Asymptotically Valid Confidence Intervals for Quantiles and Values-at-Risk When Applying Latin Hypercube Sampling

Marvin K. Nakayama
Computer Science Department
New Jersey Institute of Technology
Newark, New Jersey, 07102, USA
marvin@njit.edu

***Abstract*—Quantiles, which are also known as values-at-risk in finance, are often used as risk measures. Latin hypercube sampling (LHS) is a variance-reduction technique (VRT) that induces correlation among the generated samples in such a way as to increase efficiency under certain conditions; it can be thought of as an extension of stratified sampling in multiple dimensions. This paper develops asymptotically valid confidence intervals for quantiles that are estimated via simulation using LHS.**

***Keywords*-quantile; value-at-risk; Latin hypercube sampling; variance reduction; confidence interval.**

I. Introduction

Complex stochastic systems arise in many application areas, such as supply-chain management, transportation, networking, and finance. The size and complexity of such systems often preclude the availability of analytical methods for studying the resulting stochastic models, so simulation is frequently used.

Suppose that we are interested in analyzing the behavior of a system over a (possibly random) finite time horizon, and let X be a random variable denoting the system's (random) performance over the time interval of interest. For example, X may represent the time to complete a project, or X may be the loss of a portfolio of financial investments over the next two weeks. Most simulation textbooks focus on estimating the mean μ of X. This typically involves running independent and identically distributed (i.i.d.) replications of the system over the time horizon, and estimating μ via the sample average of the outputted performance from the replications. To provide a measure of the error in the estimate of μ, the analyst will often construct a confidence interval for μ using the simulated output; e.g., see Section 9.4.1 of [2].

In many contexts, however, performance measures other than a mean provide more useful information. One such measure is a *quantile*. For $0 < p < 1$, the p-quantile ξ_p of a random variable X is the smallest constant x such that $P(X \leq x) \geq p$. A well-known example is the median, which the 0.5-quantile. In terms of the cumulative distribution function (CDF) F of X, we can write $\xi_p = F^{-1}(p)$. Quantiles arise in many practical situations, often to measure risk. For example, in bidding on a project, a contractor may want to determine a date such that his firm has a 95% chance of finishing the project by that date, which is the 0.95-quantile. In finance, quantiles, which are known as values-at-risk, are frequently used as measures of risk of portfolios of assets [3]. For example, a portfolio manager may want to know the 0.99-quantile $\xi_{0.99}$ of the loss of his portfolio over the next two weeks, so there is a 1% chance that the loss over the next two weeks will exceed $\xi_{0.99}$.

Estimation of a quantile ξ_p is complicated by the fact that ξ_p cannot be expressed as the mean of a random variable, so one cannot estimate ξ_p via a sample average. However, the fact that $\xi_p = F^{-1}(p)$ suggests an alternative approach: develop an estimator of the CDF F, which may be a sample average, and then invert the estimated CDF.

In addition to a point estimator of ξ_p, we also would like a confidence interval (CI) of ξ_p to provide a measure of the accuracy of the point estimator. One approach to developing a CI is to first prove that the estimator of ξ_p satisfies a central limit theorem (CLT), and then construct a consistent estimator of the variance constant appearing in the CLT to obtain a CI.

Sometimes, the CI for quantile ξ_p is large, especially when $p \approx 0$ or $p \approx 1$, motivating the use of a variance-reduction technique (VRT) to obtain a quantile estimator with smaller error. VRTs that have been applied to quantile estimation include control variates (CV) [4], [5]; induced correlation, including antithetic variates (AV) and Latin hypercube sampling (LHS) [6], [7]; importance sampling (IS) [8]; and combined importance sampling and stratified sampling (IS+SS) [9]. Typically, variance reduction for quantile estimation entails applying a VRT to estimate the CDF, and then inverting the resulting CDF estimator to obtain a quantile estimator.

While most of the papers in the previous paragraph establish CLTs for the corresponding quantile estimators when applying VRTs, none of them provides a way to consistently estimate the CLTs' variance constants. Indeed, [9] states that this is "difficult and beyond the scope of this paper." To address this issue, [10] develops a general framework for analyzing some asymptotic properties (as the sample size gets large) of quantile estimators when applying

VRTs, and [10] shows how this can be exploited to construct consistent estimators of the variance constants in the CLTs. Also, [10] shows the framework encompasses CV, IS+SS and AV. In the current paper, we now do the same for LHS.

The rest of the paper is organized as follows. Section II develops the mathematical framework. We describe LHS in Section III, giving a CI for a quantile estimated via LHS. We present experimental results on a small example in Section IV. Section V provides some concluding remarks, and Section VI contains the proofs of our theorems. The current paper is based on and expands a previous conference paper [1], which includes neither the experimental results nor the proofs.

II. Background

Consider a random variable X having CDF F. For fixed $0 < p < 1$, the goal is to estimate the p-quantile $\xi_p = F^{-1}(p)$ of X, where $F^{-1}(q) = \inf\{x : F(x) \geq q\}$ for any $0 < q < 1$. We assume that X can be expressed as

$$X = g(U_1, U_2, \ldots, U_d) \tag{1}$$

for a known and given function $g : \Re^d \to \Re$, where $U_1, U_2, \ldots, U_d$ are i.i.d. unif$[0, 1)$ random numbers. Thus, generating an output X can be accomplished by transforming d i.i.d. uniforms through g. We now provide examples fitting in this framework.

Example 1: Suppose X is the time to complete a project, and we are interested in computing the 0.95-quantile $\xi_{0.95}$ of X. Assume the time to complete the project is modeled as a stochastic activity network (SAN) [11] having s activities, labeled $1, \ldots, s$. Suppose that there are r paths through the SAN, and let B_j be the set of activities on path j, $j = 1, 2, \ldots, r$. For each activity $i = 1, \ldots, s$, let A_i be its (random) duration. We allow for $A_1, \ldots, A_s$ to be dependent, and let H denote the joint distribution of $A \equiv (A_1, \ldots, A_s)$. Suppose that we can generate a sample of A from H using a fixed number d of i.i.d. unif$[0, 1)$ random variables $U_1, \ldots, U_d$. In the case when $A_1, \ldots, A_s$ are independent with each A_i having marginal distribution H_i, we can generate A_i as $A_i = H_i^{-1}(U_i)$, for $i = 1, \ldots, s$, assuming that H_i^{-1} can be computed efficiently. The length of the jth path in the SAN is $T_j = \sum_{i \in B_j} A_i$, and we can express $X = \max(T_1, \ldots, T_r)$ as the time to complete the project. Thus, the function g in (1) in this case takes the i.i.d. uniforms $U_1, \ldots, U_d$ as arguments, transforms them into $A_1, \ldots, A_s$, computes the length of each path T_j, and returns the maximum path length as X.

