A

Confidence Intervals for Quantiles When Applying Variance-Reduction Techniques

FANG CHU, New Jersey Institute of Technology
MARVIN K. NAKAYAMA, New Jersey Institute of Technology

Quantiles, which are also known as values-at-risk in finance, frequently arise in practice as measures of risk. This paper develops asymptotically valid confidence intervals for quantiles estimated via simulation using variance-reduction techniques (VRTs). We establish our results within a general framework for VRTs, which we show includes importance sampling, stratified sampling, antithetic variates, and control variates. Our method for verifying asymptotic validity is to first demonstrate that a quantile estimator obtained via a VRT within our framework satisfies a Bahadur-Ghosh representation. We then exploit this to show that the quantile estimator obeys a central limit theorem (CLT) and to develop a consistent estimator for the variance constant appearing in the CLT, which enables us to construct a confidence interval. We provide explicit formulae for the estimators for each of the VRTs considered.

Categories and Subject Descriptors: I.6.6 [**Simulation and Modeling**]: Simulation Output Analysis; G.3 [**Probability and Statistics**]: Nonparametric statistics; G.3 [**Probability and Statistics**]: Probabilistic algorithms (including Monte Carlo)

General Terms: Performance, Theory

Additional Key Words and Phrases: Antithetic variates, control variates, importance sampling, quantile, stratified sampling, value-at-risk, variance reduction

1. INTRODUCTION

For $0 < p < 1$, the p-quantile ξ_p of a random variable X is defined as the smallest value x such that $P\{X \leq x\} \geq p$. In terms of the cumulative distribution function (CDF) F of X, we can express the p-quantile as $\xi_p = F^{-1}(p)$. For example, the 0.5-quantile is the median. Quantiles arise in many practical contexts and are sometimes of more interest than means. For example, some internet service providers charge a user based on the 0.95-quantile of the user's traffic load in a billing cycle [Goldenberg et al. 2004]. In project planning, a planner may want to determine a time t such that the project has a 95% chance of completing by t, so $t = \xi_{0.95}$ is the 0.95-quantile. In finance, where a quantile is known as a value-at-risk, an analyst may be interested in the 0.99-quantile $\xi_{0.99}$ of the loss of a portfolio over a certain time period (e.g., two weeks), so there is a 1% chance that the loss over this period will be greater than $\xi_{0.99}$. Value-at-risk is widely used in the financial industry as a measure of portfolio risk; e.g., see Duffie and Pan [1997].

Authors' address: F. Chu, Department of Information Systems, New Jersey Institute of Technology, Newark, NJ 07102; M. K. Nakayama, Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102. This material is based upon work supported in part by the National Science Foundation under Grant Numbers 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.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

Often in practice, the CDF F is unknown or cannot be computed explicitly, but we still may be able to collect samples from F. We thus seek a sampling-based estimator of ξ_p. One complication arises from the fact that a quantile is not a mean (nor a function of a mean) of a random variable, so we cannot estimate a quantile using a sample average. Instead, the following approach (e.g., see Section 2.3 of Serfling [1980]) can be applied. First collect independent and identically distributed (i.i.d.) samples $X_1, X_2, \ldots, X_n$ from distribution F, and use these to construct an estimator of F. One such estimator of F is the *empirical CDF* F_n, where $F_n(x)$ is the fraction of the n samples less than or equal to x. Then the fact that $\xi_p = F^{-1}(p)$ suggests constructing a quantile estimator as $\hat{\xi}_{p,n} = F_n^{-1}(p)$. When applying simulation to generate the i.i.d. samples of X used to construct F_n, we call the method *crude Monte Carlo (CMC)*.

In addition to computing a point estimate for a quantile, it is important to also provide a confidence interval for the quantile as a way of indicating the error in the estimate. A common approach to developing a confidence interval is to first show that the quantile estimator satisfies a central limit theorem (CLT), and then replace the variance constant in the CLT with a consistent estimator of it to construct a confidence interval. For CMC, one can appeal to the CLT for quantile estimators formed from i.i.d. samples; e.g., see Section 2.3.3 of Serfling [1980].

A problem with CMC is that the resulting confidence interval may be large, which is often the case when estimating extreme quantiles (i.e., when p is close to 0 or 1). This motivates applying variance-reduction techniques (VRTs) to try to obtain more efficient quantile estimators; see Chapter 4 of Glasserman [2004] for an overview of VRTs for estimating a mean. There has been some previous work on applying VRTs to estimate a quantile. Hsu and Nelson [1990] and Hesterberg and Nelson [1998] develop quantile estimators using control variates. Avramidis and Wilson [1998] consider quantile estimation with a general class of correlation-induction techniques, which includes antithetic variates and Latin hypercube sampling (LHS). Jin et al. [2003] establish exponential convergence rates for quantile estimators, including those using LHS, and also develop a type of combined stratified-LHS quantile estimator. Glynn [1996] uses importance sampling for quantile estimation, and Glasserman et al. [2000b] combine importance sampling with stratified sampling to estimate value-at-risk. Variance reduction for quantile estimation typically entails applying VRTs to estimate the CDF F and then inverting the resulting CDF estimator $\tilde{F}_n$.

None of the previous work on estimating quantiles using VRTs provides a method to consistently estimate the variance constant κ_p^2 appearing in the CLT for the quantile estimator from which a confidence interval can be formed. It turns out that $\kappa_p = \psi_p / f(\xi_p)$, where ψ_p^2 is the variance constant in the CLT for $\tilde{F}_n(\xi_p)$ and $f(\xi_p)$ is the density function of the (unknown) CDF F evaluated at the (unknown) quantile. Glynn [1996] notes that to construct confidence intervals, "the major challenge is finding a good way of estimating" $f(\xi_p)$, "either explicitly or implicitly," but he does not provide a method for doing this. Indeed, Glasserman et al. [2000b] state (p. 1357) that estimation of $f(\xi_p)$ "is difficult and beyond the scope of this paper." In the case of CMC, several methods for consistently estimating $f(\xi_p)$ (or $1/f(\xi_p)$) have been proposed [Bloch and Gastwirth 1968; Bofinger 1975; Babu 1986], but the consistency proofs do not generalize when using VRTs. (Viewed as a function of $0 < p < 1$, $1/f(\xi_p)$ is sometimes called the *sparsity function* [Tukey 1965] or the *quantile-density function* [Parzen 1979], and these two references discuss its usefulness, apart from quantile estimation, in analyzing data and distributions.)

In our paper we provide a way to consistently estimate $1/f(\xi_p)$ and ψ_p when using VRTs, and taking the product of these estimators yields a consistent estimator of κ_p. This enables us to construct an asymptotically valid confidence interval for the quantile when applying VRTs, which is one of the main contributions of our work. We establish our results within a general framework for VRTs specified by a set of assumptions on the resulting CDF estimator $\tilde{F}_n$. We first prove the quantile estimator resulting from inverting $\tilde{F}_n$ satisfies a weaker form of a so-called Bahadur [1966] representation established by Ghosh [1971], and we call this a Bahadur-Ghosh representation. Also of independent interest, this result shows that the quantile estimator can be approximated by the sum of the true quantile ξ_p and a linear function of the CDF estimator $\tilde{F}_n$ evaluated at ξ_p, with a remainder term vanishing in probability as the sample size n grows. As $\tilde{F}_n(\xi_p)$ is typically a sample average, it correspondingly satisfies a CLT (under appropriate moment conditions). Thus, the Bahadur-Ghosh representation provides insight into why a quantile estimator, which is *not* a sample average, obeys a CLT, which we also prove. We then apply the Bahadur-Ghosh representation to derive a consistent estimator for the asymptotic variance in the CLT for the quantile estimator, leading to a confidence interval for ξ_p. We show that different VRTs, including a combination of importance sampling and stratified sampling, antithetic variates, and control variates, fit in our framework, and we provide formulae for the estimators used to build confidence intervals for quantiles estimated using these VRTs.

Rather than consistently estimating the variance constant from the CLT to develop a confidence interval, one could instead divide all the data into batches, and then produce an interval by constructing a quantile estimate from each batch and computing the sample variance of the (i.i.d.) quantile estimates; e.g., see p. 491 of Glasserman [2004]. However, a drawback of batching is that accurate quantile estimation often requires large sample sizes; e.g., see Avramidis and Wilson [1998]. Thus, it is preferable to have methods that use all of the sampled data to construct a single quantile estimator, as we do. For their control-variate quantile estimator, Hsu and Nelson [1990] instead develop a confidence interval by generalizing a technique for CMC based on the binomial distribution.

The rest of the paper has the following organization. Section 2 discusses quantile estimation and provides the background on the Bahadur-Ghosh representation for CMC. In Section 3 we establish a general framework for proving a Bahadur-Ghosh representation and for developing asymptotically valid confidence intervals for quantiles when applying a generic VRT. We then employ this framework in Sections 4–6 to examine specific VRTs (combined importance sampling and stratified sampling, antithetic variates, and control variates). Our estimator of $1/f(\xi_p)$ is a finite difference, and Section 7 discusses selection of its difference parameter, which we call the bandwidth. Section 8 presents some experimental results, and we provide concluding remarks in Section 9. The appendix contains the proofs of theorems from Section 3. An online-only appendix contains additional empirical results and the other proofs. Chu and Nakayama [2010] present the results without proofs when applying only importance sampling.

2. REVIEW OF QUANTILE ESTIMATION FOR CRUDE MONTE CARLO

Let X be a real-valued random variable with CDF F. For a real-valued function G, define $G^{-1}(a) = \inf\{x : G(x) \geq a\}$. For any $0 < q < 1$, define the q-quantile $\xi_q = F^{-1}(q)$, and we want to compute the p-quantile ξ_p for some fixed p. Suppose F is differentiable at ξ_p and $f(\xi_p) > 0$, where $f(x) = dF(x)/dx$, when it exists.

We will estimate ξ_p using simulation. CMC estimation of ξ_p entails first generating i.i.d. samples $X_1, X_2, \ldots, X_n$ from distribution F. Then we compute the *empirical distribution function* F_n with

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x) \tag{1}$$

as an estimator of $F(x)$, where $I(A)$ is the indicator function of a set A, which assumes value 1 on A and 0 on the complement A^c. We then compute the p-quantile estimator $\hat{\xi}_{p,n} = F_n^{-1}(p)$. An alternative way of computing $\hat{\xi}_{p,n}$ is in terms of order statistics. Sort the samples $X_1, X_2, \ldots, X_n$ into ascending order as $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, where $X_{(i)}$ is the ith smallest of the samples. Then $\hat{\xi}_{p,n} = X_{(\lceil np \rceil)}$, where $\lceil \cdot \rceil$ is the round-up function.

Consider the following heuristic argument. For large sample size n, the Glivenko-Cantelli Theorem (e.g., p. 61 of Serfling [1980]) implies $F_n(x) \approx F(x)$ for all x, so $\hat{\xi}_{p,n} \approx \xi_p$. Thus, since $p = F(\xi_p)$, we have

$$p \approx F(\hat{\xi}_{p,n}) \approx F(\xi_p) + f(\xi_p)(\hat{\xi}_{p,n} - \xi_p) \approx F_n(\xi_p) + f(\xi_p)(\hat{\xi}_{p,n} - \xi_p),$$

where the second step follows from a Taylor approximation, and the last step holds since $F_n(x) \approx F(x)$ for all x. Rearranging terms gives $\hat{\xi}_{p,n} \approx \xi_p - (F_n(\xi_p) - p)/f(\xi_p)$.

Bahadur [1966] makes rigorous the above heuristic argument. Specifically, assuming that $f(\xi_p) > 0$ and that the second derivative of F is bounded in a neighborhood of ξ_p, he proves the following, which is known as a *Bahadur representation*:

$$\hat{\xi}_{p,n} = \xi_p - \frac{F_n(\xi_p) - p}{f(\xi_p)} + R_n, \tag{2}$$

where almost surely (a.s.),

$$R_n = O(n^{-3/4} (\log n)^{1/2} (\log \log n)^{1/4}) \text{ as } n \to \infty. \tag{3}$$

By "$Y_n = O(g(n))$ a.s." we mean that there exists a set Ω_0 such that $P(\Omega_0) = 1$ and for each $\omega \in \Omega_0$, there exists a constant $B(\omega)$ such that $|Y_n(\omega)| \le B(\omega) g(n)$, for n sufficiently large. Kiefer [1967] shows the exact order for R_n is $O(n^{-3/4} (\log \log n)^{3/4})$.

It is well known (e.g., Section 2.3.3 of Serfling [1980]) that $\sqrt{n}(\hat{\xi}_{p,n} - \xi_p)$ converges in distribution as $n \to \infty$ to a normal random variable with mean 0 and variance $p(1-p)/f^2(\xi_p)$, and so does $\sqrt{n}(p - F_n(\xi_p))/f(\xi_p)$ since $F_n(\xi_p)$ is the average of i.i.d. indicator functions. But Bahadur's representation goes further by showing the difference between these two quantities approaches 0 a.s. and provides the rate at which the difference vanishes.

Ghosh [1971] establishes a variation of a weaker form of a Bahadur representation under weaker assumptions. Consider any $p_n = p + O(n^{-1/2})$, and working with perturbed p_n will enable us to develop consistent estimators of the sparsity function $1/f(\xi_p)$, which we use for constructing a confidence interval for ξ_p. Define $\hat{\xi}_{p_n,n} = F_n^{-1}(p_n)$. Assuming only $f(\xi_p) > 0$, Ghosh [1971] shows

$$\hat{\xi}_{p_n,n} = \dot{\xi}_{p_n} - \frac{F_n(\xi_p) - p}{f(\xi_p)} + R_n' \tag{4}$$

with $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$ and R_n' in (4) satisfying

$$\sqrt{n} R_n' \xrightarrow{P} 0 \text{ as } n \to \infty, \tag{5}$$

where $\xrightarrow{P}$ denotes convergence in probability (p. 330 of Billingsley [1995]). If in addition f is continuous in a neighborhood of ξ_p, then (4)–(5) hold with $\dot{\xi}_{p_n} = \xi_{p_n}$ for all $p_n \to p$. We call (4)–(5) a *Bahadur-Ghosh representation*, and this weaker form suffices for most applications, including ours.

One consequence of the Bahadur-Ghosh representation for CMC is that it implies a CLT for the quantile estimator $\hat{\xi}_{p,n} = F_n^{-1}(p)$, as shown in Theorem 10.3 of David and

Nagaraja [2003]. Taking $p_n = p$ in (4), we can write

$$\sqrt{n}(\hat{\xi}_{p,n} - \xi_p) = \frac{\sqrt{n}}{f(\xi_p)}(p - F_n(\xi_p)) + \sqrt{n}R'_n. \tag{6}$$

Let $\xrightarrow{L}$ denote convergence in distribution (Billingsley [1995], Section 25), and define $N(a, b^2)$ as a normal distribution with mean a and variance b^2. That $F_n(\xi_p)$ is the sample average of $I(X_i \le \xi_p)$, $i = 1, 2, \ldots, n$, which are i.i.d. with mean p and variance $p(1-p)$, implies the first term in the right-hand side (RHS) of (6) converges in distribution to $N(0, p(1-p)/f^2(\xi_p))$ as $n \to \infty$. Moreover, the second term on the RHS of (6) vanishes in probability as $n \to \infty$ by (5). Hence, $\sqrt{n}(\hat{\xi}_{p,n} - \xi_p) \xrightarrow{L} N(0, p(1-p)/f^2(\xi_p))$ as $n \to \infty$ by Slutsky's theorem (Serfling [1980], p. 19). Note that the CLT holds under the minimal assumptions of Ghosh [1971] (i.e., $f(\xi_p) > 0$) and does not require the additional conditions that Bahadur [1966] imposes. However, if we instead apply (3) from Bahadur [1966] to obtain (6), then we gain additional information about the exact rate of convergence of $\sqrt{n}R_n$ that is lost in the cruder result of (5) from Ghosh [1971].

3. BAHADUR-GHOSH REPRESENTATION WHEN APPLYING VRTS

CMC is sometimes inefficient for estimating quantiles, so we may try to obtain improved quantile estimators by applying a variance-reduction technique (VRT). VRTs often change the way samples are generated, or collect additional data, and this leads to different estimators of the CDF. We then invert the resulting estimated CDF to obtain a quantile estimator. The VRT estimators of the CDF have more complicated forms than the simple estimator in (1) for CMC, so the methods used to analyze the CMC quantile estimators need to be modified to handle VRTs.

We now establish a general framework under which a Bahadur-Ghosh representation holds. Because the goal in this section is to capture a broad range of simulation settings, the assumptions here are very general. Subsequent sections examine specific VRTs, and we then provide more readily verifiable conditions for importance sampling (IS), stratified sampling (SS), antithetic variates (AV), control variates (CV) and certain combinations of them.

Let $\tilde{F}_n$ denote a generic simulation estimator of the CDF F, where n is the "computational budget," which we define differently for various simulation methods. For example, when applying IS, n is the number of samples generated from a new distribution F_* obtained from a change of measure (see Section 4). In the case of AV, n denotes the number of antithetic pairs (Section 5). For CV, n is the number of pairs of output and control collected (Section 6).

Now set $\tilde{\xi}_{q,n} = \tilde{F}_n^{-1}(q)$ as the estimator of the q-quantile for any $0 < q < 1$. To show that $\tilde{\xi}_{p_n,n}$ with perturbed p_n satisfies a Bahadur-Ghosh representation analogous to (4) and (5), we require that $\tilde{F}_n$ satisfies the following assumptions.

ASSUMPTION A1. *$P(M_n) \to 1$ as $n \to \infty$, where M_n is the event that $\tilde{F}_n(x)$ is monotonically increasing in x.*

This assumption allows for the estimated CDF to not necessarily be monotonically increasing in x, but the probability of this occurring must vanish as n increases. For many (but not all) VRTs, $\tilde{F}_n(x)$ will always be monotonically increasing in x for each n, so Assumption A1 will trivially hold.

ASSUMPTION A2. *For every $a_n = O(n^{-1/2})$,*

$$\sqrt{n}\left[(F(\xi_p + a_n) - F(\xi_p)) - (\tilde{F}_n(\xi_p + a_n) - \tilde{F}_n(\xi_p))\right] \xrightarrow{P} 0$$

as $n \to \infty$.

