\vdots & & \\
L_{(s),n_s} & = & w_{(s),L}(V_{(s),n_s,1}, & V_{(s),n_s,2}, & \ldots, & V_{(s),n_s,d_{s,L}}) & \sim & F_{(s)},
\end{array} \tag{18}$$

where $L_{(s),i} = (L^{[1]}_{(s),i}, L^{[2]}_{(s),i}, \ldots, L^{[q]}_{(s),i})$, $i = 1,2,\ldots,n_s$. Finally, the outer expressions in (15) imply that an unbiased estimate of $\theta_{(s)}$ is $\hat{\theta}_{(s),\mathrm{SCL},\eta}(n) = (1/n_s)\sum_{i=1}^{n_s} J_{(s)}(L_{(s),i})$, where SCL denotes SS+CMC+LHS.

The dependence among $J_{(s)}(L_{(s),i})$, $i = 1,2,\ldots,n_s$, from LHS complicates the expression for and the construction of a consistent estimator of the asymptotic variance of the estimator $\hat{\theta}_{(s),\mathrm{SCL},\eta}(n)$. To avoid these issues, we instead apply rLHS. For each scenario $s = 1,2,\ldots,t$, we generate $b \geq 2$ independent LHS grids of dependent uniforms, each with $m_s = n_s/b$ rows and $d_{s,L}$ columns. For each replication $r = 1,2,\ldots,b$, let $\tilde{\theta}^{\langle r\rangle}_{(s)}(n) = (1/m_s)\sum_{i=1}^{m_s} J_{(s)}(L^{\langle r\rangle}_{(s),i})$, where $L^{\langle r\rangle}_{(s),i} = (L^{\langle r\rangle[1]}_{(s),i}, L^{\langle r\rangle[2]}_{(s),i}, \ldots, L^{\langle r\rangle[q]}_{(s),i})$ is the observation of the load vector from the ith row in replication r for scenario $s = 1,2,\ldots,t$. The estimator of the overall failure probability θ in (5) from replication r is $\tilde{\theta}^{\langle r\rangle}(n) = \sum_{s=1}^{t}\lambda_s\tilde{\theta}^{\langle r\rangle}_{(s)}(n)$. The final SS+CMC+rLHS estimator of θ from all b replications using SS allocation η across scenarios is $\hat{\theta}_{\mathrm{SCrL},\eta,b}(n) = (1/b)\sum_{r=1}^{b}\tilde{\theta}^{\langle r\rangle}(n)$, where SCrL is an abbreviation for SS+CMC+rLHS. To derive a UCB for θ when applying SS+CMC+rLHS with an overall sample size of n, let $\tilde{S}^2_b(n) = (1/(b-1))\sum_{r=1}^{b}\left[\tilde{\theta}^{\langle r\rangle}(n) - \hat{\theta}_{\mathrm{SCrL},\eta,b}(n)\right]^2$ be the sample variance of $\tilde{\theta}^{(r)}(n)$, $r = 1,2,\ldots,b$. Then we have the following result; see Alban et al. (2016) for a proof.

Theorem 2 Under Assumptions 1 and 2, an asymptotic γ-level UCB for θ when using SS+CMC+rLHS is $B_{\mathrm{SCrL},\eta,b}(n) = \hat{\theta}_{\mathrm{SCrL},\eta,b}(n) + \tau_{b-1,\gamma}\tilde{S}_b(n)/\sqrt{b}$; i.e., $\lim_{n\to\infty} P(\theta \leq B_{\mathrm{SCrL},\eta,b}(n)) = \gamma$ as in (3) for any fixed number $b \geq 2$ of replications for rLHS and any SS allocation η.

7 NUMERICAL RESULTS

We now present numerical results showing the benefits of combining the VRTs in this paper. Rather than using a computer code (Hess et al. 2009) as in an actual nuclear safety analyses (see Section 1), we instead work with a synthetic model of a hypothesized NPP station blackout having the event tree in Figure 1 with $t = 4$ scenarios. We consider $q = 3$ criteria: PCT, CWO, and MLO. For each scenario $s = 1,2,\ldots,t$, we assume that loads $L_{(s)}$ are independent of capacities $C_{(s)}$, as implied by Assumption 2, which we previously noted is reasonable in our context. The marginal load distributions vary across scenarios, but we keep the capacity distributions the same.