Example 2: Consider a financial portfolio consisting of a mix of investments, e.g., stocks, bonds and derivatives. Let $V(t)$ be the value of the portfolio at time t, and suppose the current time is $t = 0$. Let T denote two weeks, and we are interested in the 0.99-quantile of the loss X in the portfolio at the end of this period. Assume we have a stochastic model for $V(T)$, and suppose that simulating $V(T)$ given the current portfolio value $V(0)$ requires generating a fixed number d of i.i.d. unif$[0, 1)$ random variables $U_1, \ldots, U_d$; see Chapter 3 of [12] for algorithms to simulate $V(T)$ under various stochastic models describing the change in values of the investments. Thus, the function g in (1) takes the i.i.d. uniforms $U_1, \ldots, U_d$ as input, transforms them into $V(T)$, and then outputs $X = V(0) - V(T)$ as the portfolio loss. (A negative loss is a gain.) The 0.99-level value-at-risk is then the 0.99-quantile of X.

We now review how quantiles can be estimated when applying *crude Monte Carlo* (CMC) (i.e., no variance reduction). We first generate $n \times d$ i.i.d. unif$[0, 1)$ random numbers $U_{i,j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, d$, which we arrange in an $n \times d$ grid:

$$\begin{array}{|cccc|} \hline U_{1,1} & U_{1,2} & \cdots & U_{1,d} \\ U_{2,1} & U_{2,2} & \cdots & U_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ U_{n,1} & U_{n,2} & \cdots & U_{n,d} \\ \hline \end{array} \; . \tag{2}$$

Then we use the uniforms to generate n outputs $X_1, X_2, \ldots, X_n$ as follows:

$$\begin{array}{|rcl|} \hline X_1 &=& g(U_{1,1}, U_{1,2}, \ldots, U_{1,d}) \\ X_2 &=& g(U_{2,1}, U_{2,2}, \ldots, U_{2,d}) \\ & \vdots & \\ X_n &=& g(U_{n,1}, U_{n,2}, \ldots, U_{n,d}) \\ \hline \end{array} \; .$$

Thus, the ith row of uniforms $U_{i,1}, U_{i,2}, \ldots, U_{i,d}$ from (2) is used to generate the ith output X_i, and the independence of the columns of uniforms in (2) ensures that each X_i has the correct distribution F. Also, because of the independence of the rows of uniforms in (2), we have that $X_1, X_2, \ldots, X_n$ are i.i.d. We then estimate F via the *empirical CDF* $\hat{F}_n$, which is constructed as

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

where $I(A)$ is the indicator function of the event A, which takes the value 1 (resp., 0) if the event A occurs (resp., does not occur). Then

$$\hat{\xi}_{p,n} = \hat{F}_n^{-1}(p) \tag{3}$$

is the CMC estimator of the p-quantile ξ_p.

Let f denote the derivative of the CDF F, when it exists, and assume F is differentiable at ξ_p with $f(\xi_p) > 0$. The estimator $\hat{\xi}_{p,n}$ then is strongly consistent; i.e., $\hat{\xi}_{p,n} \to \xi_p$ as $n \to \infty$ with probability 1 (e.g., see p. 75 of [13]). Also, $\hat{\xi}_{p,n}$ satisfies the following CLT (Section 2.3.3 of [13]):

$$\frac{\sqrt{n}}{\kappa_p} \left(\hat{\xi}_{p,n} - \xi_p \right) \Rightarrow N(0, 1) \tag{4}$$

as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution (p. 8 of [13]) and $N(a, b^2)$ is a normal random variable

with mean a and variance b^2. The constant

$$\kappa_p = \sqrt{p(1-p)}\phi_p, \tag{5}$$

where

$$\phi_p = \frac{1}{f(\xi_p)}. \tag{6}$$

Using (4), we can construct an approximate $100(1-\alpha)\%$ CI for ξ_p as

$$\left[\hat{\xi}_{p,n} \pm z_{\alpha/2}\frac{\sqrt{p(1-p)}\,\phi_p}{\sqrt{n}}\right],$$

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ and Φ is the CDF of a $N(0,1)$. Unfortunately, the above CI is not implementable in practice since ϕ_p is unknown. Thus, we require a (weakly) consistent estimator of ϕ_p, and for the CMC case, such estimators have been proposed in the statistics literature [14]–[17]. As $\phi_p = \frac{d}{dp}F^{-1}(p)$, these papers develop finite-difference estimators of the derivative (Section 7.1 of [12]):

$$\hat{\phi}_{p,n}(h') = \frac{\hat{F}_n^{-1}(p+h') - \hat{F}_n^{-1}(p-h')}{2h'}, \tag{7}$$

where $h' \equiv h'_n > 0$ is known as the *smoothing parameter*. Consistency holds (i.e., $\hat{\phi}_{p,n}(h'_n) \Rightarrow \phi_p$ as $n \to \infty$) when $h'_n \to 0$ and $nh'_n \to \infty$ as $n \to \infty$. In this case, we obtain the following approximate $100(1-\alpha)\%$ CI for ξ_p:

$$J_n(h'_n) \equiv \left[\hat{\xi}_{p,n} \pm z_{\alpha/2}\frac{\sqrt{p(1-p)}\,\hat{\phi}_{p,n}(h'_n)}{\sqrt{n}}\right], \tag{8}$$

which is asymptotically valid in the sense that

$$P\{\xi_p \in J_n(h'_n)\} \to 1-\alpha$$

as $n \to \infty$. The consistency proofs of $\hat{\phi}_{p,n}(h'_n)$ in [15] and [16] utilize the fact that the i.i.d. outputs X_i, $i = 1, 2, \ldots, n$, can be represented as $X_i = F^{-1}(U_i)$ for $U_i \sim$ unif$[0,1)$ i.i.d. and then exploits properties of uniform order statistics. But this approach does not work when applying VRTs, including LHS, so we require another proof technique when applying LHS.

In the case of CMC, there has been some asymptotic analysis suggesting how one should choose the smoothing parameter $h' = h'_n$ in (7). To asymptotically minimize the mean square error (MSE) of $\hat{\phi}_{p,n}(h'_n)$ as an estimator of ϕ_p, one should choose $h'_n = O(n^{-1/5})$; see [16]. Alternatively, if one wants to minimize the coverage error of the confidence interval in (8), then [17] gives the asymptotically optimal rate for h'_n to be $O(n^{-1/3})$.