This assumption requires that the scaled difference in the actual CDF and the CDF estimator, both evaluated over an interval of length of order $n^{-1/2}$ with an endpoint ξ_p, vanishes in probability as $n \to \infty$. We provide a set of sufficient conditions for Assumption A2 in the on-line appendix.

We also require that a CLT holds for the CDF estimator at ξ_p.

ASSUMPTION A3. $\sqrt{n}\left[\tilde{F}_n(\xi_p) - F(\xi_p)\right] \xrightarrow{L} N(0, \psi_p^2)$ *as* $n \to \infty$ *for some* $0 < \psi_p < \infty$.

We will show in the later sections that Assumptions A1–A3 hold under certain conditions for the VRTs we consider. In the case of CMC, where F_n in (1) replaces $\tilde{F}_n$, A1 holds since $F_n(x)$ is monotonically increasing in x for each n. Moreover, Ghosh [1971] (also see David and Nagaraja [2003], p. 287) shows that A2 and A3 hold by exploiting the fact that $nF_n(x)$ has a binomial distribution with parameters n and $F(x)$. Also, for CMC, $\psi_p^2 = \mathrm{Var}[I(X \le \xi_p)] = p(1-p)$ in A3.

The following theorem shows that for a CDF estimator satisfying our assumptions, a Bahadur-Ghosh representation holds for the resulting quantile estimator.

THEOREM 3.1. *Suppose* $\tilde{F}_n$ *satisfies Assumptions A1–A3 and* $f(\xi_p) > 0$. *Then*

(i) For any $p_n = p + O(n^{-1/2})$, *the* p_n*-quantile estimator* $\tilde{\xi}_{p_n,n} = \tilde{F}_n^{-1}(p_n)$ *satisfies*

$$\tilde{\xi}_{p_n,n} = \dot{\xi}_{p_n} - \frac{\tilde{F}_n(\xi_p) - p}{f(\xi_p)} + \tilde{R}_n \tag{7}$$

with $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$ *and*

$$\sqrt{n}\tilde{R}_n \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{8}$$

(ii) If in addition f *is continuous in a neighborhood of* ξ_p *and A2 holds for all* $a_n \to 0$, *then (7)–(8) hold with* $\dot{\xi}_{p_n} = \xi_{p_n}$ *for all* $p_n \to p$.

All of the specific VRTs we consider in subsequent sections satisfy A2 for all $a_n \to 0$. Also, a simple consequence of Theorem 3.1 is that $\tilde{\xi}_{p,n}$ is a consistent estimator of ξ_p, which we can see as follows. Assumption A3 implies $\tilde{F}_n(\xi_p) \xrightarrow{P} F(\xi_p) = p$ by Theorem 2.3.4 of Lehmann [1999]. Hence, (7) ensures

$$\tilde{\xi}_{p,n} = \xi_p + \frac{p - \tilde{F}_n(\xi_p)}{f(\xi_p)} + \tilde{R}_n \xrightarrow{P} \xi_p \tag{9}$$

as $n \to \infty$ by Theorem 3.1 and Slutsky's theorem. Moreover, Theorem 3.1 also implies the following CLT for the VRT estimator of the quantile.

THEOREM 3.2. *Suppose* $\tilde{F}_n$ *satisfies Assumptions A1–A3. If* $f(\xi_p) > 0$, *then*

$$\frac{\sqrt{n}}{\kappa_p}(\tilde{\xi}_{p,n} - \xi_p) \xrightarrow{L} N(0,1) \tag{10}$$

as $n \to \infty$, *where* $\kappa_p = \psi_p \phi_p$, ψ_p *is defined in Assumption A3, and* $\phi_p = 1/f(\xi_p)$.

When quantiles are estimated by applying CMC or a VRT to estimate the CDF and then inverting the CDF estimator, the constant κ_p in (10) always has the form $\kappa_p = \psi_p \phi_p$, where ψ_p is from Assumption A3. The value of ψ_p depends on the particular VRT (and equals $\sqrt{p(1-p)}$ for CMC), but ϕ_p does not change. Thus, efficient quantile estimation typically focuses on applying a VRT to reduce ψ_p.

The CLT in Theorem 3.2 provides a way to construct confidence intervals for a quantile estimated with VRTs, if we have consistent estimators of ψ_p and ϕ_p. To handle ϕ_p, first note $\frac{d}{dp}F^{-1}(p) = 1/f(\xi_p) = \phi_p$ by the chain rule of differentiation, and we propose the following finite-difference estimators (e.g., Section 7.1 of Glasserman [2004]) of ϕ_p:

$$\tilde{\phi}_{p,n,1}(h_n) = \frac{\tilde{F}_n^{-1}(p+h_n) - \tilde{F}_n^{-1}(p)}{h_n}, \tag{11}$$

$$\tilde{\phi}_{p,n,2}(h_n) = \frac{\tilde{F}_n^{-1}(p+h_n) - \tilde{F}_n^{-1}(p-h_n)}{2h_n}, \tag{12}$$

where $h_n \neq 0$ is called the *bandwidth* or *smoothing parameter*. Note that $\tilde{\phi}_{p,n,1}(h_n)$ is a forward (resp., backward) finite-difference estimator when $h_n > 0$ (resp., $h_n < 0$), and $\tilde{\phi}_{p,n,2}(h_n)$ is a central finite difference. To define additional estimators of ϕ_p, let $h_{n,1}, \ldots, h_{n,r}$ and $w_1, \ldots, w_r$ for any $r \geq 1$ be nonzero constants (some possibly negative) with $\sum_{j=1}^{r} w_j = 1$. Then define estimators

$$\bar{\phi}_{p,n,i}(h_{n,1}, \ldots, h_{n,r}, w_1, \ldots, w_r) = \sum_{j=1}^{r} w_j\, \tilde{\phi}_{p,n,i}(h_{n,j}), \text{ for } i = 1, 2, \tag{13}$$

which are weighted combinations of the previous finite-difference estimators. Babu [1986] and p. 384 of Glasserman [2004] discuss combined estimators of this form, and they both provide some discussion on the selection of the $h_{n,j}$ and w_j; also see Section 7. The following theorem shows that all of our estimators of ϕ_p are consistent. In addition, if we also have a consistent estimator $\tilde{\psi}_{p,n}$ of ψ_p in Assumption A3, then we can consistently estimate $\kappa_p = \psi_p\phi_p$ in (10), and the CLT in (10) still holds when κ_p is replaced by its consistent estimator. Thus, we can construct a confidence interval for ξ_p. Let $z_\beta = \Phi^{-1}(1-\beta/2)$ and Φ is the CDF of a $N(0,1)$ random variable.

THEOREM 3.3. *Suppose $\tilde{F}_n$ satisfies Assumptions A1–A3 and $f(\xi_p) > 0$. Then*

(i) $\tilde{\phi}_{p,n} \to \phi_p$ as $n \to \infty$, where $\tilde{\phi}_{p,n}$ is any of the estimators in (11)–(13) with

$$h_n = \frac{c}{\sqrt{n}},\ h_{n,j} = \frac{c_j}{\sqrt{n}}, \text{ for any nonzero constants } c,\ c_j,\ j = 1, \ldots, r, \tag{14}$$

and w_j, $j = 1, \ldots, r$, are nonzero constants with $\sum_{j=1}^{r} w_j = 1$. In addition, if $\tilde{\psi}_{p,n} \xrightarrow{P} \psi_p$ as $n \to \infty$, then

$$\frac{\sqrt{n}}{\tilde{\kappa}_{p,n}}(\tilde{\xi}_{p,n} - \xi_p) \xrightarrow{L} N(0,1) \tag{15}$$

as $n \to \infty$, with $\tilde{\kappa}_{p,n} = \tilde{\psi}_{p,n}\tilde{\phi}_{p,n}$. Thus, $[\tilde{\xi}_{p,n} \pm z_\alpha \tilde{\kappa}_{p,n}/\sqrt{n}]$ is an asymptotically valid $100(1-\alpha)\%$ confidence interval for ξ_p.

(ii) If in addition f is continuous in a neighborhood of ξ_p and Assumption A2 holds for all $a_n \to 0$, then all of the above holds for any h_n and $h_{n,j}$ satisfying

$$h_n \to 0,\ 1/h_n = O(\sqrt{n}), \text{ and } h_{n,j} \to 0,\ 1/h_{n,j} = O(\sqrt{n}),\ j = 1, \ldots, r. \tag{16}$$

Hong [2009], Liu and Hong [2009] and Fu et al. [2009] develop consistent estimators for derivatives of quantiles with respect to certain model parameters θ, but their methods do not apply for estimating $\phi_p = \frac{d}{dp}F^{-1}(p)$ (nor when using VRTs). These papers assume that the (random) output $X = X(\theta)$ and its distribution $F(\cdot) = F(\cdot\,;\theta)$ depend on a parameter θ with respect to which the derivative of the quantile is taken. However, in our case, neither X nor F change when $\theta = p$ is varied, so our problem does not fit in their framework.

When applying CMC (i.e., i.i.d. sampling), Bloch and Gastwirth [1968] and Bofinger [1975] show that estimators analogous to $\tilde{\phi}_{p,n,i}(h_n)$, $i = 1, 2$, in (11) and (12) consistently estimate ϕ_p. Also, Babu [1986] considers estimators that are weighted combinations as in (13) for i.i.d. sampling. All their consistency proofs rely on representing each i.i.d. sample X_i as $X_i = F^{-1}(U_i)$, where U_i is uniformly distributed on $[0, 1]$. However, these arguments do not generalize when applying VRTs, so we require a different approach to establish the first result in Theorem 3.3(i). In particular, Corollary 2.5.2 of Serfling [1980] provides a method that exploits an a.s. Bahadur representation as in (2)–(3) but with perturbed p_n rather than fixed p for i.i.d. sampling to consistently estimate ϕ_p, and we modify this idea to work instead with a Bahadur-Ghosh representation and VRTs.

The following sections give explicit formulae for $\tilde{F}_n$ and $\tilde{\psi}_{p,n}$ for different VRTs, and we can use these estimators to compute $\tilde{\xi}_{p,n}$ and $\tilde{\phi}_{p,n}$ to build the confidence interval for ξ_p in Theorem 3.3. To simplify notation, we will continue to use the same variables $\tilde{F}_n$, $\tilde{\xi}_{p,n}$, $\tilde{\kappa}_{p,n}$, $\tilde{\psi}_{p,n}$ and $\tilde{\phi}_{p,n}$ in each case rather than develop new notation for each VRT.

4. IMPORTANCE SAMPLING AND STRATIFIED SAMPLING

IS (Section 4.6 of Glasserman [2004]) and SS (Section 4.3 of Glasserman [2004]) are two VRTs used to improve the efficiency of simulations, and combining them may further enhance the effect. Before describing a combined IS+SS quantile estimator developed by Glasserman et al. [2000b], we start by applying just IS alone without SS, as in Glynn [1996].

We first explain how to apply IS in the simple case when the output X has CDF F and density function f. Let F_* be another CDF, and let f_* be the density function of F_* with the property that for each t, $f(t) > 0$ implies that $f_*(t) > 0$. Define E_* to be expectation under CDF F_*. Also, define $L(t) = f(t)/f_*(t)$ to be the likelihood ratio at t. Then we can write

$$F(x) = \int I(t \le x)\, f(t)\, dt = \int I(t \le x)\, L(t)\, f_*(t)\, dt = E_*\left[I(X \le x)\, L(X)\right].$$

The above suggests that to estimate $F(x)$ using IS, we generate i.i.d. samples $X_1, \ldots, X_n$ of X from CDF F_* and average $I(X_i \le x)L(X_i)$, $i = 1, \ldots, n$.

As explained in Glynn and Iglehart [1989], IS applies more generally than the situation we just described. Let P be the original probability measure governing the stochastic system or process being studied, and let P_* be another probability measure such that for each (measurable) event A, $P(A) > 0$ implies $P_*(A) > 0$; i.e., P is absolutely continuous (p. 422 of Billingsley [1999]) with respect to P_*. Define E_* as the expectation operator under the IS probability measure P_*, and define the likelihood ratio $L = dP/dP_*$, which is also called the Radon-Nikodym derivative of P with respect to P_* (p. 423 of Billingsley [1999]). Then we have

$$F(x) = \int I(X \le x)\, dP = \int I(X \le x)\, L\, dP_* = E_*\left[I(X \le x)\, L\right], \tag{17}$$

which is known as applying a change of measure. This motivates estimating $F(x)$ as follows. Generate i.i.d. samples $(X_1, L_1), \ldots, (X_n, L_n)$ of (X, L) using P_*, and the IS estimator of F is then

$$\tilde{F}_{n,\mathrm{IS}}(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x)\, L_i. \tag{18}$$

Inverting $\tilde{F}_{n,\mathrm{IS}}$ results in the IS quantile estimator.

Recall the CMC estimator of F is given in (1), where the X_i in (1) are generated using the original measure P induced by CDF F. By Theorem 3.2, the key to applying IS for estimating ξ_p is choosing P_* so that $\psi_p^2 = \mathrm{Var}_*[I(X \le \xi_p)\, L] < \mathrm{Var}[I(X \le \xi_p)] = p(1-p)$ to achieve a variance reduction (relative to CMC), where Var_* denotes variance under measure P_*. Glynn [1996] and Glasserman et al. [2000b] present particular choices of P_* for various settings.

To additionally incorporate stratified sampling, we identify a *stratification variable* Y such that X and Y are dependent. We partition the support of Y into $k < \infty$ strata $S_1, \ldots, S_k$ such that each $\lambda_i \equiv P_*\{Y \in S_i\} > 0$ is known and $\sum_{i=1}^k \lambda_i = 1$. For example, the strata may be disjoint intervals, and Sections 4.3 and 9.2.3 of Glasserman [2004] discuss this and other possibilities for the strata. Therefore, by (17), we can write

$$F(x) = \sum_{i=1}^{k} \lambda_i E_* \left[I(X \le x)\, L \mid Y \in S_i \right]. \tag{19}$$

Note that we derived (19) by first applying IS and then using stratification, so Y is distributed under the IS measure P_*. (Instead applying SS first and then IS leads to a different representation for F and thus a different estimator; see Glasserman et al. [2000a] for details.)

IS+SS estimation of F entails replacing each conditional expectation in (19) with an average of samples from the corresponding stratum. We now provide details on this approach as developed in Glasserman et al. [2000a]. Define the sample size in each stratum i as $n_i = n\gamma_i$, where the $\gamma_i > 0$ are user-specified constants satisfying $\sum_{i=1}^k \gamma_i = 1$. One possible implementation of IS+SS defines the k strata to be equiprobable (i.e., each $\lambda_i = 1/k$), and also lets each $\gamma_i = \lambda_i$. (Glasserman [2004], pp. 217–218, discusses other possible choices for the γ_i.) For simplicity, we assume that n_i is always an integer, so the total number of samples across all strata is $\sum_{i=1}^k n_i = \sum_{i=1}^k n\gamma_i = n$. For each stratum $i = 1, \ldots, k$, we use the IS measure P_* to draw n_i samples Y_{ij}, $j = 1, \ldots, n_i$, of Y conditioned to lie in S_i. Then for each $j = 1, \ldots, n_i$, generate X_{ij} as a sample of X having the conditional IS distribution of X given $Y = Y_{ij}$, and let L_{ij} be the corresponding likelihood ratio. We thus have n_i i.i.d. samples (X_{ij}, Y_{ij}, L_{ij}), $j = 1, \ldots, n_i$, of (X, Y, L) from stratum i. (Glasserman et al. [2000b] employ a "bin tossing" method to generate samples of the triple (X_{ij}, Y_{ij}, L_{ij}).) The (X_{ij}, Y_{ij}, L_{ij}) sample triples across strata are generated independently. Then the IS+SS estimator of the CDF is

$$\tilde{F}_n(x) = \sum_{i=1}^{k} \lambda_i \frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{ij} \le x)\, L_{ij}. \tag{20}$$

We allow for $P_* = P$, in which case the likelihood ratio $L \equiv 1$ and we do not apply IS. Also, the number of strata may be $k = 1$, in which case there is no stratified sampling. Hence, the following result establishing Bahadur-Ghosh representations and CLTs for IS+SS encompasses CMC, IS-only and SS-only as special cases.

THEOREM 4.1. *Suppose $f(\xi_p) > 0$, and for each stratum i, suppose there exists $\epsilon > 0$ and $\delta > 0$ such that $E_*[I(X_{ij} < \xi_p + \delta) L_{ij}^{2+\epsilon}] < \infty$. Let $\tilde{F}_n$ be the IS+SS estimator of F defined in (20), and let $\tilde{\xi}_{q,n} = \tilde{F}_n^{-1}(q)$ be the q-quantile estimator for any $0 < q < 1$. Then*

(i) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$ for any $p_n = p + O(n^{-1/2})$;

(ii) CLTs (10) and (15) hold for h_n and $h_{n,j}$ satisfying (14), where $\psi_p^2 = \sum_{i=1}^k \lambda_i^2 \zeta_i^2/\gamma_i$ is the variance constant in Assumption A3 with

$$\zeta_i^2 = E_* \left[I(X_{ij} \le \xi_p) L_{ij}^2 \right] - P^2(X \le \xi_p \,|\, Y \in S_i), \tag{21}$$

and a consistent estimator of ψ_p^2 is $\tilde{\psi}_{p,n}^2 = \sum_{i=1}^k \lambda_i^2 \tilde{\zeta}_{i,n}^2/\gamma_i$ with

$$\tilde{\zeta}_{i,n}^2 = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{ij} \le \tilde{\xi}_{p,n}) L_{ij}^2 \right) - \left(\frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{ij} \le \tilde{\xi}_{p,n}) L_{ij} \right)^2 . \tag{22}$$

If in addition f is continuous in a neighborhood of ξ_p, then

(iii) (7)–(8) hold with $\dot{\xi}_{p_n} = F^{-1}(p_n)$ for all $p_n \to p$;
(iv) (15) holds for h_n and $h_{n,j}$ satisfying (16).