The initial RISMC studies in Sherry, Gabor, and Hess (2013) and Dube et al. (2014) consider only $q = 1$ criterion, PCT, and assume the CDF $G^{[1]}$ of the capacity $C^{[1]}$ for PCT is triangular with mode $c^{[1]} = 2200$ and support $[a^{[1]}, b^{[1]}] = [1800, 2600]$, which is the same for all scenarios s. The mode $c^{[1]}$ is the prescribed fixed limit of the NRC (see Section 1), and the support $[a^{[1]}, b^{[1]}]$ is symmetric around the mean with approximately 20% separation. We also adopt this structure for the marginal CDFs $G^{[2]}$ and $G^{[3]}$ of the CWO and MLO capacities $C^{[2]}$ and $C^{[3]}$, respectively. Thus, $G^{[2]}$ (resp., $G^{[3]}$) is a triangular distribution with mean $c^{[2]} = 1$ (resp., $c^{[3]} = 17$) and support $[a^{[2]}, b^{[2]}] = [0.8, 1.2]$ (resp., $[a^{[3]}, b^{[3]}] = [13.6, 20.4]$).

Nakayama (2015) considers a lognormal model for the load of the single criterion PCT, and we use the same for the PCT load in our problem with $q = 3$ criteria. A lognormal random variable for the PCT in scenario s can be obtained by exponentiating a $N(\mu_{(s)}, \sigma^2_{(s)})$ variable, and we set $\mu_{(s)} = 7.4 + 0.1s$ and $\sigma_{(s)} = 0.01 + 0.01s$ for $s = 1,2,3,4$. Let $F^{[1]}_{(s)}$ be the corresponding lognormal load CDF for scenario s.

Nutt and Wallis (2004) present PCT, CWO and MLO load data (sample sizes ≈ 180) output from an actual nuclear computer code. Based on the shapes of the data's histograms, we chose marginal Weibull CDFs for the CWO and MLO loads for each scenario s. For the CWO load in scenario s, the CDF is $F^{[2]}_{(s)}(x) = 1 - \exp\{-(x/\alpha^{[2]}_{(s)})^{\beta^{[2]}_{(s)}}\}$ for $x \geq 0$, where $\alpha^{[2]}_{(s)}$ and $\beta^{[2]}_{(s)}$ are the scale and shape parameters, respectively. We set $\alpha^{[2]}_{(s)} = 0.010 + 0.005s$ and $\beta^{[2]}_{(s)} = 0.90 + 0.02s$ for scenario $s = 1,2,3,4$, where the values for $s = 1$ are the maximum likelihood estimates (MLEs) from the Nutt and Wallis (2004) data. Applying the Kolmogorov-Smirnov (KS) and chi-squared goodness-of-fit tests to the selected CDF for $s = 1$ results in p-values of 0.22 and 0.61, respectively, so we do not reject the null hypothesis (at significance level 0.05) that the data are from our chosen distribution. The MLO load Weibull CDF $F^{[3]}_{(s)}$ has parameters

$\alpha_{(s)}^{[3]} = 0.35 + 0.3s$ and $\beta_{(s)}^{[3]} = 0.82 + 0.03s$ for each scenario s, where the values for $s = 1$ are the MLEs. The p-values for the KS and chi-squared tests for $s = 1$ are 0.82 and 0.08, so neither rejects our choice.

We have thus far specified only the marginal distributions of the loads and capacities, and we further need to identify the dependency structure among the variables. Even though the loads are independent of capacities, we still want dependence across criteria among the loads and among the capacities. We implemented the Gaussian copula (also called NORTA by Cario and Nelson 1997) to model this dependence. (In Alban et al. 2016, we also use a Student-t copula.) A *copula* (Nelsen 2006) is the joint CDF of a random vector in which each of its elements has a marginal $U[0,1)$ distribution. To describe the Gaussian copula, we start with a multivariate normal random vector $X = (X^{[1]}, X^{[2]}, \ldots, X^{[q]})$ with mean vector of all 0s and covariance matrix $\Sigma = (\Sigma^{[i,j]} : i, j = 1, 2, \ldots, q)$, where $\Sigma^{[i,j]} = \mathrm{Cov}(X^{[i]}, X^{[j]})$ and $\Sigma^{[i,i]} = 1$ for each i, so the elements of Σ are correlations. Each $X^{[k]}$ has the $N(0,1)$ CDF Φ, and we define the random vector

$$W = (W^{[1]}, W^{[2]}, \ldots, W^{[q]}) \text{ with each } W^{[k]} \equiv \Phi(X^{[k]}) \sim U[0,1). \tag{19}$$

