Efficient Quantile Estimation When Applying Stratified Sampling and Conditional Monte Carlo, With Applications to Nuclear Safety

Marvin K. Nakayama

Dept. of Computer Science
New Jersey Institute of Technology
Newark, New Jersey 07102
Email: marvin@njit.edu

***Abstract*—We describe how to estimate a quantile when applying a combination of stratified sampling and conditional Monte Carlo, which are variance-reduction techniques for Monte Carlo simulations. We establish a central limit theorem for the resulting quantile estimator. We further prove that for any fixed stratification allocation, the asymptotic variance of the quantile estimator with a combination of stratified sampling and conditional Monte Carlo is no greater than that for stratified sampling alone. We explain how the methods may be used to efficiently perform a safety analysis of a nuclear power plant.**

Keywords–Monte Carlo; Variance Reduction; Risk Analysis; Value-at-Risk.

I. Introduction

For a given constant $0 < p < 1$, the p-quantile of a continuous random variable Y is the constant ξ such that p (resp., $1 - p$) of the mass of the distribution of Y lies to the left (resp., right) of ξ. An example is the median, which is the 0.5-quantile.

In many application areas, risk is measured by a p-quantile, with p close to 0 or 1. For example, in finance, a quantile is known as a *value-at-risk*, and there are banking regulations [1] that specify required cash reserves in terms of a 0.99-quantile of a loss distribution. In safety analyses of nuclear power plants (NPPs), the U.S. Nuclear Regulatory Commission (NRC) [2] requires that for a hypothesized event, such as a loss-of-coolant accident, the 0.95-quantile of the peak cladding temperature must lie below a given threshold.

When the random variable Y is the output of a complicated stochastic model, analytically computing a quantile of Y typically presents intractable challenges, so Monte Carlo simulation is instead often applied [3]. Quantile estimation via *simple random sampling* (SRS) has been well-studied; see Sections 2.3–2.6 of [4]. But SRS can produce quantile estimators with large statistical error, motivating the use of *variance-reduction techniques* (VRTs) to obtain more statistically efficient estimators; see Chapter V of [5] and Chapter 4 of [6] for overviews of VRTs to estimate a mean. Quantile estimation has also employed VRTs, including importance sampling (IS) [7][8][9][10], control variates (CV) [11][12][10], Latin hypercube sampling (LHS) [13][14], stratified sampling (SS) [8][15][10], and conditional Monte Carlo (CMC) [16]. The use of VRTs can be especially important when each simulation run takes substantial time to execute, limiting the sample size that can be obtained.

In this paper, we consider applying a combination of stratified sampling and conditional Monte Carlo, which we denote by SS+CMC, to estimate a quantile. We give a central limit theorem for the SS+CMC quantile estimator. Moreover, we prove that the asymptotic variance of the SS+CMC quantile estimator is no larger than that of the corresponding quantile estimator with SS alone. Thus, SS+CMC is guaranteed to do at least as well as SS for quantile estimation.

Stratified sampling plays a fundamental role in the so-called *risk-informed safety-margin characterization* (RISMC) for nuclear power plants [17][18]. Developed by an international effort of the Nuclear Energy Agency, RISMC analyzes a hypothesized accident of an NPP through Monte Carlo simulation with a detailed computer code. The computer code takes as input a random vector with specified joint distribution, where the random inputs may specify the timing, size, and location of events during the postulated accident. The progression of the accident is also modeled through an *event tree*, consisting of intermediate events that determine how the accident evolves, e.g., whether or not a safety relief valve is stuck open. The intermediate events have known probabilities of occurring, and a path through the event tree partitions the sample space into *scenarios*. The probability of each scenario is known, but the distribution of the output variable Y for a scenario is not known, although we can generate observations from the distribution by simulation with the computer code. The framework fits exactly into applying stratified sampling by using the scenarios as strata. Further incorporating CMC leads to additional improvements in statistical efficiency. This is critical because each code run entails numerically solving differential equations, which is computationally expensive.