The consistency proof of $\tilde{\psi}_{p,n}^2$ is complicated by the fact that the two terms in (22) are not sums of independent quantities. Each summand depends on $\tilde{\xi}_{p,n}$, which is a function of all the samples, making the summands dependent.

We recently found out that independently of our work, Sun and Hong [2010] establish that the IS-only quantile estimator obtained by inverting $\tilde{F}_{n,\mathrm{IS}}$ in (18) satisfies an a.s. Bahadur representation analogous to (2) and (3) using a different proof technique and under a stronger set of assumptions than we use. Specifically, they further assume that the density f is positive and continuously differentiable in a neighborhood of ξ_p and that the likelihood ratio $L(x)$ is bounded in a neighborhood of ξ_p. Also, they do not consider IS+SS (nor AV and CV), as we do. Moreover, they examine only the case of fixed p and not perturbed p_n, the latter of which is essential for our approach for developing confidence intervals for ξ_p.

The right tail of $\tilde{F}_{n,\mathrm{IS}}$ in (18) may not behave as a proper CDF since it is possible (and indeed likely) for $\lim_{x\to\infty} \tilde{F}_{n,\mathrm{IS}}(x) = a$ with $a < 1$ or $a > 1$. To avoid such a situation, Glynn [1996] also proposes another IS estimator of the CDF, $\tilde{F}'_{n,\mathrm{IS}}(x) = 1 - \frac{1}{n}\sum_{i=1}^n I(X_i > x) L(X_i)$, which can be more effective when estimating a quantile for $p \approx 1$. (However, we may instead have $\lim_{x\to-\infty} \tilde{F}'_{n,\mathrm{IS}}(x) = b$ with $b < 0$ or $b > 0$, so $\tilde{F}'_{n,\mathrm{IS}}$ may not be appropriate when estimating a quantile for $p \approx 0$.) Glasserman et al. [2000b] develop the corresponding IS+SS estimator of F:

$$\tilde{F}'_n(x) = 1 - \sum_{i=1}^k \lambda_i \frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{ij} > x) L_{ij}. \tag{23}$$

The following theorem, in which primed variables replace non-primed variables from before, shows that quantile estimators based on inverting $\tilde{F}'_n$ satisfy a Bahadur-Ghosh representation and CLTs. It can be established by straightforward modifications of the proof of Theorem 4.1.

THEOREM 4.2. *Suppose $f(\xi_p) > 0$, and for each stratum i, suppose there exists $\epsilon > 0$ and $\delta > 0$ such that $E_*[I(X_{ij} > \xi_p - \delta) L_{ij}^{2+\epsilon}] < \infty$. Let $\tilde{F}'_n$ be the IS+SS estimator of F defined in (23), and let $\tilde{\xi}'_{q,n} = \tilde{F}'^{-1}_n(q)$ be the q-quantile estimator for any $0 < q < 1$. Then*

(i) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$ for any $p_n = p + O(n^{-1/2})$;

(ii) CLTs (10) and (15) hold for h_n and $h_{n,j}$ satisfying (14), where $\psi_p'^2 = \sum_{i=1}^k \lambda_i^2 \zeta_i'^2/\gamma_i$ is the variance constant in Assumption A3 with $\zeta_i'^2 = E_*\left[I(X_{ij} > \xi_p) L_{ij}^2\right] - P^2(X > \xi_p \,|\, Y \in S_i)$, and a consistent estimator of $\psi_p'^2$ is $\tilde{\psi}_{p,n}'^2 = \sum_{i=1}^k \lambda_i^2 \tilde{\zeta}_{i,n}'^2/\gamma_i$ with $\tilde{\zeta}_{i,n}'^2 = \left(\frac{1}{n_i}\sum_{j=1}^{n_i} I(X_{ij} > \tilde{\xi}_{p,n}')L_{ij}^2\right) - \left(\frac{1}{n_i}\sum_{j=1}^{n_i} I(X_{ij} > \tilde{\xi}_{p,n}')L_{ij}\right)^2$.

If in addition f is continuous in a neighborhood of ξ_p, then

(iii) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_{p_n} = F^{-1}(p_n)$ for all $p_n \to p$;
(iv) (15) holds for h_n and $h_{n,j}$ satisfying (16).

Glasserman et al. [2000b] also prove that the quantile estimator $\tilde{\xi}_{p,n}'$ obtained using IS+SS satisfies the CLT in (10) (but they do not consider the CLT in (15) with estimated variance). Also, Glynn [1996] establishes CLTs analogous to (10) (but not (15) with estimated variance) for the IS-only quantile estimator $\tilde{\xi}_{p,n}$ equal to $\tilde{F}_{n,\mathrm{IS}}^{-1}(p)$ or $\tilde{F}_{n,\mathrm{IS}}'^{-1}(p)$. In their proofs both Glasserman et al. [2000b] and Glynn [1996] apply a version of the Berry-Esséen theorem that requires a finite third absolute moment (e.g., p. 33 of Serfling [1980]), and consequently, they assume the likelihood ratio L_{ij} has a finite third moment (under the IS measure). Our proof of (10) uses a different approach employing the Bahadur-Ghosh representation, allowing us to relax the moment condition on the likelihood ratio to instead require $E_*[I(X_{ij} \le \xi_p + \delta)L_{ij}^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ and $\delta > 0$. We can modify the proofs of Glasserman et al. [2000b] and Glynn [1996] to relax their assumption to be the same as ours by instead applying a version of the Berry-Esséen theorem that requires a finite moment of only order $2+\epsilon$; see Theorem 5.7 of Petrov [1995]. However, in this case the bound on the supremum of the difference of the normalized CDF estimator and the standard normal CDF converges only as $O(n^{-\epsilon/2})$ instead of $O(n^{-1/2})$ as in the Berry-Esséen theorem requiring a finite third moment. Thus, the convergence to the normal distribution may be extremely slow for both our and their CLTs when the $2+\epsilon$ moment with $\epsilon < 1$ is finite but not the third.

Under the stronger assumptions discussed earlier, Sun and Hong [2010] establish that the IS-only quantile estimator $\tilde{F}_{n,\mathrm{IS}}^{-1}(p)$ obeys the CLT in (10), but they do not consider (15) with estimated variance nor IS+SS.

We now explain how to invert the estimated CDF $\tilde{F}_n$ in (20). We prespecify λ_i and n_i; after completing sampling, we have the X_{ij} and L_{ij}. Recall the total number of samples across all strata is n. For each $i = 1, \ldots, k$, and $j = 1, \ldots, n_i$, define $A_m = X_{ij}$ and $B_m = L_{ij}\lambda_i/n_i$, where $m = \sum_{\ell=1}^{i-1} n_\ell + j$. Then sort $A_1, A_2, \ldots, A_n$ in ascending order as $A_{(1)} \le A_{(2)} \le \cdots \le A_{(n)}$, and let $B^{(i)}$ correspond to $A_{(i)}$. For fixed $0 < q < 1$ and $\tilde{F}_n$ in (20), define the q-quantile estimator $\tilde{\xi}_{q,n}$ to be $\tilde{F}_n^{-1}(q) = A_{(i_q)}$, where i_q is the smallest integer for which $\sum_{m=1}^{i_q} B^{(m)} \ge q$. Similarly, for the estimated CDF $\tilde{F}_n'$ in (23), we compute the q-quantile estimator $\tilde{\xi}_{q,n}'$ to be $\tilde{F}_n'^{-1}(q) = A_{(i_q')}$, where i_q' is the smallest integer for which $\sum_{m=i_q'+1}^{n} B^{(m)} \le 1-q$.

5. ANTITHETIC VARIATES

In the case of estimating the mean of a random output X having CDF F, the basic idea of AV (Section 4.2 of Glasserman [2004]) is to generate two copies X and X' of the output having CDF F in such a way that X and X' are negatively correlated (we discuss below how this may be done), and we average the two outputs. Since $\mathrm{Var}((X+X')/2) = [\mathrm{Var}(X) + \mathrm{Cov}(X,X')]/2 \le \mathrm{Var}(X)/2$ when $\mathrm{Cov}(X,X') \le 0$, AV reduces variance compared to when X and X' are independent.

As in Avramidis and Wilson [1998] we implement AV to estimate a quantile of output X by generating (X_i, X_i'), $i = 1, 2, \dots, n$, as i.i.d. antithetic pairs, where X_i and X_i' each have marginal distribution F and X_i and X_i' are negatively correlated. Then the AV estimator of the CDF F is

$$\tilde{F}_n(x) = \frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x) + \frac{1}{n}\sum_{i=1}^{n} I(X_i' \le x)\right] = \frac{1}{2}\left[F_n(x) + F_n'(x)\right], \tag{24}$$

where F_n is defined in (1) and $F_n' = (1/n)\sum_{i=1}^{n} I(X_i' \le x)$. Inverting $\tilde{F}_n$ yields the AV quantile estimator, which satisfies a Bahadur-Ghosh representation and CLTs.

THEOREM 5.1. *Suppose* $f(\xi_p) > 0$. *Let* $\tilde{F}_n$ *be the AV estimator of* F *defined in (24), and let* $\tilde{\xi}_{q,n} = \tilde{F}_n^{-1}(q)$ *be the* q*-quantile estimator for any* $0 < q < 1$. *Then*

(i) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$ *for any* $p_n = p + O(n^{-1/2})$*;*
(ii) CLTs (10) and (15) hold for h_n *and* $h_{n,j}$ *satisfying (14), where*

$$\psi_p^2 = \frac{1}{2}\left[p(1-2p) + P\{X \le \xi_p, X' \le \xi_p\}\right] \tag{25}$$

is the variance constant in Assumption A3 with (X, X') *an antithetic pair, and*

$$\tilde{\psi}_{p,n}^2 = \frac{1}{2}\left[p(1-2p) + \frac{1}{n}\sum_{i=1}^{n} I(X_i \le \tilde{\xi}_{p,n}, X_i' \le \tilde{\xi}_{p,n})\right]. \tag{26}$$

If in addition f *is continuous in a neighborhood of* ξ_p*, then*

(iii) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_{p_n} = F^{-1}(p_n)$ *for all* $p_n \to p$*;*
(iv) (15) holds for h_n *and* $h_{n,j}$ *satisfying (16).*

We now describe how to invert the AV CDF estimator $\tilde{F}_n$ in (24). For each $i = 1, \dots, n$, define $A_{2i-1} = X_i$ and $A_{2i} = X_i'$. Then sort $A_1, A_2, \dots, A_{2n}$ in ascending order as $A_{(1)} \le A_{(2)} \le \cdots \le A_{(2n)}$. For fixed $0 < q < 1$, define the q-quantile estimator $\tilde{\xi}_{q,n}$ to be $\tilde{F}_n^{-1}(q) = A_{(\lceil 2nq \rceil)}$.

There are various ways in which we can generate negatively correlated X and X' with the same marginal distribution F. For example, suppose that the output X can be expressed as $X = g(U_1, \dots, U_d)$ for some function g, where $U_1, \dots, U_d$ are i.i.d. uniform random variables on $[0, 1]$. Then $X' = g(1 - U_1, \dots, 1 - U_d)$ has the same distribution as X since $1 - U_i$ is also uniform on $[0, 1]$. If g is monotonic in each of its arguments, it follows that $I(X \le \xi_p)$ is monotonic in each U_i, so $I(X \le \xi_p)$ and $I(X' \le \xi_p)$ are negatively correlated (p. 181 of Ross [1997]). This ensures that ψ_p^2 in (25) satisfies $\psi_p^2 \le p(1-p)/2$, so AV with n antithetic pairs asymptotically reduces variance compared to CMC with $2n$ independent samples.

6. CONTROL VARIATES

In many simulations, one often knows the mean of an auxiliary random variable that is generated in the process of generating the output X. For example, in a queueing simulation one knows the distributions (so also the means) of the interarrival times and service times. The method of control variates (CV; Section 4.1 of Glasserman [2004]) reduces variance by exploiting this knowledge. We now describe a CV quantile estimator, as developed by Hesterberg and Nelson [1998].

Suppose that (X, C) is a correlated pair of random variables, where we are interested in estimating the p-quantile of X. We assume that C has known mean ν and finite variance, and we will use C as a *control variate*. Since $X' \equiv I(X \le x) - \beta(C - \nu)$ has mean $F(x)$ for any constant β, we can average i.i.d. samples of X' to obtain an unbiased estimator of $F(x)$. Specifically, we generate i.i.d. samples (X_i, C_i), $i = 1, 2, \ldots, n$, of the pair (X, C). Letting β be any constant, we can define an unbiased estimator of the CDF F of X as

$$\tilde{F}'_{n,\beta}(x) = \frac{1}{n}\sum_{i=1}^{n}[I(X_i \le x) - \beta(C_i - \nu)] = F_n(x) - \beta(\bar{C}_n - \nu), \tag{27}$$

where F_n is the empirical CDF defined in (1) and $\bar{C}_n = (1/n)\sum_{i=1}^{n} C_i$.

Clearly, the variance of $\tilde{F}'_{n,\beta}(x)$ depends on the value of β. It can be shown (e.g., p. 186 of Glasserman [2004]) that the choice of β that minimizes the variance is $\beta_*(x) = \text{Cov}[I(X \le x), C]/\text{Var}[C]$, which depends on x. However, one typically does not know the value of $\text{Cov}[I(X \le x), C]$, so it must be estimated. We thus estimate $\beta_*(x)$ via

$$\hat{\beta}_n(x) = \frac{[(1/n)\sum_{i=1}^{n} I(X_i \le x)C_i] - F_n(x)\bar{C}_n}{(1/n)\sum_{j=1}^{n}(C_j - \bar{C}_n)^2}. \tag{28}$$

Replacing β in (27) with $\hat{\beta}_n(x)$ gives us the CV estimator of the CDF F as

$$\tilde{F}_n(x) = F_n(x) - \hat{\beta}_n(x)(\bar{C}_n - \nu), \tag{29}$$

which is typically no longer unbiased because of the correlation of $\hat{\beta}_n(x)$ and $\bar{C}_n$. We obtain the CV quantile estimator by inverting $\tilde{F}_n$ in (29), and the following shows the quantile estimator satisfies a Bahadur-Ghosh representation and CLTs.

THEOREM 6.1. *Suppose $f(\xi_p) > 0$, and let C be a control variate with $0 < Var[C] < \infty$. Let $\tilde{F}_n$ be the CV estimator of F defined in (29), and let $\tilde{\xi}_{q,n} = \tilde{F}_n^{-1}(q)$ be the q-quantile estimator for any $0 < q < 1$. Then*

(i) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$ for any $p_n = p + O(n^{-1/2})$;
(ii) CLTs (10) and (15) hold for h_n and $h_{n,j}$ satisfying (14), where

$$\psi_p^2 = p(1-p) - \frac{(Cov[I(X \le \xi_p), C])^2}{Var[C]} \tag{30}$$

is the variance constant in Assumption A3, and

$$\tilde{\psi}_{p,n}^2 = p(1-p) - \frac{([(1/n)\sum_{i=1}^{n} I(X_i \le x)C_i] - F_n(x)\bar{C}_n)^2}{(1/n)\sum_{j=1}^{n}(C_j - \bar{C}_n)^2}.$$

If in addition f is continuous in a neighborhood of ξ_p, then

(iii) (7)–(8) hold with $\dot{\xi}_{p_n} = \xi_{p_n} = F^{-1}(p_n)$ for all $p_n \to p$;
(iv) (15) holds for h_n and $h_{n,j}$ satisfying (16).

Inverting $\tilde{F}_n$ in (29) initially appears complicated by the fact that $\hat{\beta}_n(x)$ depends on x. However, Hesterberg and Nelson [1998] show that $\tilde{F}_n(x)$ can be rewritten as

$$\tilde{F}_n(x) = \sum_{i=1}^{n} H_i I(X_i \le x), \tag{31}$$

where

$$H_i = \frac{1}{n} + \frac{(\bar{C}_n - C_i)(\bar{C}_n - \nu)}{\sum_{j=1}^{n}(C_j - \bar{C}_n)^2}, \tag{32}$$

which does not depend on x and $\sum_{i=1}^{n} H_i = 1$. (If all the C_j are equal, then $H_i = 1/n$.) If $H_i \geq 0$ for each i, then it is clear from (31) that $\tilde{F}_n(x)$ is monotonically increasing in x. (It is possible for H_i to be negative, but Hesterberg and Nelson [1998] note that it is unlikely, in a sense they make precise.) The advantage of the representation of $\tilde{F}_n$ in (31) is that it allows evaluating $\tilde{F}_n(x)$ at different values of x without needing to recompute $\hat{\beta}_n(x)$ in (28) each time. Also, we can invert $\tilde{F}_n$ as follows. We first sort $X_1, X_2, \ldots, X_n$ in ascending order as $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, and let $H^{(i)}$ correspond to $X_{(i)}$. Then for $0 < q < 1$, we can compute the q-quantile estimator $\tilde{\xi}_{q,n}$ as $\tilde{F}_n^{-1}(q) = X_{(i_q)}$, where $i_q = \min\left\{ j : \sum_{i=1}^{j} H^{(i)} \geq q \right\}$.

We now discuss a particular choice for control variate C. Suppose that Y is a random variable correlated with the output X, and let G be the marginal CDF of Y. We can then define $C = I(Y \leq y)$ for some constant y when $G(y) = E[C]$ is known with $0 < G(y) < 1$. Thus, we collect i.i.d. pairs (X_i, Y_i), $i = 1, \ldots, n$, and define $C_i = I(Y_i \leq y)$. It is straightforward to show in this case that $H_i \geq 0$ always holds. Moreover, Hesterberg and Nelson [1998] note that H_i in (32) becomes $H_i = P(Y \leq y)/\sum_{i=1}^{n} I(Y_i \leq y)$ if $Y_i \leq y$, and $H_i = P(Y > y)/\sum_{i=1}^{n} I(Y_i > y)$ if $Y_i > y$. When estimating the p-quantile $\xi_p = F^{-1}(p)$ of X, a natural choice for y is $y = G^{-1}(p)$ (assuming this is known), but this is not required.

We can also apply CV with multiple controls $C^{(1)}, C^{(2)}, \ldots, C^{(m)}$, where each $C^{(j)}$ has known mean; see Hesterberg and Nelson [1998] for details. One possible choice is to specify constants $y_1, \ldots, y_m$ for which each $G(y_j)$ is known and set $C^{(j)} = I(Y \leq y_j)$.