The joint CDF of W is then a Gaussian copula with *input correlation matrix* Σ. If M_k is a specified marginal CDF, we can transform W into a random vector $Y = (Y^{[1]}, Y^{[2]}, \ldots, Y^{[q]})$ in which marginally each $Y^{[k]} \sim M_k$ by setting $Y^{[k]} = M_k^{-1}(W^{[k]})$. The Gaussian copula determines the dependence structure of Y.

Our models generate the $q = 3$ loads using a Gaussian copula, and independently generate the q capacities with its own Gaussian copula. As in (19), let $W_L = (W_L^{[1]}, W_L^{[2]}, W_L^{[3]})$ (resp., $W_C = (W_C^{[1]}, W_C^{[2]}, W_C^{[3]})$) be a vector with Gaussian copula having input correlation matrix $\Sigma_L = (\Sigma_L^{[i,j]} : i, j = 1, 2, 3)$ (resp., $\Sigma_C = (\Sigma_C^{[i,j]} : i, j = 1, 2, 3)$), where each entry in W_L and W_C has a marginal $U[0,1)$ CDF. We then transform W_L (resp., W_C) into loads (resp., capacities) with marginal CDFs $F_{(s)}^{[1]}, F_{(s)}^{[2]}, F_{(s)}^{[3]}$ (resp., $G^{[1]}, G^{[2]}, G^{[3]}$) via

$$(L_{(s)}^{[1]}, L_{(s)}^{[2]}, L_{(s)}^{[3]}) = ((F_{(s)}^{[1]})^{-1}(W_L^{[1]}), (F_{(s)}^{[2]})^{-1}(W_L^{[2]}), (F_{(s)}^{[3]})^{-1}(W_L^{[3]})), \tag{20}$$

$$(C^{[1]}, C^{[2]}, C^{[3]}) = ((G^{[1]})^{-1}(W_C^{[1]}), (G^{[2]})^{-1}(W_C^{[2]}), (G^{[3]})^{-1}(W_C^{[3]})). \tag{21}$$

We assume that W_L is independent of W_C, so loads are independent of capacities.

Nutt and Wallis (2004) estimate the correlations among the three criteria's loads from their data as $\rho_L^{[1,2]} = 0.85$, $\rho_L^{[1,3]} = 0.87$, and $\rho_L^{[2,3]} = 0.83$. However, these values are the *output* correlations of $L_{(s)}^{[1]}$, $L_{(s)}^{[2]}$, and $L_{(s)}^{[3]}$, and we need to specify the *input* correlations Σ_L of the multivariate normal X giving rise to the Gaussian copula. The output correlation is often a complicated and not-easily-computed function of the input correlations, and typically, it is not clear how to directly specify the input correlation to obtain the desired output correlation. Instead, we applied a search algorithm suggested by Cario and Nelson (1997) to obtain estimates of the load input correlations as $\Sigma_L^{[1,2]} = 0.92$, $\Sigma_L^{[1,3]} = 0.96$, and $\Sigma_L^{[2,3]} = 0.86$, which we used for all scenarios s. For the capacity input correlations, we set $\Sigma_C^{[1,2]} = \Sigma_C^{[1,3]} = \Sigma_C^{[2,3]} = 0.85$ for all s.