III. Latin Hypercube Sampling

LHS, which was introduced by [18] and further analyzed by [19], is a VRT that induces correlation among the simulated outputs in such a way as to increase the statistical efficiency under certain conditions. It can be viewed as an extension of stratified sampling (Section 4.3 of [12]). We next describe an approach in [6] for applying LHS to estimate the quantile ξ_p of the random variable X.

A. Single-Sample LHS

We will first generate a Latin hypercube (LH) sample of size t in dimension d as follows. Let $U_{i,j}$ for $i = 1, \ldots, t$ and $j = 1, \ldots, d$ be $t \times d$ i.i.d. unif$[0,1)$ random numbers. Let $\pi_1, \ldots, \pi_d$ be d independent random permutations of $\{1, \ldots, t\}$; i.e., for each π_j, each of the $t!$ permutations is equally likely. The $\pi_1, \ldots, \pi_d$ are generated independently of the $U_{i,j}$. For each $j = 1, 2, \ldots, d$, we have $\pi_j = (\pi_j(1), \pi_j(2), \ldots, \pi_j(t))$, and $\pi_j(i)$ is the number to which $i \in \{1, 2, \ldots, t\}$ is mapped in the jth permutation. Then define

$$V_{i,j} = \frac{\pi_j(i) - 1 + U_{i,j}}{t};\ i = 1, \ldots, t;\ j = 1, \ldots, d;$$

and arrange them into a $t \times d$ grid:

$$\begin{bmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,d} \\ V_{2,1} & V_{2,2} & \cdots & V_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ V_{t,1} & V_{t,2} & \cdots & V_{t,d} \end{bmatrix}. \tag{9}$$

For each $i = 1, \ldots, t$, it is easy to show that the ith row $V_i \equiv (V_{i,1}, V_{i,2}, \ldots, V_{i,d})$ from (9) is a vector of d i.i.d. unif$[0,1)$ numbers. But within each column j, the t uniforms $V_{1,j}, V_{2,j}, \ldots, V_{t,j}$ are dependent since they all use the same permutation π_j. Thus, the rows $V_1, V_2, \ldots, V_t$ are dependent, and we call the $t \times d$ uniforms in (9) an *LH sample* of size t in d dimensions.

The LH sample in (9) has an interesting feature, as we now explain. Partition the unit interval $[0,1)$ into t equal-length subintervals $[0, 1/t), [1/t, 2/t), \ldots, [(t-1)/t, 1)$. Then one can show that the t uniforms in any column of (9) have the property that exactly one of them lies in each subinterval. Each of the d columns thus forms a stratified sample of the unit interval in one dimension, so the LH sample can be seen as an extension of stratified sampling in d dimensions.

We use the ith row V_i from (9) to generate an output X'_i:

$$\begin{array}{rcl} X'_1 &=& g(V_{1,1}, V_{1,2}, \ldots, V_{1,d}) \\ X'_2 &=& g(V_{2,1}, V_{2,2}, \ldots, V_{2,d}) \\ & \vdots & \\ X'_t &=& g(V_{t,1}, V_{t,2}, \ldots, V_{t,d}) \end{array}. \tag{10}$$

Each X'_i has distribution F since $V_{i,1}, V_{i,2}, \ldots, V_{i,d}$ from the ith row of (9) are i.i.d. unif$[0,1)$. An estimator of F is then

$$\bar{F}_t(x) = \frac{1}{t}\sum_{i=1}^{t} I(X'_i \le x).$$

We can then invert this to obtain

$$\bar{\xi}_{p,t} = \bar{F}_{p,t}^{-1}(p),$$

which we call the *single-sample LHS (SS-LHS) quantile estimator*. As shown in [6], the estimator $\bar{\xi}_{p,t}$ satisfies the CLT

$$\frac{\sqrt{t}}{\eta_p}(\bar{\xi}_{p,t} - \xi) \Rightarrow N(0,1) \tag{11}$$

as $t \to \infty$ under certain regularity conditions, where

$$\eta_p = \zeta_p \phi_p, \tag{12}$$

ζ_p is given in [6], and ϕ_p is defined in (6). Recalling (4) and (5), which are for the case of CMC, we see that SS-LHS yields an asymptotic variance reduction when $\zeta_p < \sqrt{p(1-p)}$, and [6] provides sufficient conditions to ensure this holds.

Constructing a confidence interval for ξ_p based on the CLT (11) requires consistently estimating η_p. But the dependence of the rows $V_1, V_2, \ldots, V_t$ implies that $X'_1, X'_2, \ldots, X'_t$ are dependent, and this complicates constructing an estimator for η_p. Indeed, [6] does not develop estimators for ζ_p and ϕ_p.

B. Combined Multiple-LHS

To avoid the above complication, rather than taking a single LH sample of size t to obtain a set of t (dependent) outputs as in [6], we instead generate a total of $n = mt$ outputs in groups of t, where each group is constructed from an LH sample of size t and the m different groups are sampled independently. We then use all n samples to compute a CDF estimator, which we invert to obtain a quantile estimator. We incur some loss in statistical efficiency by using m different independent LH samples, each of size t, instead of taking one big LH sample of size n, but [19] notes the degradation is small when t/d is large.

We now provide details of our approach. Define $t \times d \times m$ i.i.d. unif$[0,1)$ random variables $U_{i,j}^{(k)}$, where $i = 1, \ldots, t$; $j = 1, \ldots, d$; and $k = 1, \ldots, m$. Also, let $\pi_j^{(k)}$ for $j = 1, \ldots, d$ and $k = 1, \ldots, m$ be $d \times m$ independent permutations of $\{1, \ldots, t\}$, and the $U_{i,j}^{(k)}$ and the $\pi_j^{(k)}$ are all mutually independent. Each $\pi_j^{(k)} = (\pi_j^{(k)}(1), \pi_j^{(k)}(2), \ldots, \pi_j^{(k)}(t))$, and $\pi_j^{(k)}(i)$ is the value to which i is mapped in permutation $\pi_j^{(k)}$. For each $k = 1, \ldots, m$, let

$$V_{i,j}^{(k)} = \frac{\pi_j^{(k)}(i) - 1 + U_{i,j}^{(k)}}{t};\ i = 1, \ldots, t;\ j = 1, \ldots, d;$$

and we arrange them in a $t \times d$ grid:

$$\begin{bmatrix} V_{1,1}^{(k)} & V_{1,2}^{(k)} & \cdots & V_{1,d}^{(k)} \\ V_{2,1}^{(k)} & V_{2,2}^{(k)} & \cdots & V_{2,d}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{t,1}^{(k)} & V_{t,2}^{(k)} & \cdots & V_{t,d}^{(k)} \end{bmatrix}, \tag{13}$$