The rest of the paper unfolds as follows. Section II provides a list of acronyms used in the paper. Section III reviews how to apply SRS for quantile estimation. Section IV describes previous work on estimating a quantile via stratified sampling. In Section V, we combine SS with conditional Monte Carlo. We provide concluding remarks in Section VI. Throughout the paper, we give details on how the methods can be applied to perform a RISMC safety analysis of a nuclear power plant.

II. List of Acronyms

CDF	cumulative distribution function
CLT	central limit theorem
CMC	conditional Monte Carlo
CV	control variates

EDF empirical distribution function
IS importance sampling
LHS Latin hypercube sampling
NPP nuclear power plant
NRC Nuclear Regulatory Commission
PCT peak cladding temperature
RISMC risk-informed safety-margin characterization
SBO station blackout
SRS simple random sampling
SS stratified sampling
VRT variance-reduction technique

III. Background and Simple Random Sampling

Let Y be a real-valued random variable with cumulative distribution function (CDF) F, i.e., $F(y) = P(Y \leq y)$. For a fixed real number p with $0 < p < 1$, we define $\xi \equiv F^{-1}(p) \equiv \inf\{y : F(y) \geq p\}$ as the p-quantile of F (or equivalently, of Y); see Fig. 1. We assume that F is not analytically nor numerically tractable, but we have a computer code that can generate independent and identically distributed (i.i.d.) observations from F. The goal is to estimate ξ via Monte Carlo simulation. The typical approach, and the one we will follow, first estimates the CDF using simulation, and then inverts it to obtain a quantile estimator. Throughout the paper, we will use the following example to motivate and explain the different methods we consider.

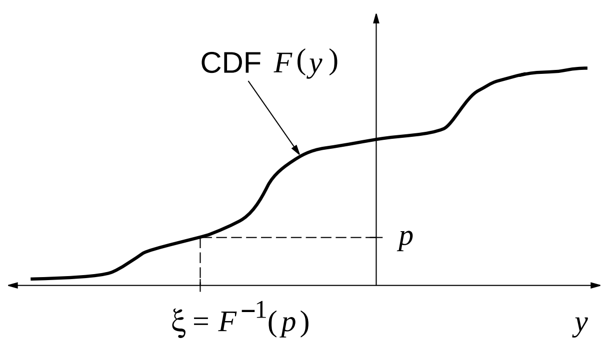

Figure 1. CDF F and p-quantile ξ.

Example 1. Consider a safety analysis of a nuclear power plant, in which a detailed computer code is used to model the progression of a hypothesized event, such as a loss-of-coolant accident or a station blackout. The computer code is run with random inputs having specified distributions, and the code outputs a *load* L representing the peak cladding temperature (PCT). The NRC [2] currently requires that the 0.95-quantile of L lies below a *fixed capacity* $C = 2200°$F. But the recent RISMC formulation [17][18] models the capacity C as a *random variable* to account for important changes in NPPs, e.g., aging components, extended operating licenses, and power uprates (i.e., operating an NPP at a higher level to produce more electricity). The papers [17][18] assume that the capacity C (in °F) has a triangular$(1800, 2200, 2600)$ distribution, and the computer code also generates an observation of C each time it is run. The RISMC problem requires that the probability that the load exceeds capacity is small, i.e., $P(L \geq C) \leq \alpha$ for some specified small α, say, $\alpha = 0.05$. We can formulate the requirement in terms of a quantile by letting $Y = C - L$, and stipulating that the α-quantile of Y is nonnegative, i.e., $\xi \geq 0$.

We start by describing how to use simple random sampling to estimate ξ; see Section 2.3 of [4] for an overview. We first generate a sample of n i.i.d. observations $Y_1, Y_2, \ldots, Y_n$ from F. Then we estimate the CDF F via the *empirical distribution function* (EDF) $\hat{F}_n$ defined by

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq y),$$

where $I(\cdot)$ is the indicator function, which takes on the value 1 (resp., 0) when its argument is true (resp., false). Because the true p-quantile is $\xi = F^{-1}(p)$, this suggests estimating it by

$$\hat{\xi}_{\text{SRS}}(n) = \hat{F}_n^{-1}(p), \tag{1}$$

which we call the ***SRS*** ***p-quantile estimator***. The SRS quantile estimator can be refined through interpolation [13] or smoothing techniques [19], but for simplicity, we only consider $\hat{\xi}_{\text{SRS}}(n)$.