Finally, we specify the structure in Assumptions 1 and 2 used for implementing LHS and CMC. As Monte Carlo methods become attractive alternatives to numerical quadrature for computing multidimensional integrals when the input dimension is high, we artificially increase the number d_s of uniform inputs for the function $w_{(s)}$ in Assumption 2 as follows. For each scenario s, we define the input dimensions for the load and capacity functions $w_{(s),L}$ and $w_{(s),C}$ in (16) as $d_{s,L} = d_{s,L}^{[1]} + d_{s,L}^{[2]} + d_{s,L}^{[3]}$ and $d_{s,C} = 3$, respectively, with $d_{s,L}^{[1]} = d_{s,L}^{[2]} = d_{s,L}^{[3]} = 10$, so $d_s = d_{s,L} + 3$. Let $U_1, U_2, \ldots, U_{d_s}$ be d_s i.i.d. $U[0,1)$ random numbers, and set $D_{L,(s)}^{[1]} = \sum_{j=1}^{d_{s,L}^{[1]}} \Phi^{-1}(U_j)/\sqrt{d_{s,L}^{[1]}}$, $D_{L,(s)}^{[2]} = \sum_{j=d_{s,L}^{[1]}+1}^{d_{s,L}^{[1]}+d_{s,L}^{[2]}} \Phi^{-1}(U_j)/\sqrt{d_{s,L}^{[2]}}$, $D_{L,(s)}^{[3]} = \sum_{j=d_{s,L}^{[1]}+d_{s,L}^{[2]}+1}^{d_{s,L}} \Phi^{-1}(U_j)/\sqrt{d_{s,L}^{[3]}}$, and $D_{C,(s)}^{[k]} = \Phi^{-1}(U_{d_{s,L}+k})$ for $k = 1, 2, 3$. The vectors $D_{L,(s)} = (D_{L,(s)}^{[k]} : k = 1, 2, 3)$ and $D_{C,(s)} = (D_{C,(s)}^{[k]} : k = 1, 2, 3)$ have i.i.d. $N(0,1)$ entries. Let Γ_L (resp., Γ_C) be a Cholesky factor of the input correlation matrix Σ_L

(resp., Σ_C), i.e., $\Sigma_L = \Gamma_L \Gamma_L^\top$ and $\Sigma_C = \Gamma_C \Gamma_C^\top$, where Γ_L and Γ_C are lower triangular. Then, $X_{L,(s)} = \Gamma_L D_{L,(s)}$ and $X_{C,(s)} = \Gamma_C D_{C,(s)}$ are independent normal random vectors with mean vectors 0 and correlation matrices Σ_L and Σ_C, respectively. Finally we use (19), (20), and (21) with $X = X_{L,(s)}$ and $X = X_{C,(s)}$ to obtain an observation of the loads and capacities.

Table 1 presents numerical results from applying four different combinations of Monte Carlo methods to estimate the failure probability θ in (1): SS (Section 4), SS+CMC (not described in the current paper), SS+rLHS (Section 5), and SS+CMC+rLHS (Section 6). We varied the total sample size $n = 4^v \times 100$ for $v = 1,2,3,4$. We set the SS allocation with $\eta_s = 0.25$ for each scenario $s = 1,2,3,4$. For rLHS we used $b = 10$ independent replicates of LHS samples within a stratum, so for each scenario s, we generate an LHS sample of size $m_s = \eta_s n/b$ for each replicate. Methods that do not use replications to create an upper confidence bound (i.e., SS and SS+CMC) have $b = 1$ in Table 1. For the safety requirement (2), we set $\theta_0 = 0.05$ and $\gamma = 0.95$. Because of the analytic tractability of our model, we are able to numerically compute the true value of the failure probability as $\theta = 0.04648$, so $\theta < \theta_0$. The results in Table 1 are from running 10^4 independent experiments for each method and total sample size n. The column labeled "AHW" gives the average half-width across the 10^4 experiments, where the half-width is the difference between the UCB and the point estimate of θ. For a UCB $B(n)$ and overall sample size n, the *coverage* is the probability $P(\theta < B(n))$. As noted throughout the paper, each method's UCB satisfies (3), so the coverage converges to the nominal level $\gamma = 0.95$ as $n \to \infty$. But for fixed n, the coverage may differ from γ. We estimate the coverage as the fraction of the 10^4 experiments in which $\theta < B(n)$. The *probability of correct decision* (PCD) is estimated as the fraction of the 10^4 experiments in which the decision rule (4) correctly determined that $\theta < \theta_0$. The column "Sample Var" gives the sample variance of the point estimate of θ across the 10^4 experiments. For each particular method x, the last column presents the *variance-reduction factor* (VRF), which for a given overall sample size n is the ratio of the sample variance for SS over the sample variance for method x.