which is an LH sample of size t in d dimensions. We have m such independent grids. Thus, for each grid $k = 1, \ldots, m$, and for each $i = 1, \ldots, t$, we have that $V_i^{(k)} \equiv (V_{i,1}^{(k)}, V_{i,2}^{(k)}, \ldots, V_{i,d}^{(k)})$ is a vector of d i.i.d. unif$[0,1)$ numbers. Also, for each $k = 1, \ldots, m$, the t vectors $V_1^{(k)}, V_2^{(k)}, \ldots, V_t^{(k)}$ are dependent. We use the ith row $V_i^{(k)}$ from (13) to generate an output $X_i^{(k)}$:

$$\boxed{\begin{array}{rcl} X_1^{(k)} &=& g(V_{1,1}^{(k)}, V_{1,2}^{(k)}, \ldots, V_{1,d}^{(k)}) \\ X_2^{(k)} &=& g(V_{2,1}^{(k)}, V_{2,2}^{(k)}, \ldots, V_{2,d}^{(k)}) \\ & \vdots & \\ X_t^{(k)} &=& g(V_{t,1}^{(k)}, V_{t,2}^{(k)}, \ldots, V_{t,d}^{(k)}) \end{array}}. \tag{14}$$

Each $X_i^{(k)}$ has distribution F since $V_{i,1}^{(k)}, V_{i,2}^{(k)}, \ldots, V_{i,d}^{(k)}$ from the ith row of (13) are i.i.d. unif$[0,1)$.

Since we independently repeat (14) for $k = 1, 2, \ldots, m$, we get $t \times m$ outputs, which we arrange in a grid:

$$\begin{bmatrix} \boxed{\begin{array}{c} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_t^{(1)} \end{array}} & \boxed{\begin{array}{c} X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_t^{(2)} \end{array}} & \cdots & \boxed{\begin{array}{c} X_1^{(m)} \\ X_2^{(m)} \\ \vdots \\ X_t^{(m)} \end{array}} \\ \end{bmatrix}. \tag{15}$$

Each boxed column k corresponds to one set of t (dependent) outputs from an LH sample of size t as in (14), so the rows of (15) are dependent. But since we generate the m LH samples independently, the columns of (15) are independent. We subsequently form an estimator of the CDF F as

$$\tilde{F}_{m,t}(x) = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{t} \sum_{i=1}^{t} I(X_i^{(k)} \le x). \tag{16}$$

For any $0 < p < 1$, we then obtain

$$\tilde{\xi}_{p,m,t} = \tilde{F}_{m,t}^{-1}(p), \tag{17}$$

which we call the *combined multiple-LHS (CM-LHS) estimator* of ξ_p.

As we will later see, the CM-LHS quantile estimator $\tilde{\xi}_{p,m,t}$ obeys a CLT, and we will use an approach developed in [10] to estimate the asymptotic variance in the CLT. To do this, first let p_m be any perturbed value of p satisfying $p_m \to p$ as $m \to \infty$, and let $\tilde{\xi}_{p_m,m,t} = \tilde{F}_{m,t}^{-1}(p_m)$. The following theorem, whose proof is in Section VI-A, establishes that $\tilde{\xi}_{p_m,m,t}$ satisfies a so-called Bahadur representation [20].

Theorem 1: If $f(\xi_p) > 0$, then for any $p_m = p + O(m^{-1/2})$,

$$\tilde{\xi}_{p_m,m,t} = \xi'_{p_m} - \frac{\tilde{F}_{m,t}(\xi_p) - p}{f(\xi_p)} + R_m \tag{18}$$

with $\xi'_{p_m} = \xi_p + (p_m - p)/f(\xi_p)$ and

$$\sqrt{m} R_m \Rightarrow 0 \tag{19}$$

as $m \to \infty$. If in addition f is continuous in a neighborhood of ξ_p, then (18)–(19) hold with $\xi'_{p_m} = F^{-1}(p_m)$ for all $p_m \to p$.

A consequence of the Bahadur representation in (18)–(19) is that the CM-LHS quantile estimator $\tilde{\xi}_{p,m,t}$ then satisfies a CLT, which we can see as follows. Take $p_m = p$ in (18), and rearranging terms and multiplying by $\sqrt{m}$ lead to

$$\sqrt{m}(\tilde{\xi}_{p,m,t} - \xi_p) = \sqrt{m}\left(\frac{p - \tilde{F}_{m,t}(\xi_p)}{f(\xi_p)}\right) + \sqrt{m}R_m. \quad (20)$$

Let

$$W^{(k)}(x) = \frac{1}{t}\sum_{i=1}^{t} I(X_i^{(k)} \le x),$$

so $\tilde{F}_{m,t}(x) = \frac{1}{m}\sum_{k=1}^{m} W^{(k)}(x)$. For any fixed x, the $W^{(k)}(x)$, $k = 1, \ldots, m$, are i.i.d., and since each $X_i^{(k)}$ has distribution F, we see that $E[W^{(k)}(x)] = F(x)$. Define

$$\psi_p^2 = \mathrm{Var}[W^{(k)}(\xi_p)],$$

which is finite since $0 \le W^{(k)}(x) \le 1$ for all x. Hence, the ordinary CLT (e.g., p. 28 of [13]) implies that for fixed t,

$$\frac{\sqrt{m}}{\psi_p}\left(F(\xi_p) - \tilde{F}_{m,t}(\xi_p)\right) \Rightarrow N(0,1) \quad (21)$$

as $m \to \infty$. Since $F(\xi_p) = p$, we then get that if $f(\xi_p) > 0$,

$$\frac{\sqrt{m}}{\psi_p / f(\xi_p)}\left(\frac{p - \tilde{F}_{m,t}(\xi_p)}{f(\xi_p)}\right) \Rightarrow N(0,1)$$

as $m \to \infty$ with t fixed. Thus, it follows from (19), (20) and Slutsky's theorem (e.g., p. 19 of [13]) that for fixed t,

$$\frac{\sqrt{m}}{\tau_p}\left(\tilde{\xi}_{p,m,t} - \xi_p\right) \Rightarrow N(0,1) \quad (22)$$

as $m \to \infty$, where

$$\tau_p = \psi_p \phi_p$$

and $\phi_p = 1/f(\xi_p)$, as in (6).