7. CHOOSING THE BANDWIDTH

Recall $\tilde{\phi}_{p,n,2}(h_n)$ from (12), the central finite-difference estimator of ϕ_p, where n is the computational budget. The estimator $\tilde{\phi}_{p,n,2}(h_n)$ scales the difference of quantile estimators at $p + h_n$ and $p - h_n$ for bandwidth $h_n \neq 0$, and we can use $\tilde{\phi}_{p,n,2}(h_n)$ to construct the $100(1-\alpha)\%$ confidence interval for ξ_p in Theorem 3.3. We now discuss the choice of h_n when Assumption A2 holds for all $a_n \to 0$ (which is the case for IS+SS, AV and CV) and f is continuous in a neighborhood of ξ_p, in which case we can choose h_n to satisfy (16).

In the case of CMC, there has been some previous work guiding the selection of h_n. Bloch and Gastwirth [1968] and Bofinger [1975] show that under certain regularity conditions, taking h_n of order $n^{-1/5}$ asymptotically minimizes the mean-square error (MSE) of $\tilde{\phi}_{p,n,2}(h_n)$ as an estimator of ϕ_p. Hall and Sheather [1988] show, under certain conditions, the coverage error of the resulting confidence interval for ξ_p is asymptotically minimized by using h_n of order $n^{-1/3}$. In each case, there is a constant $c \neq 0$ such that $h_n = cn^{-v}$ is asymptotically optimal, where c depends on the underlying CDF F and p. These authors present some data-based approaches for estimating c when applying CMC.

When combining different values of h_n as in (13), we can use the following idea from p. 384 of Glasserman [2004] for reducing bias of finite-difference estimators. Let $Q(x) = F^{-1}(x)$, so $\phi_p = Q'(p)$, where prime denotes derivative. A Taylor expansion of $Q(p)$ (assuming enough derivatives exist) yields

$$\frac{Q(p+h_n) - Q(p-h_n)}{2h_n} = Q'(p) + \frac{1}{6}Q'''(p)h_n^2 + O(h_n^4),$$

where odd powers of h_n cancel out. Similarly,

$$\frac{Q(p+2h_n) - Q(p-2h_n)}{4h_n} = Q'(p) + \frac{2}{3}Q'''(p)h_n^2 + O(h_n^4),$$

so we can cancel out the order h_n^2 term by combining the two above results as

$$\frac{4}{3}\left(\frac{Q(p+h_n)-Q(p-h_n)}{2h_n}\right)-\frac{1}{3}\left(\frac{Q(p+2h_n)-Q(p-2h_n)}{4h_n}\right)=Q'(p)+O(h_n^4).$$

This suggests that when combining $r = 2$ values of h_n in (13), selecting $h_{n,1}$ and $h_{n,2}$ with $h_{n,2} = 2h_{n,1}$, $w_1 = 4/3$ and $w_2 = -1/3$ may lead to the estimator $\bar{\phi}_{p,n,2}(h_{n,1}, h_{n,2}, w_1, w_2)$ having low bias.

8. EXPERIMENTAL RESULTS

The previous sections developed confidence intervals (CIs) for a quantile of a random variable X having CDF F, and we established the asymptotic validity of the intervals as the computational budget $n \to \infty$. However, in practice, only finite sample sizes can be used, so we now carry out an empirical study to see how well the CIs perform with finite n. Our experiments entail applying CMC, IS+SS, AV, and CV on a stochastic activity network (SAN) described in Section 8.1. We give experimental results with a small version of the model in Section 8.3 and a larger SAN in the online appendix. The goal is to study how the computational budget n and the bandwidth h_n used in the finite-difference estimators in (11) and (12) affect the coverage of the CIs. (We also ran experiments using IS-only and on a normal distribution (IS-only, AV, CMC) and a bivariate normal (IS+SS, CV), but the results are similar, so we do not include them here; Chu [2010] and Chu and Nakayama [2010] give those results.)

8.1. Stochastic Activity Network

SANs are often employed to model the time to complete a project and are useful in project planning. Consider a SAN with d activities. Let $A_1, A_2, \ldots, A_d$ be the durations of the d activities, which are mutually independent. Let f_i (resp., F_i) denote the density (resp., CDF) of A_i. There are q paths in the network, and let B_j be the set of activities on the jth path. Let $T_j = \sum_{i\in B_j} A_i$ be the (random) length of path j, and let $\mu_j = \sum_{i\in B_j} E[A_i]$ be its mean. Let $X = \max_{j=1,\ldots,q} T_j$ be the length of the longest path. Our goal is to estimate and construct CIs for the p-quantile ξ_p of X.

8.1.1. CV for SAN. We chose the control C for CV as follows. Let Y be the (random) length of the path j for which μ_j is the largest. Let G denote the CDF of Y, and we take $C = I(Y \le G^{-1}(p))$ when estimating the p-quantile of X.

8.1.2. AV for SAN. We generate the antithetic pair (X, X') as follows. Let $U_1, \ldots, U_d$ be i.i.d. unif$[0, 1]$, and for each $i = 1, \ldots, d$, set $A_i = F_i^{-1}(U_i)$ and $A'_i = F_i^{-1}(1 - U_i)$. Then for $j = 1, 2, \ldots, q$, set $T_j = \sum_{i\in B_j} A_i$ and $T'_j = \sum_{i\in B_j} A'_i$, and let $X = \max(T_1, \ldots, T_q)$ and $X' = \max(T'_1, \ldots, T'_q)$.

8.1.3. IS+SS for SAN. We now describe how IS+SS is implemented on the SAN. Since our experiments focus on $p \approx 1$, we employ the IS+SS quantile estimator in Theorem 4.2. We apply IS as in Juneja et al. [2007] by using a mixture of exponentially tilted distributions, and SS takes Y defined in Section 8.1.1 as the stratification variable. Define f_i^θ to be the exponentially tilted version of f_i under tilting parameter θ, so $f_i^\theta(t) = e^{\theta t - \chi_i(\theta)} f_i(t)$, where $\chi_i(\theta) = \ln E[e^{\theta A_i}]$ is the cumulant generating function (CGF) of A_i.

We will define the IS distribution to be a mixture of q distributions, each defined by exponentially tilting one path length T_j and not changing the distributions of activities not on that path. To do this, define positive mixture weights $\alpha_1, \ldots, \alpha_q$ summing to 1; we will later discuss how we choose specific values for α_j. For each $j = 1, \ldots, q$, let θ_j be the tilting parameter under the jth distribution in the mixture; we will give the value of θ_j later. For each $j = 1, \ldots, q$, define probability measure P_j such that $A_1, \ldots, A_d$ are mutually independent, and each A_i has density $f_i^{\theta_j}$ when $i \in B_j$ and density f_i when $i \notin B_j$. Now the IS measure P_* is defined as $P_*(A) = \sum_{j=1}^q \alpha_j P_j(A)$ for any event A. The likelihood ratio is then

$$L = \left[\sum_{j=1}^{q} \alpha_j \exp\left(\theta_j T_j - \zeta_j(\theta_j)\right) \right]^{-1}, \tag{33}$$

where $\zeta_j(\theta) = \sum_{i \in B_j} \chi_i(\theta)$ is the CGF of T_j.

We apply an approach described in Glynn [1996] to choose each tilting parameter θ_j. Large-deviations theory suggests that under certain conditions, $P(T_j > x) \approx \exp(-x\theta_x + \zeta_j(\theta_x))$ for $x \gg E[T_j]$, where θ_x is the root of the equation $\zeta_j'(\theta_x) = x$ and prime denotes derivative. Taking $P(T_j > x) = 1 - p$ gives

$$-\zeta_j'(\theta)\theta + \zeta_j(\theta) = \ln(1 - p), \tag{34}$$

and we let θ_j be its root. Thus, we obtain $\zeta_j'(\theta_j)$ as a (crude) approximation for the p-quantile of T_j (under the original measure) when $p \approx 1$.

We now describe how to choose the IS mixture weights α_j by modifying a heuristic for IS-only in Juneja et al. [2007]. Let E_j denote expectation under measure P_j, and let E_* be expectation under the mixture P_*. Because $\zeta_j'(\theta_j)$ is roughly equal to the p-quantile of T_j and $X = \max_j T_j$, we approximate ξ_p via $\bar{\xi}_p \equiv \max_j \zeta_j'(\theta_j)$, which leads to approximating the second moment $E_*[L^2 I(X > \xi_p)]$ (under IS-only) by $E_*[L^2 I(X > \bar{\xi}_p)]$. Now we develop an upper bound for this quantity. On the event $\{T_j > \bar{\xi}_p\}$, we have that $L \le K_j/\alpha_j$ for $\theta_j > 0$ by (33), where $K_j = \exp(-\theta_j \bar{\xi}_p + \zeta_j(\theta_j))$. Hence, since $\{X > \bar{\xi}_p\} = \cup_{j=1}^q \{T_j > \bar{\xi}_p\}$, we get $E_*[L^2 I(X > \bar{\xi}_p)] \le (\max_{j=1,\ldots,q} K_j/\alpha_j)^2$. Choosing $\alpha_j \ge 0$ to minimize this bound subject to $\sum_{j=1}^q \alpha_j = 1$ gives $\alpha_j^* = K_j / \sum_{l=1}^q K_l$.

We used $k = 5$ equiprobable strata defined by the intervals $S_i = (G_*^{-1}((i-1)/k), G_*^{-1}(i/k)]$ for $i = 1, \ldots, 5$, where G_* is the CDF of Y under the IS measure P_*, so each $\lambda_i = 1/k$. Also, we let $\gamma_i = \lambda_i$. In our experiments, we generate stratified samples using the "bin tossing" method of Glasserman et al. [2000b].

8.2. Small SAN Model

We now give the details of a small SAN model with which we experiment. (The online appendix contains results from a larger SAN.) Previously considered in Hsu and Nelson [1990], the SAN has $d = 5$ activities, which are i.i.d. exponentials with mean 1. There are $q = 3$ paths in the network, and $B_1 = \{1, 2\}$, $B_2 = \{1, 3, 5\}$ and $B_3 = \{4, 5\}$. The CDF of X is given by, for $x \ge 0$, $F(x) = 1 + (3 - 3x - x^2/2)e^{-x} + (-3 - 3x + x^2/2)e^{-2x} - e^{-3x}$, which has continuous and positive density $f(x)$ for all $x \ge 0$, and in particular in a neighborhood of ξ_p, as required by our theory. The CGF of A_i is $\chi_i(\theta) = -\ln(1 - \theta)$, which exists for $\theta < 1$, and f_i^θ is exponential with rate $1 - \theta$. The CGF of T_j is $\zeta_j(\theta) = \sum_{i \in B_j} -\ln(1 - \theta)$, so $\zeta_j'(\theta) = \sum_{i \in B_j} (1 - \theta)^{-1}$.

Since the second path in the SAN has the largest mean, $Y = T_2 = A_1 + A_3 + A_5$ is our stratification variable and is also used to define the CV C. We now compute the CDF G_* of Y under P_*. First let G_j be the CDF of Y under measure P_j, and define $\eta_j = 1 - \theta_j$. Under measure P_1, we have that A_1 is exponential with rate η_1 while A_3 and A_5 are both exponential with rate 1, with A_1, A_3, A_5 mutually independent. Thus, by conditioning on A_1, we can show that for $t > 0$,

$$G_1(t) = 1 - e^{-\eta_1 t}\left(1 + \frac{\eta_1}{1 - \eta_1} + \frac{\eta_1}{(1 - \eta_1)^2}\right) + e^{-t}\frac{\eta_1}{1 - \eta_1}\left(1 + t + \frac{1}{1 - \eta_1}\right).$$

Also, G_3 is the same as G_1 but with η_3 replacing η_1 throughout. Finally, under measure P_2, we have that A_1, A_3, A_5 are i.i.d. exponential with rate η_2, so Y has an Erlang-3 distribution with scale parameter η_2 and $G_2(t) = 1 - e^{-\eta_2 t} - \eta_2 t e^{-\eta_2 t} - (\eta_2 t)^2 e^{-\eta_2 t}/2$. Thus, the CDF of Y under IS measure P_* is $G_*(t) = \sum_{j=1}^{3} \alpha_j G_j(t)$.

8.3. Discussion of Empirical Results

In our experiments, we constructed CIs for the p-quantile ξ_p using simulation methods CMC, AV, CV, and IS+SS. In each case, we constructed the CI from Theorem 3.3 having nominal level $1 - \alpha = 0.9$, and we estimated coverage by running 10^4 independent replications. We varied the computational budget $n = 100 \times 4^j$ for $0 \le j \le 3$. Whenever possible, we applied common random numbers across different n and different methods, which can sharpen comparisons.

We first discuss some issues related to the bandwidth h_n of the central finite difference (CFD) estimator (12) of ϕ_p. Let $q_{1,n} = p + h_n$ and $q_{2,n} = p - h_n$, and computing the CFD entails inverting the estimated CDF at $q_{1,n}$ and $q_{2,n}$. This then requires that $0 < q_{1,n} < 1$ and $0 < q_{2,n} < 1$, but when p is close to 1, we often have $q_{1,n} \ge 1$ for small values of n when h_n shrinks slowly. Consequently, in our experiments, if $p + h_n \ge 1$, we instead set $q_{1,n} = 1-(1-p)/10$ and $q_{2,n} = 2p-1+(1-p)/10$ when constructing the CFD, where $q_{2,n}$ is chosen so that $q_{1,n}$ and $q_{2,n}$ are symmetric about p. We also make similar adjustments to the forward and backward finite difference (FFD and BFD) estimators in (11).

There are some issues in experimenting with CFD (or FFD) to estimate ϕ_p when $p \approx 1$, n is small and the bandwidth h_n shrinks slowly. (Similar issues occur for BFD when $p \approx 0$.) If we run experiments over a range of n for which $p+h_n \ge 1$, the perturbed values $q_{1,n}$ and $q_{2,n}$ of p used in the CFD do not change for these n when we apply our adjustment described above. But the validity of the CFD requires that $q_{1,n} - q_{2,n} \to 0$ as n increases, so we may not see the asymptotics taking effect unless n is large enough for $q_{1,n} - q_{2,n}$ to shrink. Thus, for extreme p, we require very large n for our CFD estimates of ϕ_p to be accurate. (Of course, we could instead use a bandwidth that decreases more quickly, but our theory only covers the case when (16) holds, limiting how fast h_n can shrink.)

Table I presents the coverage levels (and average half widths in parentheses) for CIs when $p = 0.8$, where ϕ_p is estimated using the CFD with different bandwidths $h_n = 0.5n^{-v}$ for $v = 1/2$, $1/3$ (suggested by Hall and Sheather [1988] for CMC) and $1/5$ (suggested by Bloch and Gastwirth [1968] and Bofinger [1975] for CMC). As a benchmark to show the effect of estimating ϕ_p, we also give results for CIs constructed using the exact value of ϕ_p; see the columns labeled "Exact ϕ_p". For each simulation method, all CIs are centered at the same point $\tilde{\xi}_{p,n}$ and use the same estimate for ψ_p, which is the other term in the quantile estimator's asymptotic variance in (10) and does not depend on v. For $n = 6400$, the coverages in Table I are close to the nominal level for all CIs, which demonstrates their asymptotic validity. Also, within each simulation method, all of the average run lengths are about the same for large n. But for smaller values of n, $v = 1/5$ yields CFD estimates that are on average too large, as seen by comparing the average half widths for $v = 1/5$ to those for exact ϕ_p; this results in overcoverage. For CMC and AV, $v = 1/2$ and $v = 1/3$ give similar results, but for CV and IS+SS, $v = 1/3$ seems to lead to coverage levels that are closer to nominal for small n. This is despite the fact that $v = 1/2$ leads to average half widths that are closer to those for exact ϕ_p than $v = 1/3$. Thus, $v = 1/3$ overestimates ϕ_p, which may compensate for underestimated ψ_p, resulting in better coverage. Choosing $v = 1/3$ for IS+SS and CV when $p = 0.8$ complements the analysis of Hall and Sheather [1988], which shows that $v = 1/3$ asymptotically minimizes coverage error for CMC.

Table I. Coverages (and average half widths) for CFD with bandwidths $h_n = 0.5n^{-v}$ for $p = 0.8$

	CMC				AV			
	bandwidth exponent v			Exact	bandwidth exponent v			Exact
n	1/2	1/3	1/5	ϕ_p	1/2	1/3	1/5	ϕ_p
100	0.903 (0.564)	0.899 (0.533)	0.990 (0.960)	0.898 (0.500)	0.904 (0.350)	0.920 (0.358)	0.998 (0.719)	0.900 (0.326)
400	0.880 (0.250)	0.910 (0.262)	0.960 (0.315)	0.902 (0.250)	0.891 (0.164)	0.912 (0.170)	0.959 (0.204)	0.904 (0.163)
1600	0.880 (0.122)	0.903 (0.127)	0.933 (0.139)	0.900 (0.125)	0.890 (0.081)	0.903 (0.083)	0.933 (0.091)	0.899 (0.082)
6400	0.893 (0.062)	0.902 (0.063)	0.918 (0.066)	0.900 (0.063)	0.897 (0.041)	0.900 (0.041)	0.917 (0.043)	0.899 (0.041)
	CV				IS+SS			
	bandwidth exponent v			Exact	bandwidth exponent v			Exact
n	1/2	1/3	1/5	ϕ_p	1/2	1/3	1/5	ϕ_p
100	0.865 (0.346)	0.903 (0.372)	0.978 (0.639)	0.881 (0.333)	0.864 (0.280)	0.900 (0.299)	1.000 (0.718)	0.864 (0.275)
400	0.885 (0.170)	0.907 (0.175)	0.956 (0.211)	0.899 (0.168)	0.881 (0.142)	0.905 (0.146)	0.953 (0.176)	0.889 (0.141)
1600	0.888 (0.084)	0.899 (0.085)	0.930 (0.094)	0.898 (0.084)	0.889 (0.071)	0.902 (0.072)	0.932 (0.079)	0.897 (0.071)
6400	0.896 (0.042)	0.901 (0.042)	0.920 (0.045)	0.901 (0.042)	0.890 (0.036)	0.900 (0.036)	0.916 (0.038)	0.900 (0.036)

Table II gives the same results as Table I but for $p = 0.95$. Now $v = 1/2$ seems to lead to coverages that are closer to nominal than $v = 1/3$ or $1/5$. We also observed this when p is even closer to 1 in other experiments. Thus, it appears that for moderate quantiles with $0 \ll p \ll 1$, selecting $v = 1/3$ leads to better coverage, but for extreme quantiles, choosing $v = 1/2$ is more appropriate. The apparent inconsistency with the CMC theory of Hall and Sheather [1988] seems to arise because $p \approx 1$ requires n to be very large to have $p + h_n < 1$, so their asymptotic analysis does not take effect for moderate n and large p. In our other experiments, we consider only $p \geq 0.95$, so we use only $v = 1/2$ and $h_n = 0.5n^{-1/2}$ in those cases. (In Table II many of the results are the same for small n, and this is because $p + h_n \geq 1$ for the different bandwidths, so they were all adjusted using the approach in the second paragraph of this section.)