As seen in Table 1, the VRF shows that the combination SS+CMC+rLHS can have a much lower variance than SS. It also illustrates the synergistic effect of combining CMC and LHS: the VRF for SS+CMC+rLHS greatly exceeds the product of the VRFs for SS+CMC and SS+rLHS. To explain the synergy, it is known that LHS can substantially reduce variance when the response function is well approximated by an additive (or separable) function of the input variables; e.g., see p. 241 of Glasserman (2004). When CMC is not applied, the response function in (6) is an indicator, which may be poorly approximated by an additive function, so LHS without CMC may not reduce variance by much. But the conditioning in (15) when incorporating CMC leads to a smoother response function $J_{(s)}(L_{(s)})$, which can have a better additive approximation, so combining CMC and LHS can lead to additional variance reduction.

Across methods for fixed sample size n, Table 1 shows that as the VRF increases, so does the PCD, which demonstrates a benefit of the methods presented in this paper. Additionally, the PCD approaches 1 as n grows larger, as expected. The asymptotic validity of our UCBs is seen by the coverages approaching the nominal level $\gamma = 0.95$ as $n \to \infty$. SS+rLHS coverage converges most slowly, but when additionally combined with CMC, the coverage comes close to γ even at small sample sizes.

8 CONCLUDING REMARKS

Motivated by a recent safety-analysis framework for nuclear power plants, we presented efficient Monte Carlo methods to estimate a failure probability θ based on multiple criteria. We developed asymptotically valid upper confidence bounds for θ when applying combinations of SS, rLHS, and CMC, and numerical experiments show that the combination of all three can lead to substantial variance reduction compared to SS alone. We applied Gaussian copulas to model dependence among the loads and among the capacities across criteria. Alban et al. (2016) present additional numerical experiments with a Student-t copula and also with different parameters for the PCT load distribution. Changing the copula led to slightly different values for θ but sometimes significant changes in the performance of the Monte Carlo methods. The appropriate modeling of capacity dependencies deserves further study.

Table 1: Comparison of results for three different combination of methods.

Method	n	b	AHW	Coverage	PCD	Sample Var	VRF SS/x
SS	400	1	9.19E-03	0.861	0.346	3.94E-05	1.00
	1600	1	5.02E-03	0.884	0.402	9.99E-06	1.00
	6400	1	2.60E-03	0.920	0.698	2.57E-06	1.00
	25600	1	1.30E-03	0.935	0.993	6.37E-07	1.00
SS+CMC	400	1	2.56E-03	0.947	0.730	2.41E-06	16.32
	1600	1	1.30E-03	0.950	0.998	6.16E-07	16.22
	6400	1	6.38E-04	0.945	1.000	1.57E-07	16.44
	25600	1	3.21E-04	0.947	1.000	3.88E-08	16.40
SS+rLHS	400	10	7.40E-03	0.777	0.616	2.83E-05	1.39
	1600	10	4.43E-03	0.793	0.395	6.75E-06	1.48
	6400	10	2.20E-03	0.921	0.801	1.54E-06	1.67
	25600	10	1.08E-03	0.936	1.000	3.68E-07	1.73
SS+CMC+rLHS	400	10	8.59E-04	0.933	1.000	2.44E-07	161.39
	1600	10	3.45E-04	0.942	1.000	3.72E-08	268.13
	6400	10	1.62E-04	0.940	1.000	8.60E-09	299.20
	25600	10	7.91E-05	0.942	1.000	2.07E-09	307.71

Although we developed the methods in the context of NPP safety, they can also be employed in a variety of other application domains. For example, civil engineers design systems/structures with random loads and capacities that require the chance of failure to be small. Moreover, financial institutions are interested in estimating the probability that cash outflows (i.e., loads) do not exceed inflows (capacities).

ACKNOWLEDGMENTS

This work has been supported in part by the National Science Foundation under Grants No. CMMI-1200065, DMS-1331010, and CMMI-1537322. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. An NJIT Provost Undergraduate Summer Research Fellowship provided additional funding.

REFERENCES

Alban, A., H. Darji, A. Imamura, and M. K. Nakayama. 2016. "Efficient Monte Carlo Methods for Estimating a Failure Probability with Multiple Criteria". In preparation.

Asmussen, S., and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.