To construct a CI for ξ_p based on the CLT in (22) for the CM-LHS quantile estimator, we now want consistent estimators of ψ_p and ϕ_p. We can estimate ψ_p^2 via

$$\tilde{\psi}_{p,m,t}^2 = \frac{1}{m-1}\sum_{k=1}^{m}\left(W^{(k)}(\tilde{\xi}_{p,m,t}) - \bar{W}_m\right)^2,$$

where

$$\bar{W}_m = \frac{1}{m}\sum_{k=1}^{m} W^{(k)}(\tilde{\xi}_{p,m,t}).$$

Even though $W^{(k)}(x)$, $k = 1, \ldots, m$, are i.i.d. for any fixed x, each $W^{(k)}(\tilde{\xi}_{p,m,t})$ depends on $\tilde{\xi}_{p,m,t}$, which is a function of *all* of the samples by (16) and (17), and this induces dependence among $W^{(k)}(\tilde{\xi}_{p,m,t})$, $k = 1, 2, \ldots, m$. This complicates the analysis of $\tilde{\psi}_{p,m,t}$, but we can apply the techniques in [10] to prove that

$$\tilde{\psi}_{p,m,t} \Rightarrow \psi_p \quad (23)$$

as $m \to \infty$ for fixed $t \ge 1$. An estimator for $\phi_p = \frac{d}{dp}F^{-1}(p)$ is the finite difference

$$\tilde{\phi}_{p,m,t}(h_m) = \frac{\tilde{F}_{m,t}^{-1}(p + h_m) - \tilde{F}_{m,t}^{-1}(p - h_m)}{2h_m} \quad (24)$$

for smoothing parameter $h_m > 0$. In the proof of Theorem 2 below, we prove that

$$\tilde{\phi}_{p,m,t}(h_m) \Rightarrow \phi_p \quad (25)$$

as $m \to \infty$ for fixed $t \ge 1$ under certain conditions (given in Theorem 2) on h_m and the CDF F.

When (23) and (25) hold, Slutsky's theorem guarantees the CLT in (22) remains valid when we replace $\tau_p = \psi_p \phi_p$ with its consistent estimator $\tilde{\psi}_{p,m,t}\tilde{\phi}_{p,m,t}(h_m)$. We then obtain the following approximate $100(1-\alpha)\%$ CI for ξ_p when applying CM-LHS:

$$\tilde{J}_{m,t}(h_m) \equiv \left[\tilde{\xi}_{p,m,t} \pm z_{\alpha/2}\frac{\tilde{\psi}_{p,m,t}\tilde{\phi}_{p,m,t}(h_m)}{\sqrt{m}}\right]. \quad (26)$$

The following theorem, whose proof is in Section VI-B, establishes the asymptotic validity of the above CI.

Theorem 2: If $f(\xi_p) > 0$, then for any fixed $t \ge 1$,

$$\lim_{m\to\infty} P\{\xi_p \in \tilde{J}_{m,t}(h_m)\} = 1 - \alpha \quad (27)$$

for $h_m = cm^{-1/2}$ and any constant $c > 0$. If in addition f is continuous in a neighborhood of ξ_p, then (27) holds for any $h_m > 0$ satisfying $h_m \to 0$ and $1/h_m = O(m^{1/2})$.

The range of valid values for the smoothing parameter h_m in the second case of Theorem 2 is of particular interest since it covers the asymptotically optimal rates for CMC, as discussed at the end of Section II.

We close this section describing how to invert $\tilde{F}_{m,t}$, which is needed to compute the CM-LHS quantile estimator $\tilde{\xi}_{p,m,t}$ in (17) and the finite difference $\tilde{\phi}_{p,m,t}(h_m)$ in (24). First take the $n = mt$ values $X_i^{(k)}$ for $i = 1, \ldots, t$ and $k = 1, \ldots, m$, and sort them in ascending order as $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. Then for $0 < q < 1$, we can compute $\tilde{F}_{m,t}^{-1}(q) = X_{(\lceil nq \rceil)}$, where $\lceil \cdot \rceil$ is the round-up function.

IV. Experimental Results

We now present some results from running experiments on a small SAN, and below we follow the notation developed in Example 1 from Section II. Previously studied in [4] and [10], the SAN has $s = 5$ activities, where the activity durations are i.i.d. exponential with mean 1. Since the activities are independent, we generate the s activity durations using $d = s$ i.i.d. unif$[0, 1)$ random variables via inversion; e.g., see Section 2.2.1 of [12]. There are $r = 3$ paths in the SAN, with $B_1 = \{1, 2\}$, $B_2 = \{1, 3, 5\}$, and $B_3 = \{4, 5\}$ as the collections of activities on the 3 paths. The goal is to estimate the p-quantile ξ_p of the length X of the longest path for different values of p.

		CM-LHS					
		normal critical point			Student-T critical point		
n	CMC	$t = 10$	$t = 20$	$t = 50$	$t = 10$	$t = 20$	$t = 50$
100	0.899	0.877	0.838	0.618	0.906	0.904	0.739
	(0.359)	(0.229)	(0.212)	(0.171)	(0.255)	(0.275)	(0.658)
400	0.881	0.879	0.867	0.838	0.887	0.883	0.883
	(0.162)	(0.106)	(0.101)	(0.098)	(0.108)	(0.106)	(0.112)
1600	0.885	0.887	0.879	0.879	0.889	0.884	0.889
	(0.081)	(0.053)	(0.051)	(0.050)	(0.053)	(0.052)	(0.052)
6400	0.898	0.895	0.891	0.897	0.895	0.893	0.899
	(0.041)	(0.027)	(0.026)	(0.025)	(0.027)	(0.026)	(0.025)

Table I
COVERAGES (AND AVERAGE HALF WIDTHS) FOR CONFIDENCE INTERVALS USING CMC AND CM-LHS FOR $p = 0.5$

		CM-LHS					
		normal critical point			Student-T critical point		
n	CMC	$t = 10$	$t = 20$	$t = 50$	$t = 10$	$t = 20$	$t = 50$
100	0.868	0.861	0.810	0.549	0.891	0.883	0.616
	(0.706)	(0.578)	(0.512)	(0.362)	(0.644)	(0.663)	(0.391)
400	0.885	0.877	0.869	0.846	0.886	0.885	0.891
	(0.348)	(0.285)	(0.260)	(0.230)	(0.292)	(0.274)	(0.265)
1600	0.899	0.891	0.888	0.878	0.893	0.892	0.889
	(0.173)	(0.142)	(0.130)	(0.117)	(0.143)	(0.131)	(0.121)
6400	0.898	0.902	0.895	0.890	0.903	0.896	0.893
	(0.086)	(0.071)	(0.065)	(0.059)	(0.071)	(0.065)	(0.059)