We now more closely analyze the results for exact ϕ_p. When $p = 0.8$, Table I shows the coverages for exact ϕ_p are close to nominal for all values of n and all simulation methods, and these coverages are no worse (and sometimes much better) than for estimated ϕ_p. When $p = 0.95$ (Table II), CMC, AV and IS+SS again show the same behavior, but this is not the case for CV when n is small. This may be due to the estimator of ψ_p being too small on average, which leads to undercoverage when using the exact value of ϕ_p, but the CFD estimator with $v = 1/2$ overestimates ϕ_p (larger average half width) to compensate. Moreover, to compute $\tilde{F}_n$, CV requires estimating a second moment and a mixed moment to calculate $\hat{\beta}_n(x)$ in (28) (or H_i in (32)), whereas the other simulation methods require estimating only first moments. Higher-order moments typically require larger sample sizes than lower-order ones for accurate estimation, and this seems to be another reason why the CV quantile estimators do not perform as well as the others for small n.

Table III presents additional results for $p = 0.95$ and $h_n = 0.5n^{-1/2}$. It gives results for FFD and BFD, as well as the same CFD and exact ϕ_p results from Table II for easy comparison. In terms of coverage we see that in general, CFD appears to do about as well or slightly better than FFD, both of which do better than BFD. Also, CFD has smaller average half widths than FFD, but BFD is even smaller, which appears to explain its undercoverage. This complements the MSE analysis in Section 7.1 of

Table II. Coverages (and average half widths) for CFD with bandwidths $h_n = 0.5n^{-v}$ for $p = 0.95$

	CMC				AV			
	bandwidth exponent v			Exact	bandwidth exponent v			Exact
n	1/2	1/3	1/5	ϕ_p	1/2	1/3	1/5	ϕ_p
100	0.947 (1.443)	0.947 (1.443)	0.947 (1.443)	0.907 (0.951)	0.950 (0.910)	0.950 (0.910)	0.950 (0.910)	0.907 (0.659)
400	0.901 (0.506)	0.973 (0.700)	0.973 (0.700)	0.904 (0.476)	0.915 (0.355)	0.982 (0.502)	0.982 (0.502)	0.904 (0.330)
1600	0.895 (0.241)	0.980 (0.345)	0.987 (0.368)	0.901 (0.238)	0.896 (0.168)	0.978 (0.237)	0.987 (0.258)	0.897 (0.165)
6400	0.900 (0.119)	0.934 (0.132)	0.989 (0.187)	0.905 (0.119)	0.904 (0.083)	0.935 (0.091)	0.990 (0.130)	0.903 (0.082)
	CV				IS+SS			
	bandwidth exponent v			Exact	bandwidth exponent v			Exact
n	1/2	1/3	1/5	ϕ_p	1/2	1/3	1/5	ϕ_p
100	0.802 (0.869)	0.802 (0.869)	0.802 (0.869)	0.763 (0.598)	0.982 (0.531)	0.982 (0.531)	0.982 (0.531)	0.872 (0.336)
400	0.892 (0.335)	0.950 (0.481)	0.950 (0.481)	0.868 (0.299)	0.923 (0.189)	0.989 (0.275)	0.989 (0.275)	0.897 (0.174)
1600	0.891 (0.155)	0.975 (0.220)	0.985 (0.240)	0.891 (0.152)	0.904 (0.090)	0.982 (0.127)	0.990 (0.139)	0.900 (0.088)
6400	0.897 (0.076)	0.931 (0.084)	0.991 (0.120)	0.901 (0.076)	0.897 (0.044)	0.931 (0.049)	0.991 (0.070)	0.898 (0.044)

Glasserman [2004] showing that CFD estimators of the derivative of a mean have asymptotically smaller MSE than FFD estimators.

The columns labeled “Comb. CFD” give results using the estimator (13) combining $r = 2$ CFD estimators, applying the strategy described at the end of Section 7 with $h_{n,1} = h_n$. For the middle range of n (400 and 1600), combining appears to reduce the average half width but lowers coverage, compared to CFD.

For comparison, we also used batching (with $b_0 = 10$ batches) as an alternative approach to construct CIs; see the columns labeled “Batch” in Table III. (In batching we allocate a computation budget of n/b_0 to each batch. We form an estimate of the quantile from each batch, and then take the sample mean and sample variance of the resulting b_0 quantile estimates to construct a CI for ξ_p using the $1 - \alpha/2$ critical point of a t-distribution with $b_0 - 1$ degrees of freedom.) In terms of coverage for small n, batching perhaps slightly outperforms the finite-difference estimators for IS+SS, but the opposite is true for AV, CV and CMC. To understand why, note that IS shifts a significant portion of the distribution's mass to be around ξ_p, thus enabling reasonable estimation of the quantile, even with a small sample size. However, for the other methods, very few samples are near ξ_p for small n because p is close to 1, and batching exacerbates this problem since quantile estimates from each batch are based on an even smaller sample size of n/b_0. Thus, batching for AV, CV and CMC results in poor coverage for small n. This can also be explained in terms of the bias of quantile estimators. Empirical results in Avramidis and Wilson [1998] show that bias can be an issue when estimating extreme quantiles with small sample sizes, so single-sample estimators often perform better than batched ones. In all cases for large n, the CIs for batching have larger average half widths than for CFD and exact ϕ_p because batching CIs use a t critical value rather than a normal's.

We also constructed CIs for the mean $E[X]$ as another benchmark for comparison; see the columns labeled “Est. Mean” in Table III. In general, we see that for small n and the different VRTs, coverage levels for the CIs for the mean are closer to the nominal level than those for the quantile. However, for CMC, using the exact ϕ_p in the CI for ξ_p leads to very good coverage, even for small n. This is because for CMC, the

Table III. Coverage levels (average half widths) for $h_n = 0.5n^{-1/2}$ and $p = 0.95$

n	CFD	FFD	BFD	Comb. CFD	Exact ϕ_p	Batch	Est. Mean
	CMC						
100	0.947 (1.443)	0.978 (2.295)	0.685 (0.633)	0.947 (1.443)	0.907 (0.951)	0.858 (0.910)	0.896 (0.278)
400	0.901 (0.506)	0.935 (0.629)	0.792 (0.383)	0.839 (0.442)	0.904 (0.476)	0.670 (0.457)	0.896 (0.140)
1600	0.895 (0.241)	0.918 (0.269)	0.845 (0.213)	0.883 (0.235)	0.901 (0.238)	0.835 (0.250)	0.903 (0.070)
6400	0.900 (0.119)	0.911 (0.126)	0.872 (0.112)	0.896 (0.119)	0.905 (0.119)	0.881 (0.127)	0.900 (0.035)
	AV						
n	CFD	FFD	BFD	Comb. CFD	Exact ϕ_p	Batch	Est. Mean
100	0.950 (0.910)	0.982 (1.354)	0.730 (0.453)	0.950 (0.910)	0.907 (0.659)	0.509 (0.569)	0.890 (0.150)
400	0.915 (0.355)	0.953 (0.442)	0.809 (0.269)	0.857 (0.306)	0.904 (0.330)	0.779 (0.336)	0.900 (0.075)
1600	0.896 (0.168)	0.924 (0.187)	0.842 (0.148)	0.884 (0.164)	0.897 (0.165)	0.859 (0.175)	0.898 (0.038)
6400	0.904 (0.083)	0.918 (0.088)	0.876 (0.078)	0.900 (0.082)	0.903 (0.082)	0.894 (0.089)	0.896 (0.019)
	CV						
n	CFD	FFD	BFD	Comb. CFD	Exact ϕ_p	Batch	Est. Mean
100	0.802 (0.869)	0.786 (1.221)	0.741 (0.511)	0.802 (0.869)	0.763 (0.598)	0.739 (0.841)	0.879 (0.133)
400	0.892 (0.335)	0.910 (0.410)	0.810 (0.261)	0.827 (0.287)	0.868 (0.299)	0.668 (0.410)	0.896 (0.068)
1600	0.891 (0.155)	0.912 (0.173)	0.845 (0.137)	0.880 (0.151)	0.891 (0.152)	0.883 (0.175)	0.899 (0.034)
6400	0.897 (0.076)	0.909 (0.081)	0.870 (0.072)	0.894 (0.076)	0.901 (0.076)	0.899 (0.083)	0.901 (0.017)
	IS+SS						
n	CFD	FFD	BFD	Comb. CFD	Exact ϕ_p	Batch	Est. Mean
100	0.982 (0.531)	0.997 (0.817)	0.726 (0.239)	0.982 (0.531)	0.872 (0.336)	0.879 (0.428)	0.881 (0.392)
400	0.923 (0.189)	0.963 (0.236)	0.807 (0.142)	0.862 (0.160)	0.897 (0.174)	0.897 (0.191)	0.891 (0.201)
1600	0.904 (0.090)	0.934 (0.100)	0.856 (0.079)	0.894 (0.087)	0.900 (0.088)	0.896 (0.095)	0.894 (0.101)
6400	0.897 (0.044)	0.913 (0.047)	0.871 (0.042)	0.894 (0.044)	0.898 (0.044)	0.895 (0.048)	0.895 (0.051)

variance constant in the CLT for the quantile estimator is $\kappa_p^2 = p(1-p)\phi_p^2$, so using the exact value of ϕ_p means there is nothing left to estimate in κ_p^2.

Table IV presents results from experiments with IS+SS to estimate extreme quantiles with $p = 1 - 10^{-e}$ for $e = 2, 3, 4, 5$. For exact ϕ_p, there is a small increase in the average half widths as p increases. Also, coverages for exact ϕ_p are all fairly close to nominal for all n, with a slight decrease as p increases for $n = 100$. Thus, increasing p seems to only slightly degrade the normal approximation from the CLT (10) for IS+SS. However, CFD gives significant overcoverage for all n. This is because the CFD consistently overestimates ϕ_p, as seen by comparing the average half widths for CFD and exact ϕ_p. The overestimation of ϕ_p by CFD seems to arise because $p + h_n \geq 1$ for all entries in Table IV except when $p = 1 - 10^{-2}$ and $n = 6400$, which leads to the issues described in the third paragraph of this section. To study this further, we ran another

Table IV. Coverage (and average half widths) of confidence intervals for IS+SS as p approaches 1

	$p = 1 - 10^{-2}$			$p = 1 - 10^{-3}$			$p = 1 - 10^{-4}$			$p = 1 - 10^{-5}$		
n	CFD	BFD	Exact ϕ_p	CFD	BFD	Exact ϕ_p	CFD	BFD	Exact ϕ_p	CFD	BFD	Exact ϕ_p
100	0.983	0.436	0.874	0.979	0.099	0.858	0.971	0.017	0.854	0.963	0.002	0.837
	(0.617)	(0.142)	(0.386)	(0.702)	(0.036)	(0.436)	(0.767)	(0.006)	(0.475)	(0.814)	(0.001)	(0.506)
400	0.989	0.598	0.896	0.989	0.176	0.894	0.989	0.029	0.888	0.989	0.004	0.888
	(0.323)	(0.103)	(0.201)	(0.371)	(0.031)	(0.230)	(0.412)	(0.006)	(0.254)	(0.447)	(0.001)	(0.275)
1600	0.993	0.720	0.903	0.991	0.276	0.900	0.992	0.053	0.897	0.992	0.008	0.896
	(0.163)	(0.067)	(0.102)	(0.189)	(0.025)	(0.117)	(0.209)	(0.005)	(0.129)	(0.228)	(0.001)	(0.140)
6400	0.943	0.803	0.900	0.993	0.407	0.899	0.991	0.092	0.896	0.992	0.014	0.902
	(0.059)	(0.040)	(0.051)	(0.095)	(0.019)	(0.058)	(0.105)	(0.004)	(0.065)	(0.114)	(0.001)	(0.070)

experiment for $p = 1 - 10^{-2}$ with $n = 25600$, and CFD (resp., exact ϕ_p) then resulted in a coverage of 0.909 (resp., 0.898) and an average half width of 0.026 (resp., 0.025). Thus, we can obtain valid CIs with CFD for extreme quantiles, but this requires very large sample sizes.

BFD evaluates the inverse of the estimated CDF at p and $p - h_n$, so unlike CFD (and FFD), BFD does not require making an adjustment to avoid evaluating the inverse outside of its domain $(0, 1)$ when $p \approx 1$. Even so, in Table IV, BFD performs poorly with extremely low coverage levels as p approaches 1, and this is due to BFD significantly underestimating ϕ_p.

9. CONCLUDING REMARKS

We have developed a general framework for producing an asymptotically valid confidence interval for a quantile ξ_p estimated using a VRT. Our framework, which requires the CDF estimator $\tilde{F}_n$ obtained by applying a VRT to satisfy Assumptions A1–A3, encompasses IS+SS, AV and CV, and we presented explicit formulae to construct confidence intervals for quantiles obtained using each of these methods.

We derived a consistent estimator of $\phi_p = 1/f(\xi_p)$ by first establishing that the quantile estimator satisfies a Bahadur-Ghosh representation, and then exploited this to estimate ϕ_p. The quantity ϕ_p can be expressed as $\frac{d}{dp}F^{-1}(p)$, and we estimated it via a finite difference of the inverse of the estimated CDF at points that are h_n or $2h_n$ apart, where $h_n \neq 0$ is a user-specified bandwidth and n is the computational budget. We studied empirically how the choice of h_n affects coverage and average half width. When $0 \ll p \ll 1$, choosing h_n of order $n^{-1/3}$ seems to be a good choice. But when $p \approx 1$, we suggest taking h_n of order $n^{-1/2}$. How to select h_n when applying VRTs is an ongoing topic for further research.

Kernel density estimators (e.g., Wand and Jones [1995]) might also be used to estimate ϕ_p, which is also another future research topic. Finally, note that Assumptions A1–A3 do not require that $\tilde{F}_n$ is constructed from independent samples, so one can also apply the framework in Section 3 to study quantile estimators for infinite-horizon problems. This is a problem for future work.

APPENDIX

Here we provide the proofs of the theorems from Section 3. To prove Theorem 3.1, we will apply the following lemma established by Ghosh [1971]; also see pp. 286–287 of David and Nagaraja [2003] for more details. We then transform $\tilde{R}_n$ in (7) in terms of the two sets of variates described in the lemma to complete the proof as required.

LEMMA A.1. *Let* (V_n, W_n), $n = 1, 2, \ldots$, *be a sequence of pairs of random variables such that*

(*1*) *for all* $\delta > 0$*, there exists* $\gamma \equiv \gamma(\delta)$ *and* $n_0 \equiv n_0(\gamma, \delta)$ *such that* $P\{|W_n| > \gamma\} < \delta$ *for* $n \geq n_0$ *(i.e,* $W_n = O_p(1)$*);*
(*2*) *for every* y *and every* $\epsilon > 0$*,*
(*a*) $\lim_{n\to\infty} P\{V_n \leq y, W_n \geq y + \epsilon\} = 0$*,*
(*b*) $\lim_{n\to\infty} P\{V_n \geq y, W_n \leq y + \epsilon\} = 0$*.*

Then

$$V_n - W_n \xrightarrow{P} 0 \;\; as \;\; n \to \infty. \tag{35}$$

PROOF OF THEOREM 3.1. We first prove (i). Set

$$\begin{aligned} V_n &= \sqrt{n}(\tilde{\xi}_{p_n,n} - \dot{\xi}_{p_n}), \\ W_n &= \frac{\sqrt{n}}{f(\xi_p)}[p - \tilde{F}_n(\xi_p)], \end{aligned} \tag{36}$$

where we recall that $\dot{\xi}_{p_n} = \xi_p + (p_n - p)/f(\xi_p)$, so (7) implies

$$\sqrt{n}\tilde{R}_n = V_n - W_n. \tag{37}$$

We now show that (V_n, W_n) satisfies conditions 1 and 2 of Lemma A.1 to establish our theorem. First, Assumption A3 implies condition 1 of Lemma A.1 by Theorem 2.3.2 of Lehmann [1999]. We next prove condition 2(a) of Lemma A.1 holds. Recall M_n is the event that $\tilde{F}_n(x)$ is monotonically increasing in x, and let M_n^c be its complement. Then

$$\begin{aligned} \{V_n \leq y\} &= \{\tilde{\xi}_{p_n,n} \leq \dot{\xi}_{p_n} + yn^{-1/2}, M_n\} \cup \{V_n \leq y, M_n^c\} \\ &\subseteq \{\tilde{F}_n(\dot{\xi}_{p_n} + yn^{-1/2}) \geq p_n\} \cup M_n^c = \{Z_n \leq y_n\} \cup M_n^c, \end{aligned}$$

where

$$\begin{aligned} Z_n &= \frac{\sqrt{n}}{f(\xi_p)} \left[F(\dot{\xi}_{p_n} + yn^{-1/2}) - \tilde{F}_n(\dot{\xi}_{p_n} + yn^{-1/2})\right], \\ y_n &= \frac{\sqrt{n}}{f(\xi_p)} \left[F(\dot{\xi}_{p_n} + yn^{-1/2}) - p_n\right]. \end{aligned}$$

Fix $\epsilon > 0$, and since

$$P\{V_n \leq y, W_n \geq y + \epsilon\} \leq P\{Z_n \leq y_n, W_n \geq y + \epsilon\} + P(M_n^c), \tag{38}$$

establishing condition 2(a) of Lemma A.1 reduces to proving the first term on the RHS of (38) approaches 0 as $n \to \infty$ because of Assumption A1. Since F is assumed to be differentiable at ξ_p, Young's form of Taylor's theorem (Hardy [1952], p. 278) implies

$$\begin{aligned} y_n &= \frac{\sqrt{n}}{f(\xi_p)} \left\{F(\xi_p) + \left(\frac{p_n - p}{f(\xi_p)} + yn^{-1/2}\right)[f(\xi_p) + o(1)] - p_n\right\} \\ &= y + \frac{\sqrt{n}}{f(\xi_p)} \left\{p + p_n - p + \left(\frac{p_n - p}{f(\xi_p)} + yn^{-1/2}\right)o(1) - p_n\right\} \end{aligned}$$

since $F(\xi_p) = p$. It is thus clear that $y_n \to y$ as $n \to \infty$ since $p_n - p = O(n^{-1/2})$. Therefore, there exists n_0 such that $|y - y_n| < \epsilon/2$ for all $n \geq n_0$, so taking the difference of the two inequalities in the first term of the RHS of (38) gives

$$P\{Z_n \leq y_n, W_n \geq y + \epsilon\} \leq P\{|W_n - Z_n| \geq \epsilon + y - y_n\} \leq P\left\{|W_n - Z_n| \geq \frac{\epsilon}{2}\right\}$$

for $n \geq n_0$. Hence, to show the left-hand side of (38) vanishes as $n \to \infty$, it suffices to prove $Z_n - W_n \xrightarrow{P} 0$ as $n \to \infty$. Note that

$$Z_n - W_n = \frac{\sqrt{n}}{f(\xi_p)} \left[(F(\dot{\xi}_{p_n} + y/\sqrt{n}) - F(\xi_p)) - (\tilde{F}_n(\dot{\xi}_{p_n} + y/\sqrt{n}) - \tilde{F}_n(\xi_p)) \right],$$

and we have $p_n - p = O(n^{-1/2})$ by assumption, so $\dot{\xi}_{p_n} + yn^{-1/2} = \xi_p + O(n^{-1/2})$. Consequently, $Z_n - W_n \xrightarrow{P} 0$ holds by Assumption A2 since $f(\xi_p) > 0$. Thus, the (V_n, W_n) pair satisfies condition 2(a) in Lemma A.1; condition 2(b) of the lemma is similarly established, so (35) holds. Recalling (37) completes the proof of (i).