Avramidis, A. N., and J. R. Wilson. 1996. "Integrated Variance Reduction Strategies for Simulation". *Operations Research* 44:327–346.

Billingsley, P. 1995. *Probability and Measure*. Third ed. New York: John Wiley and Sons.

Cario, M. C., and B. L. Nelson. 1997. "Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix.". Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Dube, D. A., R. R. Sherry, J. R. Gabor, and S. M. Hess. 2014. "Application of risk informed safety margin characterization to extended power uprate analysis". *Reliability Engineering and System Safety* 129:19–28.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. New York: Springer.

Helton, J. C., and F. J. Davis. 2003. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems". *Reliability Engineering and System Safety* 81:23–69.

Hess, S., J. Gaertner, J. Gabor, L. Shanley, L. Enos-Sylla, S. Prasad, and S. Hollingsworth. 2009. "Framework for Risk-Informed Safety Margin Characterization". Technical Report 1019206, Electric Power Research Institute, Palo Alto, California.

Iman, R. L. 1981. "Statistical Methods for Including Uncertainties Associated With the Geologic Isolation of Radioactive Waste Which Allow for a Comparison With Licensing Criteria.". In *Proceedings of the Symposium on Uncertainties Associated with the Regulation of the Geologic Disposal of High-Level Radioactive Waste*, edited by D. C. Kocher, 145–157. Washington, DC: US Nuclear Regulatory Commission, Directorate of Technical Information and Document Control.

Lehmann, E. L. 1999. *Elements of Large-Sample Theory*. New York: Springer.

Nakayama, M. K. 2015. "Estimating a Failure Probability Using a Combination of Variance-Reduction Tecniques". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 621–632: IEEE.

Nelsen, R. B. 2006. *An Introduction to Copulas*. 2 ed. New York: Springer.

Nuclear Energy Agency Committee on the Safety of Nuclear Installations 2007. "Safety Margins Action Plan (SMAP) – Final Report". Technical Report NEA/CSNI/R(2007)9, Organization for Economic Cooperation and Development.

Nutt, W. T., and G. B. Wallis. 2004. "Evaluation of nuclear safety from the outputs of computer codes in the presence of uncertainties". *Reliability Engineering and System Safety* 83:57–77.

Pál, L., and M. Makai. 2013. "Remarks on the Statistical Aspects of the Safety Analysis". Technical report, arXiv. 1336v1 [physics.data-an] 6 Jun 2013.

Sherry, R. R., J. R. Gabor, and S. M. Hess. 2013. "Pilot Application of Risk Informed Safety Margin Characterization to a Total Loss of Feedwater Event". *Reliability Engineering and System Safety* 117:65–72.

U.S. Nuclear Regulatory Commission 2010. "Acceptance criteria for emergency core cooling systems for light-water nuclear power reactors". Title 10, Code of Federal Regulations Section 50.46 (10CFR50.46), U.S. Nuclear Regulatory Commission, Washington, DC.

U.S. Nuclear Regulatory Commission 2011. "Applying Statistics". U.S. Nuclear Regulatory Commission Report NUREG-1475, Rev 1, U.S. Nuclear Regulatory Commission, Washington, DC.

AUTHOR BIOGRAPHIES

ANDRES ALBAN is an undergraduate student of applied mathematics at the New Jersey Institute of Technology. His email address is aa655@njit.edu.

HARDIK A. DARJI is an undergraduate student of applied mathematics and mechanical engineering at the New Jersey Institute of Technology. His email address is had8@njit.edu.

ATSUKI IMAMURA is an undergraduate student of computer science at the New Jersey Institute of Technology. His email address is ai64@njit.edu.

MARVIN K. NAKAYAMA is a professor in the Department of Computer Science at the New Jersey Institute of Technology. He received his Ph.D. in operations research from Stanford University and a B.A. in mathematics-computer science from U.C. San Diego. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. A paper he co-authored received the Best Theoretical Paper Award for the 2014 Winter Simulation Conference. He served as the simulation area editor for the *INFORMS Journal on Computing* from 2007–2016, and is an associate editor for *ACM Transactions on Modeling and Computer Simulation*. His research interests include simulation, modeling, statistics, risk analysis, and energy. His email address is marvin@njit.edu.