Table II
COVERAGES (AND AVERAGE HALF WIDTHS) FOR CONFIDENCE INTERVALS USING CMC AND CM-LHS FOR $p = 0.9$

Tables I and II present results for $p = 0.5$ and $p = 0.9$, respectively. In our experiments we constructed nominal $100(1-\alpha)\% = 90\%$ confidence intervals (CIs) for ξ_p based on the CMC and the CM-LHS quantile estimators from (3) and (17), respectively. We ran 10^4 independent replications to estimate coverage levels and average half widths (in parentheses) of the confidence intervals. We varied the total sample size $n = 4^a \times 100$ for $a = 0, 1, 2, 3$. When applying CM-LHS, we varied the LH sample size as $t = 10$, 20 and 50. We set the smoothing parameter for CMC to be $h'_n = 0.5n^{-1/2}$; for CM-LHS, we chose $h_m = 0.5(tm)^{-1/2}$, so $h'_n = h_m$ in all our experiments when $n = mt$. Column 2 in the tables gives the results for the CMC CIs from (8). Columns 3–5 correspond to the CM-LHS CIs in (26), where we recall $z_{\alpha/2}$ is the $1 - \alpha/2$ critical point of a standard normal. For columns 6–8, we replace $z_{\alpha/2}$ in (26) with the $1 - \alpha/2$ critical point $T_{m-1,\alpha/2}$ of a Student-T distribution with $m - 1$ degrees of freedom (df); i.e., if G_{m-1} is the CDF of a Student-T random variable with $m - 1$ df, then $T_{m-1,\alpha/2} = G^{-1}_{m-1}(1 - \alpha/2)$. Since $T_{m-1,\alpha/2} \to z_{\alpha/2}$ as $m \to \infty$, the resulting CI with critical point $T_{m-1,\alpha/2}$ is also asymptotically valid as $m \to \infty$ with fixed t. But $T_{m-1,\alpha/2} > z_{\alpha/2}$ for all m and α, so CIs with Student-T critical points are wider than those with the normal critical point, which can lead to higher coverage.

We first discuss the results for $p = 0.5$ (Table I). CMC gives close to nominal coverage (0.9) for all sample sizes n. The same is true for CM-LHS with $t = 10$ and the normal critical point $z_{\alpha/2}$, but now the average half width decreased by about 35%. For CM-LHS with $t = 20$ and $t = 50$ and the normal critical point, coverage is less than 0.9 for small n, but when n is large, coverage is close to nominal and the average half width is about the same as for $t = 10$. The poor coverage for small n and large t arises because the value of $m = n/t$ is then small; e.g., when $t = 50$ and $n = 100$, then m is only 2. But the validity of our CM-LHS confidence interval in (26) requires $m \to \infty$ (see Theorem 2), so we see poor coverage when t is large and n is small. Coverage is improved for small n when we instead use the Student-T critical point $T_{m-1,\alpha/2}$ to construct the CI, but it is still significantly below the nominal level for $t = 50$. However, this problem goes away when n is large for both the normal and Student-T critical points since then m is also large, so the asymptotic validity takes effect.

The results for $p = 0.9$ (Table II) exhibit similar qualities to those for $p = 0.5$, but there are some important differences. For $p = 0.9$, we see that the amount that CM-LHS decreases the average half width depends on the choice of t. For $t = 10$, the average half width shrunk by about 17% compared to CMC. Average half width is even smaller when using CM-LHS with larger t, with more than 30% reduction for $t = 50$. Thus, while the choice of LH sample size t does not appear to have much affect on the amount of variance reduction when estimating quantiles in the middle of the distribution, it can have a large impact when estimating more extreme quantiles.

Recall that the CM-LHS quantile estimator $\tilde{\xi}_{p,m,t}$ in (17) is computed by inverting $\tilde{F}_{m,t}$ in (16), which depends on *all* $n = mt$ samples generated. An alternative approach is to use *batching*, where we compute m different independent quantile estimates, one from each of the m columns of size t in (15). Since the m columns are i.i.d., we can then compute the average and sample variance of the m i.i.d. quantile estimates to construct a confidence interval for ξ_p. This approach is discussed on p. 242 of [12] for the case of estimating a mean using LHS, and now we provide details on how one could apply batching for estimating a quantile; see also p. 491 of [12]. For each $k = 1, 2, \ldots, m$, let $\dot{\xi}_{p,k,t} = \dot{F}^{-1}_{k,t}(p)$, where

$$\dot{F}_{k,t}(x) = \frac{1}{t} \sum_{i=1}^{t} I(X_i^{(k)} \le x),$$

which is the estimated CDF from using only the kth boxed column of LH samples in (15). Then the *batched LHS quantile estimator* is

$$\bar{\xi}_{p,m,t} = \frac{1}{m} \sum_{k=1}^{m} \dot{\xi}_{p,k,t}, \tag{28}$$

	$p = 0.5$			$p = 0.9$		
n	$t = 10$	$t = 20$	$t = 50$	$t = 10$	$t = 20$	$t = 50$
100	0.587	0.817	0.891	0.437	0.733	0.876
	(0.218)	(0.242)	(0.607)	(0.470)	(0.546)	(1.327)
400	0.093	0.531	0.836	0.042	0.411	0.768
	(0.103)	(0.103)	(0.111)	(0.222)	(0.234)	(0.245)
1600	0.000	0.066	0.652	0.000	0.021	0.487
	(0.051)	(0.050)	(0.051)	(0.110)	(0.114)	(0.113)
6400	0.000	0.000	0.178	0.000	0.000	0.046
	(0.025)	(0.025)	(0.025)	(0.055)	(0.057)	(0.056)

Table III
COVERAGES (AND AVERAGE HALF WIDTHS) FOR LHS CONFIDENCE INTERVALS USING BATCHING WITH $m = n/t$ BATCHES AND LH SAMPLE SIZE t

n	$p = 0.5$	$p = 0.9$
100	0.587	0.437
	(0.218)	(0.470)
400	0.807	0.720
	(0.109)	(0.241)
1600	0.879	0.850
	(0.055)	(0.118)
6400	0.889	0.888
	(0.027)	(0.060)

Table IV
COVERAGES (AND AVERAGE HALF WIDTHS) FOR LHS CONFIDENCE INTERVALS USING BATCHING WITH $m = 10$ BATCHES AND LH SAMPLE SIZE $t = n/m$

and an approximate $100(1-\alpha)\%$ confidence interval for ξ_p is

$$\left[\bar{\xi}_{p,m,t} \pm T_{m-1,\alpha/2}\frac{S_m}{\sqrt{m}}\right], \tag{29}$$

where

$$S_m^2 = \frac{1}{m-1}\sum_{k=1}^{m}(\dot{\xi}_{p,k,t} - \bar{\xi}_{p,m,t})^2.$$

As in our experiments with the CM-LHS quantile estimator in Tables I and II, we also ran experiments with the batched CI in (29) for $p = 0.5$ and 0.9 to study its behavior as the total sample size n increases. Since $n = mt$, increasing n requires correspondingly increasing the number m of batches and/or the LH sample size t.