We can equivalently compute $\hat{\xi}_{\text{SRS}}(n)$ via *order statistics*. Let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ be the order statistics of the sample $Y_1, Y_2, \ldots, Y_n$. Then $\hat{\xi}_{\text{SRS}}(n) = Y_{(\lceil np \rceil)}$, where $\lceil \cdot \rceil$ is the ceiling function; see Fig. 2.

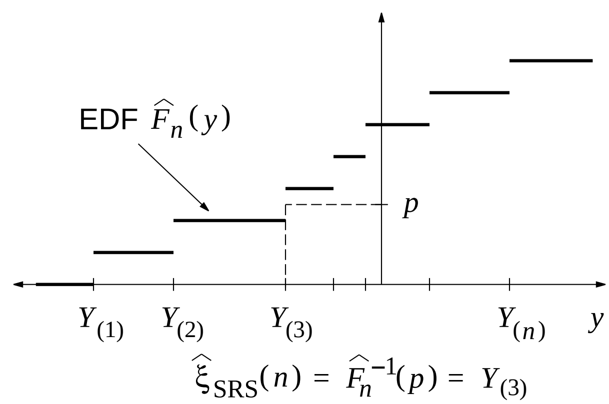

Figure 2. EDF $\hat{F}_n$ and the SRS p-quantile estimator $\hat{\xi}_{\text{SRS}}(n)$.

The SRS quantile estimator $\hat{\xi}_{\text{SRS}}(n)$ satisfies a central limit theorem (CLT), for which we give the following heuristic derivation. Let f denote the derivative (when it exists) of the CDF F, and suppose that $f(\xi) > 0$. For large n, we have that $\hat{F}_n \approx F$, so it is plausible that $\hat{\xi}_{\text{SRS}}(n) = \hat{F}_n^{-1}(p) \approx F^{-1}(p) = \xi$. Consequently,

$$\begin{aligned} F(\xi) &\approx F(\hat{\xi}_{\text{SRS}}(n)) \approx F(\xi) + f(\xi)[\hat{\xi}_{\text{SRS}}(n) - \xi] \\ &\approx \hat{F}_n(\xi) + f(\xi)[\hat{\xi}_{\text{SRS}}(n) - \xi], \end{aligned}$$

where the second step uses a Taylor approximation, and the last step follows because $\hat{F}_n \approx F$. Rearranging terms and scaling by $\sqrt{n}$ then yields

$$\sqrt{n}[\hat{\xi}_{\text{SRS}}(n) - \xi] \approx \frac{\sqrt{n}}{f(\xi)}[F(\xi) - \hat{F}_n(\xi)]. \tag{2}$$

 ISBN: 978-1-61208-501-2

The ordinary CLT (e.g., Theorem 1.9.1A of [4]) ensures that

$$\sqrt{n}[F(\xi) - \hat{F}_n(\xi)] \Rightarrow N(0, \psi^2_{\text{SRS}}) \tag{3}$$

as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution (see Section 1.2.4 of [4]),

$$\psi^2_{\text{SRS}} = \text{Var}[I(Y \le \xi)] = p(1-p), \tag{4}$$

and $N(a, b^2)$ represents a normal random variable with mean a and variance b^2. Finally, dividing the left side of (3) by $f(\xi)$ gives the right side of (2), suggesting that $\hat{\xi}_{\text{SRS}}(n)$ obeys the CLT

$$\sqrt{n}[\hat{\xi}_{\text{SRS}}(n) - \xi] \Rightarrow N(0, \kappa^2_{\text{SRS}}) \tag{5}$$

as $n \to \infty$, where

$$\kappa^2_{\text{SRS}} = \eta^2 \psi^2_{\text{SRS}} \tag{6}$$

is the *asymptotic variance* in the CLT, and

$$\eta = \frac{1}{f(\xi)} \tag{7}$$

is known as the *quantile density*. For a rigorous proof of the CLT in (5), see, e.g., p. 77 of [4].