The proof of (ii) is similar. Here, we also use the fact that $\xi_{p_n} \to \xi_p$ as $n \to \infty$, which follows since the continuity of f in a neighborhood of ξ_p implies $f(x) > 0$ for all x in a neighborhood of ξ_p so F is strictly increasing in a neighborhood of ξ_p, and then apply Theorems 6.8, 6.20 and 4.17 of Rudin [1976]. □

PROOF OF THEOREM 3.2. By (7), we have

$$\frac{\sqrt{n}}{\kappa_p}(\tilde{\xi}_{p,n} - \xi_p) = \frac{\sqrt{n}}{\psi_p}(p - \tilde{F}_n(\xi_p)) + \frac{\sqrt{n}}{\kappa_p}\tilde{R}_n. \tag{39}$$

The first term on the RHS of (39) converges in distribution to a standard normal by Assumption A3, and the second term converges in probability to 0 by (8). Thus, the result follows from Slutsky's theorem. □

PROOF OF THEOREM 3.3. Since $\tilde{\xi}_{q,n} = \tilde{F}_n^{-1}(q)$ for all q, Theorem 3.1 and (7) imply

$$\tilde{F}_n^{-1}(p) = \xi_p - \frac{\tilde{F}_n(\xi_p) - p}{f(\xi_p)} + \tilde{R}_{n,1},$$

$$\tilde{F}_n^{-1}(p + h_n) = \dot{\xi}_{p+h_n} - \frac{\tilde{F}_n(\xi_p) - p}{f(\xi_p)} + \tilde{R}_{n,2},$$

where $\sqrt{n}\tilde{R}_{n,i} \xrightarrow{P} 0$ as $n \to \infty$ for $i = 1, 2$. Thus,

$$\tilde{\phi}_{p,n,1}(h_n) = \frac{1}{h_n}(\dot{\xi}_{p+h_n} - \xi_p + \tilde{R}_{n,2} - \tilde{R}_{n,1}).$$

For part (i), we have $h_n = c/\sqrt{n}$ by (14), so $\dot{\xi}_{p+h_n} = \xi_p + h_n/f(\xi_p)$ by Theorem 3.1(i). Hence,

$$\tilde{\phi}_{p,n,1}(h_n) = \frac{\sqrt{n}}{c}\left(\frac{c}{\sqrt{n}f(\xi_p)} + \tilde{R}_{n,2} - \tilde{R}_{n,1}\right) \xrightarrow{P} \frac{1}{f(\xi_p)} = \phi_p$$

as $n \to \infty$. For part (ii), the additional conditions imposed allow us to set $\dot{\xi}_{p+h_n} = \xi_{p+h_n}$ by Theorem 3.1(ii), so $\frac{d}{dp}F^{-1}(p) = \phi_p$ implies

$$\tilde{\phi}_{p,n,1}(h_n) = \frac{\xi_{p+h_n} - \xi_p}{h_n} + \frac{1}{h_n}\left(\tilde{R}_{n,2} - \tilde{R}_{n,1}\right) \xrightarrow{P} \phi_p$$

as $n \to \infty$ by (16). This completes the proof of the consistency of (11), and the consistency of (12) similarly holds. Also, since $\sum_{j=1}^{r} w_j = 1$, applying Slutsky's theorem yields the consistency of (13). Finally, (15) follows rom (10) and Slutsky's theorem. □

ELECTRONIC APPENDIX

The electronic appendix for this article can be accessed in the ACM Digital Library.

ACKNOWLEDGMENTS

The authors would like to thank the Editor in Chief, an associate editor and referees for providing comments that improved the paper.

REFERENCES

AVRAMIDIS, A. N. AND WILSON, J. R. 1998. Correlation-induction techniques for estimating quantiles in simulation. *Operations Research 46*, 574–591.

BABU, G. J. 1986. Efficient estimation of the reciprocal of the density quantile function at a point. *Statistics and Probability Letters 4*, 133–139.

BAHADUR, R. R. 1966. A note on quantiles in large samples. *Annals of Mathematical Statistics 37*, 577–580.

BILLINGSLEY, P. 1995. *Probability and Measure*, Third ed. John Wiley & Sons, New York.

BILLINGSLEY, P. 1999. *Convergence of Probability Measures*, Second ed. John Wiley & Sons, New York.

BLOCH, D. A. AND GASTWIRTH, J. L. 1968. On a simple estimate of the reciprocal of the density function. *Annals of Mathematical Statistics 39*, 1083–1085.

BOFINGER, E. 1975. Estimation of a density function using order statistics. *Australian Journal of Statistics 17*, 1–7.

CHU, F. 2010. Variance reduction techniques for estimating quantiles and value-at-risk. Ph.D. thesis, New Jersey Institute of Technology, Information Systems Department.

CHU, F. AND NAKAYAMA, M. K. 2010. Confidence intervals for quantiles and value-at-risk when applying importance sampling. In *Proceedings of the 2010 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 2751–2761.

DAVID, H. A. AND NAGARAJA, H. 2003. *Order Statistics*, Third ed. Wiley, Hoboken, NJ.

DUFFIE, D. AND PAN, J. 1997. An overview of value at risk. *Journal of Derivatives 4*, 7–49.

FU, M. C., HONG, L. J., AND HU, J.-Q. 2009. Conditional Monte Carlo estimation of quantile sensitivities. *Management Science 55*, 2019–2027.

GHOSH, J. K. 1971. A new proof of the Bahadur representation of quantiles and an application. *Annals of Mathematical Statistics 42*, 1957–1961.

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

GLASSERMAN, P., HEIDELBERGER, P., AND SHAHABUDDIN, P. 2000a. Importance sampling and stratification for value-at-risk. In *Computational Finance 1999, Proceedings of the Sixth International Conference on Computational Finance*, Y. S. Abu-Mostafa, B. L. Baron, A. W. Lo, and A. S. Weigend, Eds. 7–24.

GLASSERMAN, P., HEIDELBERGER, P., AND SHAHABUDDIN, P. 2000b. Variance reduction techniques for estimating value-at-risk. *Management Science 46*, 1349–1364.

GLYNN, P. W. 1996. 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, 180–185.

GLYNN, P. W. AND IGLEHART, D. L. 1989. Importance sampling for stochastic systems. *Management Science 35*, 1367–1393.

GOLDENBERG, D. K., QIU, L., XIE, H., YANG, Y. R., AND ZHANG, Y. 2004. Optimizing cost and performance for multihoming. In *SIGCOMM '04 Proceedings*. 79–92.

HALL, P. AND SHEATHER, S. J. 1988. On the distribution of a Studentized quantile. *Journal of the Royal Statistical Society B 50*, 381–391.

HARDY, G. H. 1952. *A Course in Pure Mathematics*, 10 ed. Cambridge University Press, New York.

HESTERBERG, T. C. AND NELSON, B. L. 1998. Control variates for probability and quantile estimation. *Management Science 44*, 1295–1312.

HONG, L. J. 2009. Estimating quantile sensitivities. *Operations Research 57*, 118–130.

HSU, J. C. AND NELSON, B. L. 1990. Control variates for quantile estimation. *Management Science 36*, 835–851.

JIN, X., FU, M. C., AND XIONG, X. 2003. Probabilistic error bounds for simulation quantile estimation. *Management Science 49*, 230–246.

JUNEJA, S., KARANDIKAR, R., AND SHAHABUDDIN, P. 2007. Asymptotics and fast simulation for tail probabilities of maximum of sums of few random variables. *ACM Transactions on Modeling and Computer Simulation 17*, article 2, 35 pages.

KIEFER, J. 1967. On Bahadur's representation of sample quantiles. *Annals of Mathemetical Statistics 38*, 1350–1353.

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

LIU, G. AND HONG, L. J. 2009. Kernel estimation of quantile sensitivities. *Naval Research Logistics 56*, 511–525.

PARZEN, E. 1979. Density quantile estimation approach to statistical data modelling. In *Smoothing Techniques for Curve Estimation*. Springer, Berlin.

PETROV, V. V. 1995. *Limit Theorems in Probability Theory*. Oxford University Press, England.

ROSS, S. 1997. *Simulation*, Second ed. Academic Press, San Diego, CA.

RUDIN, W. 1976. *Principles of Mathematical Analysis*, Third ed. McGraw-Hill, New York.

SERFLING, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

SUN, L. AND HONG, L. J. 2010. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters*, to appear.

TUKEY, J. W. 1965. Which part of the sample contains the information? *Proc Natl Acad Sci USA 53*, 127–134.

WAND, M. P. AND JONES, M. C. 1995. *Kernel Smoothing*. Chapman & Hall, London.

Online Appendix to: Confidence Intervals for Quantiles When Applying Variance-Reduction Techniques

FANG CHU, New Jersey Institute of Technology
MARVIN K. NAKAYAMA, New Jersey Institute of Technology

A. EMPIRICAL RESULTS FOR A LARGER SAN MODEL

A.1. Details of Model and VRTs

We experimented with another SAN, which is larger than that in Section 8.2 and which was previously considered in Juneja et al. [2007]. (We use the same notation as in Section 8.1.) This SAN has $d = 15$ activities, and the distribution of the duration A_i of activity i for $i \leq 8$ (resp., $i \geq 9$) is exponential with rate 0.5 (resp., 1). Let f_i denote the density function of A_i, so $f_i(t) = \lambda_i e^{-\lambda_i t}$ for $t \geq 0$, where $\lambda_i = 0.5$ for $i \leq 8$ and $\lambda_i = 1$ for $i \geq 9$. There are $q = 10$ paths through the network, with the B_j defined as follows:

j	B_j	j	B_j
1	$\{1, 4, 11, 15\}$	2	$\{1, 4, 12\}$
3	$\{2, 5, 11, 15\}$	4	$\{2, 5, 12\}$
5	$\{2, 6, 13\}$	6	$\{2, 7, 14\}$
7	$\{3, 8, 11, 15\}$	8	$\{3, 8, 12\}$
9	$\{3, 9, 15\}$	10	$\{3, 10, 14\}$

In general the distribution of $T_j = \sum_{i \in B_j} A_i$ is given by a convolution of exponentials. The CDF of $X = \max_j T_j$ is very cumbersome to work out, so we instead estimated the true values of the quantile ξ_p, sparsity function $\phi_p = 1/f(\xi_p)$ and mean $E[X]$ using a single simulation run with CMC and $n = 5 \times 10^7$, which is about the largest value of n our Matlab program could handle before running out of memory. We estimated ϕ_p using CFD with bandwidth $h_n = 0.5 n^{-1/2}$. For $p = 0.95$, the estimates are $\xi_p = 15.3478$ and $\phi_p = 48.5718$. For $p = 0.99$, the estimates are $\xi_p = 19.1259$ and $\phi_p = 225.2248$. Our estimate for $E[X] = 9.3435$. We used these values in our coverage experiments, described in Section A.2.

The cumulant generating function of A_i evaluated at parameter θ is $\chi_i(\theta) = \ln[\lambda_i/(\lambda_i - \theta)]$, which exists for $\theta < \lambda_i$. The exponentially tilted density of A_i is $f_i^\theta(t) = (\lambda_i - \theta) e^{-(\lambda_i - \theta)t}$ for $t \geq 0$, and 0 otherwise; i.e., f_i^θ is the density of an exponential with rate $\lambda_i - \theta$. The cumulant generating function of T_j is $\zeta_j(\theta) = \sum_{i \in B_j} \ln[\lambda_i/(\lambda_i - \theta)]$, so $\zeta_j'(\theta) = \sum_{i \in B_j} (\lambda_i - \theta)^{-1}$. Specializing (34) to the current setting gives the value of θ_j as the solution in θ to

$$-\theta \sum_{i \in B_j} \frac{1}{\lambda_i - \theta} + \sum_{i \in B_j} \ln\left[\frac{\lambda_i}{\lambda_i - \theta}\right] = \ln(1 - p).$$

Also, we have $\zeta_j'(\theta_j) = \sum_{i \in B_j} (\lambda_i - \theta_j)^{-1}$, $K_l = \exp(-\theta_l \bar{\xi}_p) \prod_{i \in B_l} \lambda_i/(\lambda_i - \theta_l)$ and $\bar{\xi}_p = \max_r \sum_{i \in B_r} (\lambda_i - \theta_r)^{-1}$.

We incorporate SS as follows. Recall μ_j is the expected length of the jth path. Paths 1, 3 and 7 all have $\mu_j = 6$, while the other paths have strictly smaller expected lengths. So we want to choose T_1, T_3 or T_7 to be the stratification variable Y. To come up with

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

a heuristic method to determine which one to select, let $c_j = \sum_{k=1}^{q} \sum_{i \in B_k} I(i \in B_j)$, which is the number of times an activity from path j appears in any of the paths. It is easy to compute $c_1 = 11$ and $c_3 = c_7 = 13$. It seems reasonable that we want tilting to affect as often as possible the activities on the path corresponding to the stratification variable, so this means we should choose the path j that maximizes c_j as the stratification variable. This narrows it down to path 3 or 7, and we arbitrary choose $Y = T_3 = A_2 + A_5 + A_{11} + A_{15}$, which is the sum of four independent exponentials, two with rate 0.5 and two with rate 1. We also used the same Y in defining the CV $C = I(Y \leq G^{-1}(p))$.

We now want to work out the CDF G_* of Y under IS measure P_*. First let G_j be the CDF of Y under measure P_j. Under any P_j, the A_i, $i \in B_3$, are still independent exponentials, but now A_i has rate $\lambda_i - \theta_j$ if $i \in B_j$, and rate λ_i if $i \notin B_j$. Examining all of the sets B_j and seeing what each has in common with B_3, we see that there are two possible general forms for the distribution of Y under P_j as the sum of four independent exponentials:

— two of the exponentials have one rate a and the other two have another rate b;
— two of the exponentials have one rate a, another has a different rate b, and the last has yet another rate c.

Each of the individual rates a, b and c is either some λ_i or some $\lambda_i - \theta_j$. To handle these separate situations in a uniform manner, we now write $Y = Z_1 + Z_2$, where

— Z_1 is the sum of two independent exponentials, each with common rate a (i.e., Z_1 is Erlang-2),
— Z_2 is the sum of two independent exponentials, either with the same rate b or different rates b and c,

with Z_1 and Z_2 independent. Let M_i (resp., m_i) be the CDF (resp., density) of Z_i. Then,

$$\begin{aligned} G_j(t) = P_j(Z_1 + Z_2 \leq t) &= \int_0^t M_1(t-z) m_2(z)\, dz \\ &= P_j(Z_2 \leq t) - \int_0^t e^{-a(t-z)} m_2(z)\, dz - \int_0^t a(t-z) e^{-a(t-z)} m_2(z)\, dz. \end{aligned} \tag{40}$$

In the case that Z_2 is the sum of i.i.d. exponentials with rate b, then m_2 is just the density of an Erlang-2. When Z_2 is the sum of two independent exponentials with different rates b and c, then

$$m_2(z) = \int_0^z b e^{-bx} c e^{-c(z-x)}\, dx = \frac{bc}{c-b}\left(e^{-bz} - e^{-cz}\right).$$

We then used Matlab to work out the details of G_j in (40) in both cases. Finally, $G_*(t) = \sum_{j=1}^{q} \alpha_j G_j(t)$.