Table III presents results using batching in which we keep t fixed at 10, 20 or 50 so $m = n/t$ increases as n grows. For $n = 100$, coverage is close to nominal for $t = 50$, but coverage is significantly low for $t = 10$ and $t = 20$. As n increases, coverage actually *worsens* for each t. This occurs because quantile estimators are biased, with bias decreasing as the sample size increases; see [6]. In each column of Table III, t is fixed and does not increase, so $\bar{\xi}_{p,m,t}$ is the average of m biased quantile estimators $\dot{\xi}_{p,k,t}$, $k = 1, 2, \ldots, m$, each computed from a single LH sample of size t, with more bias for small t. If $E[|\dot{\xi}_{p,1,t}|] < \infty$, then

$$\bar{\xi}_{p,m,t} = \frac{1}{m}\sum_{k=1}^{m}\dot{\xi}_{p,k,t} \Rightarrow E[\dot{\xi}_{p,1,t}]$$

as $m \to \infty$ by the law of large numbers since $\dot{\xi}_{p,k,t}$, $k = 1, 2, \ldots, m$, are i.i.d. But $E[\dot{\xi}_{p,1,t}] \neq \xi_p$ because of the bias for fixed t. Thus, as n and m increase, the CI in (29) is shrinking at rate $m^{-1/2}$, but it is centered at an estimate whose bias is not decreasing since t is fixed. This results in the poor coverage. For a batching approach to work, we instead need the LH sample size t to increase as the total sample size n increases to ensure the bias of $\bar{\xi}_{p,m,t}$ decreases.

Table IV presents results with batching in which the number of batches is fixed at $m = 10$, so the LH sample size $t = n/m$ grows as the total sample size n increases. In contrast to the case when t was fixed as n increases (Table III), we now see in Table IV that the coverage levels of the CIs in (29) converge to the nominal level 0.9 as n increases. However, compared to the results in Tables I and II for the CM-LHS quantile estimator, we see that batching with t increasing in n gives lower coverage when n is small. This occurs because of the bias of quantile estimators, as we discussed in the previous paragraph. The batched LHS estimator in (28) averages m i.i.d. quantile estimators, each based on an LH sample of size t, so the bias of the batched quantile estimator is determined by t. On the other hand, CM-LHS computes a single quantile estimator (17) based on inverting the CDF estimator (16) from all of the $n = mt$ samples. Because the bias of quantile estimators decreases as the sample size grows, the batched LHS quantile estimator has larger bias than the CM-LHS quantile estimator. This leads to the coverage of batched CI in (29) converging more slowly to the nominal level as the total sample size n grows than the CI in (26) for CM-LHS. This property is more pronounced for $p = 0.9$ than for $p = 0.5$, so batching seems to do worse for extreme quantiles than for those in the middle of the distribution.

V. CONCLUSION

We presented an asymptotically valid CI for a quantile estimated using LHS. We developed a combined multiple-LHS approach in which one generates a total of n samples in m independent groups, where each group is generated from an LH sample of size t. Using a general framework developed in [10] for quantile estimators from applying VRTs, we proved that the resulting CM-LHS quantile estimator satisfies a Bahadur representation, which provides an asymptotic approximation for the estimator. The Bahadur representation implies a CLT for the CM-LHS quantile estimator and also allows us to construct a consistent estimator for the asymptotic variance in the CLT.

We ran experiments on a small SAN, and our results demonstrate the asymptotic validity of our CM-LHS CIs. Also, the results show that CM-LHS can decrease the aver-

age half width of confidence intervals relative to CMC, but the amount of decrease can depend of the LH sample size t when estimating an extreme quantile. We also experimented with an alternative approach based on batching with m independent batches, each consisting of t samples constructed from an LH sample of size t. To lead to asymptotically valid CIs, batching requires t to increase as the total sample size $n = mt$ grows. Compared to CM-LHS, batching needs larger samples sizes n for the CIs to have close to nominal coverage. Further work is needed to study the empirical performance of the proposed method when simulating other larger stochastic models.

VI. Appendix

A. Proof of Theorem 1

Chu and Nakayama [10] develop a general framework giving sufficient conditions for quantile estimators obtained when applying VRTs to satisfy a Bahadur representation, and we now establish (18)–(19) by showing that our CM-LHS approach fits into the framework. Specifically, for the first result in Theorem 1 (i.e., for $p_m = p + O(m^{-1/2})$), we need to show that $\tilde{F}_{m,t}$ in (16) satisfies the following assumptions from [10]:

Assumption A1: $P(M_{m,t}) \to 1$ as $m \to \infty$, where $M_{m,t}$ is the event that $\tilde{F}_{m,t}(x)$ is monotonically increasing in x for fixed m and t.

Assumption A2: $E[\tilde{F}_{m,t}(x)] = F(x)$ for all x, and for every $a_m = O(m^{-1/2})$,

$$E\left[\tilde{F}_{m,t}(\xi_p + a_m) - \tilde{F}_{m,t}(\xi_p)\right]^2 = [F(\xi_p + a_m) - F(\xi_p)]^2 + s_m(a_m)/m$$

with $s_m(a_m) \to 0$ as $m \to \infty$.

Assumption A3: $\sqrt{m}\left[\tilde{F}_{m,t}(\xi_p) - F(\xi_p)\right] \Rightarrow N(0, \psi_p^2)$ as $n \to \infty$ for some $0 < \psi_p < \infty$.

As shown in [10], if $f(\xi_p) > 0$, then Assumptions A1–A3 imply that (18) and (19) hold for any $p_m = p + O(m^{-1/2})$. Also, [10] proves that if we additionally strengthen A2 to be true for all $a_m \to 0$ and f is continuous in a neighborhood of ξ_p, then (18) and (19) hold for any $p_m \to p$, which is what we need to show for the second part of Theorem 1. Thus, we now prove that $\tilde{F}_{m,t}$ in (16) satisfies A1–A3, with A2 holding for all $a_m \to 0$.

Examining (16), we see that Assumption A1 holds since $\tilde{F}_{m,t}(x)$ is always monotonically increasing in x because each $I(X_i^{(k)} \le x)$ has this property. Also, we previously showed Assumption A3 holds in (21), so it remains to prove Assumption A2 holds, which we show is true for any $a_m \to 0$.