IV. Stratified Sampling

Stratified sampling partitions the sample space into strata, and then allocates a fixed proportion of the overall sample size to each stratum. Section 4.3 of [6] provides an overview of SS to estimate a mean, and [15] considers quantile estimation combining SS with CV. Also, [8][10] combine SS with IS to estimate a quantile.

Suppose there is an auxiliary random variable Z, which could be generated in the process of generating the output variable Y, and we will use Z as a *stratification variable*. One possibility is $Z = Y$. Another is $Z = h(X)$ when the output variable $Y = v(X)$, where h and v are real-valued functions and X is some multidimensional random variable with known joint distribution; here, the function h may be more analytically tractable than v.

We partition the support R of Z into $R = \cup_{s=1}^t R_s$ for some fixed $t \ge 1$, with $R_s \cap R_{s'} = \emptyset$ for $s \ne s'$. Assume that we know the value of $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)$, where $\lambda_s = P(Z \in R_s)$ for $s = 1, 2, \ldots, t$. We call each R_s (or s) a *stratum*, which is also known as a *scenario*. Thus, for each $y \in \Re$, the CDF F of Y satisfies

$$\begin{aligned} F(y) = P(Y \le y) &= \sum_{s=1}^t P(Z \in R_s) P(Y \le y | Z \in R_s) \\ &= \sum_{s=1}^t \lambda_s F_{[s]}(y) \end{aligned} \tag{8}$$

by the law of total probability, where

$$F_{[s]}(y) = P(Y \le y | Z \in R_s) \tag{9}$$

is the conditional CDF of Y given $Z \in R_s$. In (8), the λ_s are known, but not the $F_{[s]}$, which we will estimate via Monte Carlo. Define a random variable $Y_{[s]} \sim F_{[s]}$, i.e., $Y_{[s]}$ has the conditional distribution of Y given $Z \in R_s$. We thus estimate $F_{[s]}$ by generating observations of $Y_{[s]}$, which we assume can be done for each stratum s, and using an empirical distribution.

Example 1 (cont). *Event trees* play an important role in a RISMC study, and Fig. 3 depicts an event tree from [17] of a hypothesized station blackout (SBO) at a nuclear power plant. The intermediate events E_1, E_2, E_3, which have known

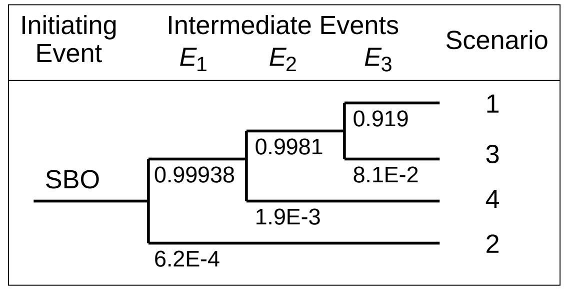

Figure 3. Event tree of a hypothesized station blackout at a nuclear power plant.

branching probabilities, as shown, determine how the accident progresses. For example, the lower (resp., upper) branch of E_2 represents the event that a safety relief valve is stuck open (resp., closes properly), which occurs with probability 0.0019 (resp., 0.9981). Paths from left to right through the event tree partition the state space into scenarios, and let Z denote a random chosen scenario. The support of Z is the set $R = \{1, 2, 3, 4\}$, and we can partition R into $t = 4$ strata $R_s = \{s\}$, $s = 1, 2, 3, 4$. We compute the probability λ_s of each scenario by multiplying the branching probabilities along its path, e.g., $\lambda_4 = 0.99938 \times 0.0019$. Each scenario s has a computer code that generates an observation of a load $L_{[s]}$ and a capacity $C_{[s]}$. Thus, we define the output $Y_{[s]} \sim F_{[s]}$ as $Y_{[s]} = C_{[s]} - L_{[s]}$ for scenario s.