A.2. Empirical Results

Table V gives coverages (and average half widths) for $p = 0.95$ when applying CFD with bandwidth $h_n = 0.5n^{-v}$ for $v = 1/2$, $1/3$ and $1/5$. As we previously saw in Table II for the smaller SAN, $v = 1/2$ gives the coverages closest to nominal. We also point out that the average half widths for CV in Table V are only slightly smaller than for CMC, which is in contrast to Table II where CV reduced the average half width by about 36% relative to CMC for $v = 1/2$. This is because for the larger SAN, the CV $C = I(Y \leq G^{-1}(p))$ is not very highly correlated with $I(X \leq \xi_p)$ as the larger SAN has three paths that are tied with the largest expected length, whereas the smaller SAN

Table V. Coverages (and average half widths) for larger SAN using CFD with bandwidths $h_n = 0.5n^{-v}$ for $p = 0.95$

	CMC				AV			
	bandwidth exponent v			Exact	bandwidth exponent v			Exact
n	1/2	1/3	1/5	ϕ_p	1/2	1/3	1/5	ϕ_p
100	0.949 (2.663)	0.949 (2.663)	0.949 (2.663)	0.907 (1.741)	0.951 (1.688)	0.951 (1.688)	0.951 (1.688)	0.902 (1.217)
400	0.900 (0.928)	0.974 (1.289)	0.974 (1.289)	0.902 (0.871)	0.914 (0.658)	0.981 (0.934)	0.981 (0.934)	0.901 (0.609)
1600	0.891 (0.443)	0.979 (0.636)	0.987 (0.679)	0.898 (0.435)	0.895 (0.311)	0.979 (0.442)	0.989 (0.480)	0.894 (0.305)
6400	0.897 (0.219)	0.931 (0.242)	0.990 (0.344)	0.900 (0.218)	0.896 (0.153)	0.930 (0.169)	0.988 (0.242)	0.898 (0.152)

	CV				IS+SS			
	bandwidth exponent v			Exact	bandwidth exponent v			Exact
n	1/2	1/3	1/5	ϕ_p	1/2	1/3	1/5	ϕ_p
100	0.928 (2.445)	0.928 (2.445)	0.928 (2.445)	0.873 (1.615)	0.984 (1.213)	0.984 (1.213)	0.984 (1.213)	0.881 (0.766)
400	0.907 (0.902)	0.977 (1.313)	0.977 (1.313)	0.896 (0.821)	0.920 (0.435)	0.988 (0.634)	0.988 (0.634)	0.894 (0.399)
1600	0.893 (0.422)	0.980 (0.600)	0.989 (0.655)	0.899 (0.412)	0.902 (0.206)	0.982 (0.294)	0.991 (0.321)	0.894 (0.202)
6400	0.897 (0.208)	0.932 (0.229)	0.992 (0.327)	0.900 (0.206)	0.904 (0.102)	0.934 (0.112)	0.992 (0.161)	0.903 (0.101)

had a unique path with largest expected length. Similarly, the average half widths for IS+SS are about the same as those for IS-only (not shown).

Table VI present more detailed results for $p = 0.95$ when the bandwidth $h_n = 0.5n^{-1/2}$. These results are qualitatively similar to those in Table III for the smaller SAN: CFD outperforms FFD and BFD for smaller n; for the middle range of n, combined CFD gives smaller average half widths but worse coverage; and batching (with $b_0 = 10$ batches) has worse coverage for small n than the methods that estimate the variance.

Table VII gives results for $p = 0.99$. As previously seen in Table IV for the smaller SAN, we observe that for the largest n, the coverages for CFD, FFD and BFD are still not quite at the nominal level for the larger SAN. But coverage for exact ϕ_p are close to nominal for all but the smallest value of n.

B. OTHER PROOFS

B.1. Sufficient Conditions for Assumption A2

As an alternative to directly showing Assumption A2 holds, we now provide a set of sufficient conditions for the assumption.

CONDITION C1. *$E[\tilde{F}_n(x)] = F(x)$ for all x, and for every $a_n = O(n^{-1/2})$,*

$$E\left[\tilde{F}_n(\xi_p + a_n) - \tilde{F}_n(\xi_p)\right]^2 = [F(\xi_p + a_n) - F(\xi_p)]^2 + s_n(a_n)/n \tag{41}$$

with $s_n(a_n) \to 0$ as $n \to \infty$.

Equation (41) considers the second moment of the difference in the CDF estimator at ξ_p and a perturbation from ξ_p of order $n^{-1/2}$. The condition requires that the second moment can be expressed as the square of the difference in probabilities due to this perturbation with a remainder approaching 0 sufficiently fast.

PROPOSITION B.1.

(i) Condition C1 implies Assumption A2.

(ii) If Condition C1 holds for all $a_n \to 0$, then Assumption A2 holds for all $a_n \to 0$.

PROOF. Let $a_n = O(n^{-1/2})$ or $a_n \to 0$. Define W_n as in (36), and let

$$Z_n = \frac{\sqrt{n}}{f(\xi_p)} \left[F(\xi_p + a_n) - \tilde{F}_n(\xi_p + a_n) \right].$$

To establish Assumption A2, it suffices to prove $E[(Z_n - W_n)^2] \to 0$ as $n \to \infty$; e.g., see Theorem 2.1.1 of Lehmann [1999]. Let $d_n = F(\xi_p + a_n) - F(\xi_p)$ and $D_n = \tilde{F}_n(\xi_p + a_n) - \tilde{F}_n(\xi_p)$. Then

$$E[(Z_n - W_n)^2] = \frac{n}{f^2(\xi_p)} \left[d_n^2 - 2d_n E(D_n) + E(D_n^2) \right]. \tag{42}$$

Condition C1 implies $E[D_n] = d_n$ and $E[D_n^2] = d_n^2 + s_n(a_n)/n$. Putting these into (42) yields

$$E[(Z_n - W_n)^2] = \frac{s_n(a_n)}{f^2(\xi_p)} \to 0 \tag{43}$$

as $n \to \infty$ by the property of s_n in Condition C1.

B.2. Proof of Theorem 4.1(i) and (iii)

We will show that the IS+SS CDF estimator $\tilde{F}_n$ in (20) satisfies Assumptions A1, A2 for all $a_n \to 0$, and A3, so the results will then follow from Theorem 3.1. Since L_{ij}, $I(X_{ij} \le x)$, n_i and λ_i are all nonnegative, $\tilde{F}_n(x)$ is monotonically increasing in x for each n, so Assumption A1 is satisfied.

Now we will show that Assumption A2 holds for all $a_n \to 0$ by verifying Condition C1 for all $a_n \to 0$ and applying Proposition B.1. For all $x \in \Re$, we have

$$\begin{aligned} E_*[\tilde{F}_n(x)] &= E_* \left[\sum_{i=1}^{k} \lambda_i \frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{ij} \le x) L_{ij} \right] = \sum_{i=1}^{k} \lambda_i \frac{1}{n_i} \sum_{j=1}^{n_i} E_* \left[I(X_{ij} \le x) L_{ij} \right] \\ &= \sum_{i=1}^{k} \lambda_i E_* \left[I(X \le x) L \mid Y \in S_i \right] = F(x) \end{aligned} \tag{44}$$

by (19), so $\tilde{F}_n(x)$ is unbiased for all x.

We now show (41) holds with $s_n(a_n) \to 0$ as $n \to \infty$ for all $a_n \to 0$. Define $\rho_n = \min(\xi_p, \xi_p + a_n)$ and $\rho_n' = \max(\xi_p, \xi_p + a_n)$. Then

$$b_n \equiv E_*[\tilde{F}_n(\xi_p + a_n) - \tilde{F}_n(\xi_p)]^2 = E_* \left(\sum_{i=1}^{k} \frac{\lambda_i}{n_i} \sum_{j=1}^{n_i} C_{ij} \right)^2, \tag{45}$$

where $C_{ij} = I(\rho_n < X_{ij} \le \rho'_n)L_{ij}$. The C_{ij}, $j = 1, \ldots, n_i$, are i.i.d., and C_{ij} and $C_{i'j'}$ are independent for $i \ne i'$, so expanding the square in (45) leads to

$$\begin{aligned} b_n &= \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i^2} \sum_{j=1}^{n_i} E_*[C_{ij}^2] + \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i^2} \sum_{j=1}^{n_i} \sum_{\substack{j'=1 \\ j' \ne j}}^{n_i} E_*[C_{ij}]E_*[C_{ij'}] \\ &+ \sum_{i=1}^{k} \sum_{\substack{i'=1 \\ i' \ne i}}^{k} \frac{\lambda_i \lambda_{i'}}{n_i n_{i'}} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} E_*[C_{ij}]E_*[C_{i'j'}] \\ &= \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i} E_*[C_{i1}^2] + \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i^2}(n_i(n_i-1))E_*^2[C_{i1}] + \sum_{i=1}^{k} \sum_{\substack{i'=1 \\ i' \ne i}}^{k} \lambda_i \lambda_{i'} E_*[C_{i1}]E_*[C_{i'1}] \\ &= \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i} E_*[C_{i1}^2] + \left(\sum_{i=1}^{k} \lambda_i E_*[C_{i1}] \right)^2 - \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i} E_*^2[C_{i1}] \;\equiv\; t_{1,n} + t_{2,n} - t_{3,n}. \qquad (46) \end{aligned}$$

Because $t_{2,n} = (F(\xi_p + a_n) - F(\xi_p))^2$ by (19), establishing (41) reduces to verifying that $s_n(a_n) \equiv n t_{1,n} - n t_{3,n} \to 0$ as $n \to \infty$. Since $k < \infty$, there exist $\epsilon > 0$ and $\delta > 0$ by assumption such that $\upsilon_i \equiv E_*[I(X_{ij} < \xi_p + \delta)L_{ij}^{2+\epsilon}] < \infty$ for all strata $i = 1, 2, \ldots, k$, and let $\tau = \max_{i=1,\ldots,k} \upsilon_i^{1/(1+\epsilon)} < \infty$. Let $\lambda'_i = P(Y \in S_i)$, and we may assume $\lambda'_i > 0$ since otherwise, $L_{i1} = 0$ so the ith summands in $t_{1,n}$ and $t_{3,n}$ vanish. Recall $n_i = \gamma_i n$. Then applying a change of measure and Hölder's inequality yield

$$\begin{aligned} t_{1,n} &= \frac{1}{n} \sum_{i=1}^{k} \frac{\lambda_i^2}{\gamma_i} E\left[I^2(\rho_n < X_{i1} \le \rho'_n) L_{i1} \right] \\ &\le \frac{1}{n} \sum_{i=1}^{k} \frac{\lambda_i^2}{\gamma_i} E^{\frac{\epsilon}{1+\epsilon}} \left[I^{\frac{1+\epsilon}{\epsilon}}(\rho_n < X_{i1} \le \rho'_n) \right] E^{\frac{1}{1+\epsilon}} \left[I^{1+\epsilon}(\rho_n < X_{i1} \le \rho'_n) L_{i1}^{1+\epsilon} \right] \\ &= \frac{1}{n} \sum_{i=1}^{k} \frac{\lambda_i^2}{\gamma_i} P^{\frac{\epsilon}{1+\epsilon}}(\rho_n < X \le \rho'_n \mid Y \in S_i) \, E_*^{\frac{1}{1+\epsilon}} \left[I(\rho_n < X_{i1} \le \rho'_n) L_{i1}^{2+\epsilon} \right] \\ &\le \frac{\tau}{n} \sum_{i=1}^{k} \frac{\lambda_i^2}{\gamma_i (\lambda'_i)^{\epsilon/(1+\epsilon)}} P^{\epsilon/(1+\epsilon)}(\rho_n < X \le \rho'_n, Y \in S_i) \end{aligned}$$

for n sufficiently large since then $I(\rho_n < X_{ij} \le \rho'_n) \le I(X_{ij} \le \xi_p + \delta)$ because $\rho'_n = \max(\xi_p, \xi_p + a_n)$ and $a_n \to 0$. Also, $t_{3,n} = \frac{1}{n} \sum_{i=1}^{k} \frac{\lambda_i^2}{\gamma_i \lambda_i'^2} P^2(\rho_n < X \le \rho'_n, Y \in S_i)$, which leads to

$$\begin{aligned} &|s_n(a_n)| \le |n t_{1,n}| + |n t_{3,n}| \\ &\le \sum_{i=1}^{k} \left[\frac{\tau \lambda_i^2}{\gamma_i} \left(\frac{P(\rho_n < X \le \rho'_n, Y \in S_i)}{\lambda'_i} \right)^{\epsilon/(1+\epsilon)} + \frac{\lambda_i^2}{\gamma_i} \left(\frac{P(\rho_n < X \le \rho'_n, Y \in S_i)}{\lambda'_i} \right)^2 \right]. \end{aligned}$$

Because the differentiability of F at ξ_p implies F is continuous at ξ_p, it follows that $P(\rho_n < X \le \rho'_n, Y \in S_i) \le P(\rho_n < X \le \rho'_n) \to 0$ since $\rho_n \to \xi_p$ and $\rho'_n \to \xi_p$ as $n \to \infty$. Thus, $s_n(a_n) \to 0$ as $n \to \infty$ because $\gamma_i > 0$, $\lambda_i \le 1$ and $\tau < \infty$, and then (41) follows from (46). Consequently, Assumption A2 holds for all $a_n \to 0$ by Proposition B.1.

Finally we need to show Assumption A3 holds. Since $n_i = n\gamma_i$,

$$\sqrt{n}\left[\tilde{F}_n(\xi_p) - F(\xi_p)\right] = \sqrt{n}\left\{\sum_{i=1}^{k}\frac{\lambda_i}{n_i}\sum_{j=1}^{n_i} I(X_{ij} \le \xi_p)L_{ij} - \sum_{i=1}^{k}\lambda_i E_*\left[I(X_{i1} \le \xi_p)L_{i1}\right]\right\}$$
$$= \sum_{i=1}^{k}\frac{\lambda_i}{\sqrt{\gamma_i}}G_{i,n}, \tag{47}$$

where $G_{i,n} = \sqrt{n_i}\left[\frac{1}{n_i}\sum_{j=1}^{n_i} I(X_{ij} \le \xi_p)L_{ij} - P(X_{i1} \le \xi_p)\right]$. Also, $I(X_{ij} \le \xi_p)L_{ij}$, $j = 1, 2, \ldots, n_i$, are i.i.d. with finite variance under P_* since $\upsilon_i < \infty$. Hence, $G_{i,n} \xrightarrow{L} N(0, \zeta_i^2) \equiv N_i$ as $n \to \infty$ for each i, where $\zeta_i^2 = \mathrm{Var}_*[I(X_{ij} \le \xi_p)L_{ij}]$, which equals (21) by a change of measure. The samples across strata are independent, so the independence of $G_{i,n}$, $i = 1, \ldots, k$, implies the independence of N_i, $i = 1, \ldots, k$. It follows that $(G_{i,n}, i = 1, \ldots, k) \xrightarrow{L} (N_i, i = 1, \ldots, k)$ as $n \to \infty$ by Example 3.2 of Billingsley [1999]. Then we apply the continuous mapping theorem (Theorem 5.1 of Billingsley [1999]) to (47) and obtain $\sqrt{n}\left(\tilde{F}_n(\xi_p) - F(\xi_p)\right) \xrightarrow{L} N\left(0, \psi_p^2\right)$ with

$$\psi_p^2 = \sum_{i=1}^{k}\lambda_i^2\zeta_i^2/\gamma_i, \tag{48}$$

which completes the proof.

B.3. Proof of Theorem 4.1(ii) and (iv)

By applying Theorem 3.2, we see that (10) follows from Theorem 4.1(i), and the formula for ψ_p^2 is given in (48). To establish (15), we will show

$$\tilde{\psi}_{p,n} \xrightarrow{P} \psi_p \text{ as } n \to \infty \tag{49}$$

as required by Theorem 3.3. Recall ζ_i^2 in (21), and we defined $\tilde{\zeta}_{i,n}^2$ in (22) as an estimator of ζ_i^2. Now define $Z_{i,n}(x) = \frac{1}{n_i}\sum_{j=1}^{n_i} I(X_{ij} \le x)L_{ij}^2$ and $z_i(x) = E_*[I(X_{ij} \le x)L_{ij}^2]$, so $Z_{i,n}(\tilde{\xi}_{p,n})$ is the first term in $\tilde{\zeta}_{i,n}^2$ and $z_i(\xi_p)$ is the first term in ζ_i^2. To prove (49), we first show that for each $i = 1, \ldots, k$,

$$Z_{i,n}(\tilde{\xi}_{p,n}) \xrightarrow{P} z_i(\xi_p) \text{ as } n \to \infty \tag{50}$$

under the IS measure P_*. Fix an arbitrary $a_0 > 0$, and establishing (50) requires proving that $P_*\left\{|Z_{i,n}(\tilde{\xi}_{p,n}) - z_i(\xi_p)| > a_0\right\} \to 0$ as $n \to \infty$. We assumed that $\upsilon_i \equiv E_*[I(X_{ij} \le \xi_p + \delta)L_{ij}^{2+\epsilon}] = E[I(X_{ij} \le \xi_p + \delta)L_{ij}^{1+\epsilon}] < \infty$ for some positive ϵ and δ, where we can use the same ϵ and δ (the smallest ones) for all i since $k < \infty$. Hence, since F is differentiable at ξ_p, there exists $0 < \delta' < \delta$ such that

$$P\{\xi_p < X \le \xi_p + \delta'\} \le \lambda_i' \frac{(a_0/2)^{(1+\epsilon)/\epsilon}}{\upsilon_i^{1/\epsilon}}, \tag{51}$$

where we recall $\lambda_i' = P[Y \in S_i]$. Then

$$P_*\left\{|Z_{i,n}(\tilde{\xi}_{p,n}) - z_i(\xi_p)| > a_0\right\} = P_*\left\{|Z_{i,n}(\tilde{\xi}_{p,n}) - z_i(\xi_p)| > a_0, |\tilde{\xi}_{p,n} - \xi_p| \le \delta'\right\} + P_*\left\{|Z_{i,n}(\tilde{\xi}_{p,n}) - z_i(\xi_p)| > a_0, |\tilde{\xi}_{p,n} - \xi_p| > \delta'\right\} \equiv q_{1,n} + q_{2,n}, \quad (52)$$

so we want to show $q_{1,n} \to 0$ and $q_{2,n} \to 0$ as $n \to \infty$ to establish (50). Since $\tilde{\xi}_{p,n} \xrightarrow{P} \xi_p$ by (9),

$$q_{2,n} \le P_*\left\{|\tilde{\xi}_{p,n} - \xi_p| > \delta'\right\} \to 0 \text{ as } n \to \infty. \quad (53)$$