Since each $X_i^{(k)}$ has distribution F, we have that

$$E[\tilde{F}_{m,t}(x)] = \frac{1}{m}\sum_{k=1}^{m}\frac{1}{t}\sum_{i=1}^{t}E[I(X_i^{(k)} \le x)] = F(x)$$

for all x, so the first part of A2 holds. To establish the second part of A2, let $\rho_m = \min(\xi_p, \xi_p + a_m)$ and $\rho'_m = \max(\xi_p, \xi_p + a_m)$ for any $a_m \to 0$. Then

$$b_m \equiv E\left[\tilde{F}_{m,t}(\xi_p + a_m) - \tilde{F}_{m,t}(\xi_p)\right]^2 = E\left(\frac{1}{m}\sum_{k=1}^{m}\frac{1}{t}\sum_{i=1}^{t}C_i^{(k)}\right)^2, \quad (30)$$

where $C_i^{(k)} = I(\rho_m < X_i^{(k)} \le \rho'_m)$. Note that $C_i^{(k)}$ and $C_{i'}^{(k')}$ are independent for $k \ne k'$ and any i and i' since $C_i^{(k)}$ and $C_{i'}^{(k')}$ correspond to outputs from different LH samples. Thus, expanding the square in (30) gives

$$b_m = \frac{1}{(mt)^2}\sum_{k=1}^{m}\sum_{i=1}^{t}E[(C_i^{(k)})^2] + \frac{1}{(mt)^2}\sum_{k=1}^{m}\sum_{\substack{k'=1\\k'\ne k}}^{m}\sum_{i=1}^{t}\sum_{i'=1}^{t}E[C_i^{(k)}]E[C_{i'}^{(k')}] + c_m,$$

where

$$c_m = \frac{1}{(mt)^2}\sum_{k=1}^{m}\sum_{i=1}^{t}\sum_{\substack{i'=1\\i'\ne i}}^{t}E[C_i^{(k)}C_{i'}^{(k)}].$$

For all i and k, we have $E[C_i^{(k)}] = F(\rho'_m) - F(\rho_m) \equiv d_m$ and $(C_i^{(k)})^2 = C_i^{(k)}$. Thus,

$$b_m = \frac{1}{mt}d_m + \frac{m-1}{m}d_m^2 + c_m = d_m^2 + e_m + c_m,$$

where

$$e_m = \frac{1}{mt}d_m - \frac{1}{m}d_m^2.$$

Since $[F(\xi_p + a_m) - F(\xi_p)]^2 = d_m^2$, A2 holds if we show that $me_m \to 0$ and $mc_m \to 0$ as $m \to \infty$.

Now $me_m \to 0$ holds since

$$d_m \to 0 \quad (31)$$

because $a_m \to 0$ and F is continuous at ξ_p. To show $mc_m \to 0$, note that

$$E[C_i^{(k)}C_{i'}^{(k)}] = P(\rho_m < X_i^{(k)} \le \rho'_m,\ \rho_m < X_{i'}^{(k)} \le \rho'_m) \le P(\rho_m < X_i^{(k)} \le \rho'_m) = d_m,$$

so it follows that

$$|mc_m| \le \frac{m}{(mt)^2}\sum_{k=1}^{m}\sum_{j=1}^{t}\sum_{\substack{j'=1\\j'\ne j}}^{t}d_m = \frac{t-1}{t}d_m \to 0$$

as $m \to \infty$ by (31). Thus, the proof is complete.

B. Proof of Theorem 2

All that remains is to prove (25) for the two different cases given in the theorem. For the first case, it is shown in [10] that if $f(\xi_p) > 0$, then it follows from the first Bahadur representation (i.e., with $\xi'_{p_m} = \xi_p + (p_m - p)/f(\xi_p)$) in Theorem 1 that (25) holds when $h_m = cm^{-1/2}$ for any constant $c > 0$. Moreover, [10] shows that when f is continuous in a neighborhood of ξ_p, the second Bahadur representation in Theorem 1 in which $\xi'_{p_m} = F^{-1}(p_m)$ implies (25) holds for any h_m satisfying $h_m \to 0$ and $1/h_m = O(m^{1/2})$, which is what we need to show for the second case of Theorem 2.

ACKNOWLEDGMENT

This work has been supported in part by the National Science Foundation under Grant No. CMMI-0926949. 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 National Science Foundation.

REFERENCES

[1] M. K. Nakayama, "Confidence Intervals for Quantiles When Applying Latin Hypercube Sampling," in *Proceedings of the Second International Conference on Advances in System Simulation (SIMUL 2010)*, pp. 78–81, 2010.

[2] A. M. Law, *Simulation Modeling and Analysis*, 4th ed. New York: McGraw-Hill, 2006.

[3] D. Duffie and J. Pan, "An overview of value at risk," *Journal of Derivatives*, vol. 4, pp. 7–49, 1997.

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

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

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

[7] X. Jin, M. C. Fu, and X. Xiong, "Probabilistic error bounds for simulation quantile estimation," *Management Science*, vol. 49, pp. 230–246, 2003.

[8] 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*. Publishing House of St. Petersburg University, St. Petersburg, Russia, 1996, pp. 180–185.

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

[10] F. Chu and M. K. Nakayama, "Confidence intervals for quantiles when applying variance-reduction techniques," *submitted*, 2010.

[11] V. G. Adlakha and V. G. Kulkarni, "A classified bibliography of research on stochastic PERT networks," *INFOR*, vol. 27, pp. 272–296, 1989.

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

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

[14] M. M. Siddiqui, "Distribution of quantiles in samples from a bivariate population," *Journal of Research of the National Bureau of Standards B*, vol. 64, pp. 145–150, 1960.

[15] D. A. Bloch and J. L. Gastwirth, "On a simple estimate of the reciprocal of the density function," *Annals of Mathematical Statistics*, vol. 39, pp. 1083–1085, 1968.

[16] E. Bofinger, "Estimation of a density function using order statistics," *Australian Journal of Statistics*, vol. 17, pp. 1–7, 1975.

[17] P. Hall and S. J. Sheather, "On the distribution of a Studentized quantile," *Journal of the Royal Statistical Society B*, vol. 50, pp. 381–391, 1988.

[18] M. D. McKay, W. J. Conover, and R. J. Beckman, "A comparison of three methods for selecting input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, pp. 239–245, 1979.

[19] M. Stein, "Large sample properties of simulations using Latin hypercube sampling," *Technometrics*, vol. 29, pp. 143–151, 1987, correction 32:367.

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