To apply SS with an overall sample size n to estimate ξ, we allocate a fraction γ_s to stratum s, where $0 < \gamma_s < 1$ and $\sum_{s=1}^t \gamma_s = 1$. One possibility is to take $\gamma_s = \lambda_s$ for each s, but we also allow other allocations. Let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_t)$, and $n_s = \gamma_s n$ be the sample size for stratum s, where we assume that n_s is integer-valued; if not, we set $n_s = \lfloor \gamma_s n \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. Let $Y_{[s],1}, Y_{[s],2}, \ldots, Y_{[s],n_s}$ be a sample of n_s i.i.d. observations of $Y_{[s]}$. Then, we can estimate $F_{[s]}$ via

$$\hat{F}_{[s],n,\gamma}(y) = \frac{1}{n_s} \sum_{i=1}^{n_s} I(Y_{[s],i} \le y)$$

for each y. By (8), we then obtain the SS estimator $\tilde{F}_{n,\gamma}$ of the CDF F as

$$\tilde{F}_{n,\gamma}(y) = \sum_{s=1}^t \lambda_s \hat{F}_{[s],n,\gamma}(y).$$

The *SS quantile estimator* is then $\hat{\xi}_{\text{SS},\gamma}(n) = \tilde{F}^{-1}_{n,\gamma}(p)$. When there is only $t = 1$ stratum for SS with $\lambda_1 = \gamma_1 = 1$, the SS quantile estimator $\hat{\xi}_{\text{SS},\gamma}(n)$ reduces to the SRS quantile estimator $\hat{\xi}_{\text{SRS}}(n)$.

The SS quantile estimator $\hat{\xi}_{\text{SS},\gamma}(n)$ satisfies a CLT

$$\sqrt{n}[\hat{\xi}_{\text{SS},\gamma}(n) - \xi] \Rightarrow N(0, \kappa^2_{\text{SS},\gamma}) \tag{10}$$

as $n \to \infty$, where the asymptotic variance is

$$\kappa^2_{\text{SS},\gamma} = \eta^2 \psi^2_{\text{SS},\gamma}, \tag{11}$$

η is the quantile density in (7),

$$\psi^2_{\mathrm{SS},\gamma} = \sum_{s=1}^{t} \frac{\lambda_s^2}{\gamma_s} \zeta^2_{\mathrm{SS},[s]}, \tag{12}$$

$$\zeta^2_{\mathrm{SS},[s]} = \mathrm{Var}[I(Y_{[s]} \leq \xi)] = F_{[s]}(\xi)[1 - F_{[s]}(\xi)]; \tag{13}$$

see [15][8][10]. (The last two papers consider the combination of importance sampling and SS for quantile estimation, but SRS is a special case of IS, so they cover the setting of SS-alone.)

The SS asymptotic variance $\kappa^2_{\mathrm{SS},\gamma}$ in (11) is the product of two terms. The first, η^2, is the same as in the SRS asymptotic variance κ^2_{SRS} in (6), and the value of η^2 is unaffected by the particular Monte Carlo method employed to estimate ξ. But the second factor $\psi^2_{\mathrm{SS},\gamma}$ does depend on how ξ is estimated. The choice of the stratification allocation γ_s, $s = 1, 2, \ldots, t$, also affects the asymptotic variance $\kappa^2_{\mathrm{SS},\gamma}$ in the CLT (10) through $\psi^2_{\mathrm{SS},\gamma}$ in (12).

One possible choice for γ is the *proportional allocation*, in which $\gamma = \lambda$. As shown on p. 217 of [6], the proportional allocation for any choice of $t \geq 2$ strata $R_1, R_2, \ldots, R_t$, is guaranteed to reduce variance compared to SRS. To see why, let S be a discrete random variable such that $S = s$ if and only if $Z \in R_s$, $s = 1, 2, \ldots, t$, so $P(S = s) = \lambda_s$. (In Example 1 we have $S = Z$.) Thus, because $\zeta^2_{\mathrm{SS},[s]} = \mathrm{Var}[I(Y \leq \xi)|S = s]$, it follows that when $\gamma = \lambda$, we have that

$$\begin{aligned}\psi^2_{\mathrm{SS},\lambda} &= \sum_{s=1}^{t} \lambda_s \zeta^2_{\mathrm{SS},[s]} = E[\mathrm{Var}[I(Y \leq \xi)|S]] \\ &\leq E[\mathrm{Var}[I(Y \leq \xi)|S]] + \mathrm{Var}[E[I(Y \leq \xi)|S]] \\ &= \mathrm{Var}[I(Y \leq \xi)] = \psi^2_{\mathrm{SRS}},\end{aligned}$$

where the inequality holds because of the nonnegativity of a variance, the next step follows from a variance decomposition, and the last equality holds by (4). Hence, (6) and (11) imply the proportional allocation for SS leads to no larger asymptotic variance for the quantile estimator than SRS; also see [15].

For a given set of t strata $R_1, R_2, \ldots, R_t$, the optimal allocation γ that minimizes the asymptotic variance $\kappa^2_{\mathrm{SS},\gamma}$ of the SS quantile estimator is $\gamma^* = (\gamma^*_1, \gamma^*_2, \ldots, \gamma^*_t)$ with

$$\gamma^*_s = \frac{\lambda_s \zeta_{\mathrm{SS},[s]}}{\sum_{s'=1}^{t} \lambda_{s'} \zeta_{\mathrm{SS},[s']}}, \quad s = 1, 2, \ldots, t;$$

see, e.g., [15] and p. 217 of [6]. The allocation γ^* typically cannot be implemented directly in practice because $\zeta_{\mathrm{SS},[s]}$ and $F_{[s]}(\xi)$, $s = 1, 2, \ldots, t$, are unknown. The paper [15] provides an adaptive two-stage approach to asymptotically achieve the minimal SS asymptotic variance, where the first stage estimates the optimal γ^*, which is then used for the sampling in the second stage.

V. Combining SS With Conditional Monte Carlo

Conditional Monte Carlo, which is also known as the conditional-expectation method, reduces variance by analytically integrating out some of the variability; see Section V.4 of [5] for an overview of applying CMC to estimate a mean. Recall that for SS, we assumed that Z was a stratification variable with strata R_s for $s = 1, 2, \ldots, t$. Now, we assume that for each s, we have another auxiliary random variable $X_{[s]}$. We can then write the (conditional) CDF $F_{[s]}$ in (9) of $Y_{[s]}$ as

$$F_{[s]}(y) = P(Y_{[s]} \leq y) = E[P(Y_{[s]} \leq y | X_{[s]})] \tag{14}$$

by conditioning on $X_{[s]}$. Thus, assuming that

$$\begin{aligned} q_{[s]}(x, y) &\equiv P(Y_{[s]} \leq y | X_{[s]} = x) &(15)\\ &= E[I(Y_{[s]} \leq y) | X_{[s]} = x] &(16)\end{aligned}$$

can be computed, analytically or numerically, then (14) and (15) suggest that we can estimate $F_{[s]}(y)$ by averaging copies of $q_{[s]}(X_{[s]}, y)$, which we note is only a function of the conditioning variable $X_{[s]}$ and y as $Y_{[s]}$ has been integrated out through the conditional probability.

Example 1 (cont). The initial RISMC studies [17][18] have that the load $L_{[s]}$ and the capacity $C_{[s]}$ are independent random variables, which we will also assume. The independence assumption is reasonable from a modeling standpoint because the load is determined by how the hypothesized accident progresses, whereas the capacity depends on material properties of the components. Let $G_{[s]}$ denote the marginal CDF of the capacity $C_{[s]}$ in scenario s, i.e., $G_{[s]}(z) = P(C_{[s]} \leq z)$. As noted before, [17][18] assume that $G_{[s]}$ is a triangular distribution; the papers actually further assume that $G_{[s]}$ is the same for all scenarios s, but we do not require that here. For each scenario s, take the conditioning variable as $X_{[s]} = L_{[s]}$, and because the output is $Y_{[s]} = C_{[s]} - L_{[s]}$, we can write (15) as

$$\begin{aligned} q_{[s]}(x, y) &= P(C_{[s]} - L_{[s]} \leq y | L_{[s]} = x) \\ &= P(C_{[s]} \leq L_{[s]} + y | L_{[s]} = x) = G_{[s]}(x + y)\end{aligned}$$

by the independence of $L_{[s]}$ and $C_{[s]}$. In this case, $q_{[s]}(x, y)$ is only a function of the observed load $L_{[s]} = x$ and y because the random capacity $C_{[s]}$ has been integrated out, replaced by its marginal CDF $G_{[s]}$. When $G_{[s]}$ is a triangular CDF, as in [17][18], the function $q_{[s]}$ can be easily computed, as we previously required.

To implement the combination SS+CMC to estimate ξ, let $X_{[s],i}$, $i = 1, 2, \ldots, n_s$, be i.i.d. copies of $X_{[s]}$, where $n_s = \gamma_s n$ as before with SS allocation γ_s. Then, as suggested by (14) and (15), our CMC estimator of the CDF $F_{[s]}$ is given by

$$\check{F}_{[s],n,\gamma}(y) = \frac{1}{n_s} \sum_{i=1}^{n_s} q_{[s]}(X_{[s],i}, y).$$

Then, as in (8), we combine the $\check{F}_{[s],n}$, $s = 1, 2, \ldots, t$, to obtain the SS+CMC estimator of the CDF F of Y as

$$\check{F}_{n,\gamma}(y) = \sum_{s=1}^{t} \lambda_s \check{F}_{[s],n,\gamma}(y).$$

We finally obtain the SS+CMC p-quantile estimator as $\hat{\xi}_{\mathrm{SS+CMC},\gamma}(n) = \check{F}^{-1}_{n,\gamma}(p)$, which satisfies the following result.

Theorem 1. *If $f(\xi) > 0$, then*

$$\sqrt{n}[\hat{\xi}_{SS+CMC,\gamma}(n) - \xi] \Rightarrow N(0, \kappa^2_{SS+CMC,\gamma}) \tag{17}$$

as $n \to \infty$ for any SS allocation γ, where

$$\kappa^2_{SS+CMC,\gamma} = \eta^2 \psi^2_{SS+CMC,\gamma}, \quad (18)$$

$$\psi^2_{SS+CMC,\gamma} = \sum_{s=1}^{t} \frac{\lambda_s^2}{\gamma_s} \zeta^2_{SS+CMC,[s]}, \quad (19)$$

$$\zeta^2_{SS+CMC,[s]} = Var[q_{[s]}(X_{[s]}, \xi)], \quad (20)$$

and η is the quantile density in (7)*. Moreover, when SS and SS+CMC use the same stratification allocation γ, we have that*

$$\kappa^2_{SS+CMC,\gamma} \le \kappa^2_{SS,\gamma}, \quad (21)$$

where $\kappa^2_{SS,\gamma}$ in (11) *is the asymptotic variance in the CLT* (10) *for the SS quantile estimator.*

Proof: By applying ideas from the proofs in [10], we can formally show that the SS+CMC quantile estimator satisfies a Bahadur representation [20], which then implies the CLT in (17). To establish (21), we apply a variance decomposition to (13) to obtain

$$\begin{aligned} \zeta^2_{\text{SS},[s]} &= \text{Var}[I(Y_{[s]} \le \xi)] \\ &= \text{Var}[E[I(Y_{[s]} \le \xi)|X_{[s]}]] + E[\text{Var}[I(Y_{[s]} \le \xi)|X_{[s]}]] \\ &\ge \text{Var}[E[I(Y_{[s]} \le \xi)|X_{[s]}]] = \text{Var}[q_{[s]}(X_{[s]}, \xi)] \\ &= \zeta^2_{\text{SS+CMC},[s]} \end{aligned}$$

by (16) and (20). Thus, (19) and (12) imply that $\psi^2_{\text{SS+CMC},\gamma} \le \psi^2_{\text{SS},\gamma}$, from which (21) follows by (11) and (18). ∎

VI. Conclusion and Future Work

We described how to estimate a quantile when applying a combination SS+CMC of stratified sampling and conditional Monte Carlo. We provided a central limit theorem for the SS+CMC quantile estimator. We further proved that the SS+CMC quantile estimator has asymptotic variance that is no greater than that of the SS quantile estimator, when both approaches use the same stratification allocation. We also explained how SS+CMC can be employed to efficiently perform a risk-informed safety-margin characterization of a nuclear power plant.

A direction for future work is to develop confidence intervals for ξ when applying SS+CMC. One approach is to use a finite difference to consistently estimate the asymptotic variance $\kappa^2_{\text{SS+CMC},\gamma}$ in (18) in the CLT (17), as is done in [10] for other variance-reduction techniques. Another possibility applies *sectioning*, an approach that is closely related to batching (also known as subsampling) and was originally proposed in Section III.5a of [5] for SRS; [21] extends sectioning to IS and CV.

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 author and do not necessarily reflect the views of the National Science Foundation.

References

[1] Basel Committee on Banking Supervision, "Basel II: International convergence of capital measurement and capital standards: a revised framework," tech. rep., Bank for International Settlements, Basel, Switzerland, 2004.

[2] U.S. Nuclear Regulatory Commission, "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, 2010.

[3] L. J. Hong, Z. Hu, and G. Liu, "Monte Carlo methods for value-at-risk and conditional value-at-risk: A review," *ACM Trans. Mod. Comp. Sim.*, vol. 24, p. Article 22 (37 pages), 2014.

[4] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons, 1980.

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

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

[7] P. W. Glynn, "Importance sampling for Monte Carlo estimation of quantiles," in *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*, pp. 180–185, Publishing House of St. Petersburg Univ., St. Petersburg, Russia, 1996.

[8] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Variance reduction techniques for estimating value-at-risk," *Management Science*, vol. 46, pp. 1349–1364, 2000.

[9] L. Sun and L. J. Hong, "Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk," *Operations Research Letters*, vol. 38, pp. 246–251, 2010.

[10] F. Chu and M. K. Nakayama, "Confidence intervals for quantiles when applying variance-reduction techniques," *ACM Transactions On Modeling and Computer Simulation*, vol. 36, pp. Article 7 (25 pages plus 12–page online–only appendix), 2012.

[11] J. C. Hsu and B. L. Nelson, "Control variates for quantile estimation," *Management Science*, vol. 36, pp. 835–851, 1990.

[12] T. C. Hesterberg and B. L. Nelson, "Control variates for probability and quantile estimation," *Management Science*, vol. 44, pp. 1295–1312, 1998.

[13] A. N. Avramidis and J. R. Wilson, "Correlation-induction techniques for estimating quantiles in simulation," *Operations Research*, vol. 46, pp. 574–591, 1998.

[14] H. Dong and M. K. Nakayama, "Constructing confidence intervals for a quantile using batching and sectioning when applying Latin hypercube sampling," in *Proceedings of the 2014 Winter Simulation Conference* (A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.), pp. 640–651, Institute of Electrical and Electronics Engineers, 2014.

[15] C. Cannamela, J. Garnier, and B. Iooss, "Controlled stratification for quantile estimation," *Annals of Applied Statistics*, vol. 2, no. 4, pp. 1554–1580, 2008.

[16] M. K. Nakayama, "Quantile estimation when applying conditional Monte Carlo," in *SIMULTECH 2014 Proceedings*, pp. 280–285, 2014.

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

[18] R. R. Sherry, J. R. Gabor, and S. M. Hess, "Pilot application of risk informed safety margin characterization to a total loss of feedwater event," *Reliability Engineering and System Safety*, vol. 117, pp. 65–72, 2013.

[19] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London: Chapman and Hall, 1995.

[20] R. R. Bahadur, "A note on quantiles in large samples," *Annals of Mathematical Statistics*, vol. 37, pp. 577–580, 1966.

[21] M. K. Nakayama, "Confidence intervals using sectioning for quantiles when applying variance-reduction techniques," *ACM Transactions on Modeling and Computer Simulation*, vol. 24, p. Article 19, 2014.