We now handle $q_{1,n}$ in (52). Because $Z_{i,n}(x)$ is monotonically increasing in x, we have $|Z_{i,n}(\tilde{\xi}_{p,n}) - z_i(\xi_p)| \le \max(|Z_{i,n}(\xi_p + \delta') - z_i(\xi_p)|, |Z_{i,n}(\xi_p - \delta') - z_i(\xi_p)|)$ when $|\tilde{\xi}_{p,n} - \xi_p| \le \delta'$. Hence,

$$q_{1,n} \le P_*\left\{|Z_{i,n}(\xi_p + \delta') - z_i(\xi_p)| > a_0, |\tilde{\xi}_{p,n} - \xi| \le \delta'\right\} + P_*\left\{|Z_{i,n}(\xi_p - \delta') - z_i(\xi_p)| > a_0, |\tilde{\xi}_{p,n} - \xi| \le \delta'\right\} \equiv r_{1,n} + r_{2,n}. \quad (54)$$

Observe that

$$r_{1,n} \le P_*\left\{|Z_{i,n}(\xi_p + \delta') - z_i(\xi_p)| > a_0\right\} \le P_*\left\{|Z_{i,n}(\xi_p + \delta') - z_i(\xi_p + \delta')| + |z_i(\xi_p + \delta') - z_i(\xi_p)| > a_0\right\} \quad (55)$$

by the triangle inequality. Also, a change of measure and the fact $I^2(\cdot) = I(\cdot)$ yield

$$\begin{aligned}
&|z_i(\xi_p + \delta') - z_i(\xi_p)| \\
&= |E[I(X_{ij} \le \xi_p + \delta')L_{ij}] - E[I(X_{ij} \le \xi_p)L_{ij}]| \\
&= E[I^2(\xi_p < X_{ij} \le \xi_p + \delta')L_{ij}] \\
&\le P^{\epsilon/(1+\epsilon)}\{\xi_p < X_{ij} \le \xi_p + \delta'\}\, E^{1/(1+\epsilon)}[I^{1+\epsilon}(\xi_p < X_{ij} \le \xi_p + \delta')L_{ij}^{1+\epsilon}] \\
&\le \left(\frac{P\{\xi_p < X \le \xi_p + \delta', Y \in S_i\}}{\lambda_i'}\right)^{\epsilon/(1+\epsilon)} \upsilon_i^{1/(1+\epsilon)} \\
&\le \left(\frac{P\{\xi_p < X \le \xi_p + \delta'\}}{\lambda_i'}\right)^{\epsilon/(1+\epsilon)} \upsilon_i^{1/(1+\epsilon)} \ \le\ a_0/2,
\end{aligned} \quad (56)$$

where the third step follows from Hölder's inequality, the fourth step holds because $I(\xi_p < X_{ij} \le \xi_p + \delta') \le I(X_{ij} \le \xi_p + \delta)$ as $\delta' < \delta$, and (51) implies the last step. Thus, using (56) in (55) gives

$$r_{1,n} \le P_*\left\{|Z_{i,n}(\xi_p + \delta') - z_i(\xi_p + \delta')| > a_0/2\right\}. \quad (57)$$

Now $Z_{i,n}(\xi_p+\delta')$ is the average of $I(X_{ij} \le \xi_p+\delta')L_{ij}^2$, $j = 1, 2, \ldots, n_i$, which are i.i.d. with mean $z_i(\xi_p + \delta') = E_*[I(X_{ij} \le \xi_p + \delta')L_{ij}^2]$, which is finite since $\delta' < \delta$ and $\upsilon_i < \infty$. Thus, the weak law of large numbers ensures the RHS of (57) converges to 0 as $n \to \infty$, so $r_{1,n} \to 0$ as $n \to \infty$. We can similarly prove that $r_{2,n}$ in (54) satisfies $r_{2,n} \to 0$, so $q_{1,n} \to 0$ by (54). Combining this with (52) and (53) then establishes (50). For the second term on the RHS of (22), similar arguments show that $\frac{1}{n_i}\sum_{j=1}^{n_i} I(X_{ij} \le \tilde{\xi}_{p,n})L_{ij} \xrightarrow{P} E_*[I(X_{ij} \le \xi_p)L_{ij}] = P\{X \le \xi_p \mid Y \in S_i\}$ as $n \to \infty$, so (49) holds by the continuous-mapping theorem, completing the proof.

B.4. Proof of Theorem 5.1

We will show that the AV CDF estimator $\tilde{F}_n$ in (24) satisfies Assumptions A1, A2 for all $a_n \to 0$, and A3, so the result will then follow from Theorem 3.1. Since both $I(X_i \le x)$ and $I(X_i' \le x)$ are nonnegative, $\tilde{F}_n(x)$ is monotonically increasing in x for each n, so Assumption A1 holds.

Next we will show Assumption A2 is satisfied for all $a_n \to 0$ by verifying Condition C1 for all $a_n \to 0$ and applying Proposition B.1. Observe that $F_n(x)$ is a special case of the IS+SS estimator with a single stratum (i.e., no SS) and IS measure $P_* = P$ (so the likelihood ratio $L \equiv 1$, which means no IS). Also, $F_n'(x) \stackrel{D}{=} F_n(x)$ since each $X_i' \stackrel{D}{=} X_i$, where $\stackrel{D}{=}$ denotes equality in distribution, so

$$E[\tilde{F}_n(x)] = \frac{1}{2}[E(F_n(x)) + E(F_n'(x))] = F(x)$$

by (44). Therefore, the AV CDF estimator is unbiased.

We now show (41) holds with $s_n(a_n) \to 0$ for all $a_n \to 0$. Note that (24) implies

$$E\left[\tilde{F}_n(\xi_p + a_n) - \tilde{F}_n(\xi_p)\right]^2 = \frac{1}{4}(b_{1,n} + b_{2,n} + b_{3,n}), \tag{58}$$

where

$$\begin{aligned} b_{1,n} &= E\left[F_n(\xi_p + a_n) - F_n(\xi_p)\right]^2, \\ b_{2,n} &= E\left[F_n'(\xi_p + a_n) - F_n'(\xi_p)\right]^2, \\ b_{3,n} &= 2E\left(\left[F_n(\xi_p + a_n) - F_n(\xi_p)\right]\left[F_n'(\xi_p + a_n) - F_n'(\xi_p)\right]\right). \end{aligned}$$

As a special case of the IS+SS CDF estimator, $F_n(x)$ also satisfies (41) with $s_n(a_n) \to 0$ for all $a_n \to 0$, as shown in the proof of Theorem 4.1. Therefore, since $F_n'(x) \stackrel{D}{=} F_n(x)$ for all x, we have

$$b_{1,n} = b_{2,n} = d_n^2 + \frac{s_n(a_n)}{n} \tag{59}$$

with $s_n(a_n) \to 0$ as $n \to \infty$, where $d_n = F(\xi_p + a_n) - F(\xi_p)$. To handle $b_{3,n}$, let $\rho_n = \min(\xi_p, \xi_p + a_n)$ and $\rho_n' = \max(\xi_p, \xi_p + a_n)$. Then

$$b_{3,n} = \frac{2}{n^2} E\left[\left(\sum_{i=1}^n I(\rho_n < X_i \le \rho_n')\right)\left(\sum_{j=1}^n I(\rho_n < X_j' \le \rho_n')\right)\right] = \frac{2}{n^2}(c_{1,n} + c_{2,n}), \tag{60}$$

where

$$\begin{aligned} c_{1,n} &= \sum_{i=1}^n E\left[I(\rho_n < X_i \le \rho_n')\, I(\rho_n < X_i' \le \rho_n')\right] \\ &= nP(\rho_n < X \le \rho_n',\ \rho_n < X' \le \rho_n') \end{aligned} \tag{61}$$

with (X, X') a generic antithetic pair and

$$c_{2,n} = E\left[\sum_{i \ne j} I(\rho_n < X_i \le \rho_n')\, I(\rho_n < X_j' \le \rho_n')\right] = n(n-1)d_n^2 \tag{62}$$

since X_i and X'_j are independent for $i \neq j$. Hence, (58)–(62) imply

$$\begin{aligned} & E\left[\tilde{F}_n(\xi_p + a_n) - \tilde{F}_n(\xi_p)\right]^2 \\ & = d_n^2 + \frac{1}{2n}\left[s_n(a_n) + P(\rho_n < X \le \rho'_n,\, \rho_n < X' \le \rho'_n) - d_n^2\right] \\ & \equiv d_n^2 + \frac{1}{2n} t_n. \end{aligned}$$

Since $P(\rho_n < X \le \rho'_n,\, \rho_n < X' \le \rho'_n) \le P(\rho_n < X \le \rho'_n) = |d_n|$, the triangle inequality yields $|t_n| \le |s_n(a_n)| + |d_n| + d_n^2$. Because the differentiability of F at ξ_p implies F is continuous at ξ_p, it follows that $d_n \to 0$ as $n \to \infty$ since $\rho_n \to \xi_p$ and $\rho'_n \to \xi_p$ as $n \to \infty$. Thus, $t_n \to 0$ as $n \to \infty$ since $s_n(a_n) \to 0$, which shows Condition C1 holds for all $a_n \to 0$. Hence, Assumption A2 follows by Proposition B.1.

Lastly, we need to show Assumption A3 holds. We express (24) as $\tilde{F}_n(\xi_p) = (1/n)\sum_{i=1}^n A_i$, where $A_i = [I(X_i \le \xi_p) + I(X'_i \le \xi_p)]/2$, $i = 1, 2, \ldots, n$, are i.i.d. as (X_i, X'_i), $i = 1, 2, \ldots, n$, are i.i.d. pairs. Also, $E[A_i] = F(\xi_p)$, and

$$\begin{aligned} \text{Var}(A_i) &= \frac{1}{4}\left\{\text{Var}[I(X_i \le \xi_p)] + \text{Var}[I(X'_i \le \xi_p)] + 2\text{Cov}[I(X_i \le \xi_p), I(X'_i \le \xi_p)]\right\} \\ &= \frac{1}{2}\left[p(1-2p) + P(X \le \xi_p, X' \le \xi_p)\right], \end{aligned}$$

which is finite. Thus, the CLT in Assumption A3 holds, and the variance ψ_p^2 in the CLT is given by $\text{Var}(A_i)$, which completes the proof of (i) and (iii). We omit the proof of (ii) and (iv) as they can be established by applying arguments similar to those employed in the proof of Theorem 4.1(ii) and (iv).

B.5. Proof of Theorem 6.1

We will show that the CV CDF estimator $\tilde{F}_n$ in (29) satisfies Assumptions A1, A2 for all $a_n \to 0$, and A3, so the result then follows from Theorem 3.1. The alternative representation of $\tilde{F}_n$ in (31) shows that $\tilde{F}_n(x)$ is monotonically increasing in x when the weights H_i in (32) are nonnegative. As noted by Hesterberg and Nelson [1998], the probability of any H_i being negative is $o(n^{-1})$ when $\text{Var}[C] < \infty$, as we have assumed, so Assumption A1 holds.

We next show Assumption A2 holds for all $a_n \to 0$. Define

$$\begin{aligned} W_n &= \sqrt{n}[p - \tilde{F}_n(\xi_p)], \\ Z_n &= \sqrt{n}\left[F(\xi_p + a_n) - \tilde{F}_n(\xi_p + a_n)\right], \end{aligned}$$

so we need to verify $Z_n - W_n \xrightarrow{P} 0$ as $n \to \infty$. Recalling (1) and (29), we can write

$$\begin{aligned} Z_n - W_n = {} & \sqrt{n}\left[(F(\xi_p + a_n) - F(\xi_p)) - (F_n(\xi_p + a_n) - F_n(\xi_p))\right] \\ & + \sqrt{n}[\bar{C}_n - \nu][\hat{\beta}_n(\xi_p + a_n) - \hat{\beta}_n(\xi_p)] \;\equiv\; A_n + B_n. \end{aligned} \tag{63}$$

Since F_n is a special case of the IS+SS CDF estimator in (20) with $P_* = P$ (so the likelihood ratio $L \equiv 1$, which means no IS) and only $k = 1$ stratum (i.e., no SS), the proof of Theorem 4.1 shows that F_n satisfies Assumption A2 for all $a_n \to 0$, which is exactly that $A_n \xrightarrow{P} 0$ as $n \to \infty$. By Slutsky's theorem, it then suffices to show that $B_n \xrightarrow{P} 0$ as $n \to \infty$ to verify that Assumption A2 holds for all $a_n \to 0$ for CV.

Let $\sigma_C^2 = \text{Var}[C]$. Since we assumed $0 < \sigma_C^2 < \infty$, the CLT then implies

$$\sqrt{n}(\bar{C}_n - \nu) \xrightarrow{L} N(0, \sigma_C^2) \text{ as } n \to \infty. \tag{64}$$

Thus, if we prove that $Q_n \equiv \hat{\beta}_n(\xi_p + a_n) - \hat{\beta}_n(\xi_p) \xrightarrow{P} 0$ as $n \to \infty$, then $B_n \xrightarrow{P} 0$ by Slutsky's theorem. Note that (28) implies

$$Q_n = \frac{\frac{1}{n}\sum_{i=1}^{n}[I(X_i \le \xi_p + a_n) - I(X_i \le \xi_p)](C_i - \bar{C}_n)}{\frac{1}{n}\sum_{k=1}^{n}(C_k - \bar{C}_n)^2} \equiv \frac{N_n}{D_n},$$

and $D_n \xrightarrow{P} \sigma_C^2 > 0$ as $n \to \infty$ since the C_i are i.i.d.; e.g., see p. 69 of Serfling [1980]. Thus, to prove that $Q_n \xrightarrow{P} 0$, it is sufficient to show that $N_n \xrightarrow{P} 0$ as $n \to \infty$ by Slutsky's theorem. We accomplish this by next verifying that $E(|N_n|) \to 0$ as $n \to \infty$ and applying Theorem 1.3.2 of Serfling [1980].

Let $\rho_n = \min(\xi_p, \xi_p + a_n)$ and $\rho_n' = \max(\xi_p, \xi_p + a_n)$, so the triangle inequality yields

$$\begin{aligned} E(|N_n|) &\le \frac{1}{n}\sum_{i=1}^{n} E\left[I(\rho_n < X_i \le \rho_n')\left|C_i - \bar{C}_n\right|\right] \\ &\le \frac{1}{n}\sum_{i=1}^{n} P^{1/2}(\rho_n < X_i \le \rho_n')\, E^{1/2}[(C_i - \bar{C}_n)^2] \end{aligned}$$

by the Cauchy-Schwarz inequality. Because $E[(C_i - \bar{C}_n)^2] = (n-1)\sigma_C^2/n$, we have $E(|N_n|) < P^{1/2}(\rho_n < X \le \rho_n')\sigma_C$ since the X_i are i.i.d. Also, $\rho_n \to \xi_p$ and $\rho_n' \to \xi_p$ as $n \to \infty$, so $P(\rho_n < X \le \rho_n') \to 0$ as $n \to \infty$ because F is differentiable at ξ_p. Hence, $E(|N_n|) \to 0$ as $n \to \infty$ since $\sigma_C < \infty$. Consequently, $B_n \xrightarrow{P} 0$ as $n \to \infty$, showing that Assumption A2 is satisfied for any $a_n \to 0$ by (63).

To prove Assumption A3 holds for CV, note that $\hat{\beta}_n(\xi_p) \xrightarrow{P} \beta_*(\xi_p)$ as $n \to \infty$ by the weak law of large numbers and Slutsky's theorem. Thus,

$$\begin{aligned} &\sqrt{n}[F_n(\xi_p) - \beta_*(\xi_p)(\bar{C}_n - \nu)] - \sqrt{n}[F_n(\xi_p) - \hat{\beta}_n(\xi_p)(\bar{C}_n - \nu)] \\ &= [\hat{\beta}_n(\xi_p) - \beta_*(\xi_p)]\sqrt{n}(\bar{C}_n - \nu) \xrightarrow{L} 0 \cdot N(0, \sigma_C^2) = 0 \text{ as } n \to \infty \end{aligned}$$

by (64) and Slutsky's theorem, so $\sqrt{n}[F_n(\xi_p) - \beta_*(\xi_p)(\bar{C}_n - \nu)]$ and $\sqrt{n}[F_n(\xi_p) - \hat{\beta}_n(\xi_p)(\bar{C}_n - \nu)]$ have the same weak limit as $n \to \infty$ by the converging-together lemma (e.g., Theorem 25.4 of Billingsley [1995]). Since the summands in (27) with β replaced with $\beta_*(\xi_p)$ are i.i.d. with finite variance (because we assumed Var$[C] < \infty$), the CLT gives the weak limit as $N(0, \psi_p^2)$, with $\psi_p^2 = \text{Var}[I(X \le \xi_p) - \beta_*(\xi_p)(C - \nu)]$, which works out to be (30). This completes the proof of (i) and (iii). We omit the proof of (ii) and (iv) as they can be established by applying arguments similar to those employed in the proof of Theorem 4.1(ii) and (iv).

Received R; revised e; accepted c
eived Month Year; revised Month Year; accepted Month Year

Table VI. Coverage levels (average half widths) for larger SAN wtih $h_n = 0.5n^{-1/2}$ and $p = 0.95$

n	CFD	FFD	BFD	Comb. CFD	Exact ϕ_p	Batch	Est. Mean
CMC							
100	0.949 (2.663)	0.979 (4.243)	0.685 (1.160)	0.949 (2.663)	0.907 (1.741)	0.856 (1.676)	0.894 (0.530)
400	0.900 (0.928)	0.937 (1.149)	0.789 (0.706)	0.836 (0.807)	0.902 (0.871)	0.679 (0.842)	0.903 (0.266)
1600	0.891 (0.443)	0.920 (0.495)	0.836 (0.391)	0.881 (0.433)	0.898 (0.435)	0.833 (0.457)	0.898 (0.133)
6400	0.897 (0.219)	0.910 (0.232)	0.870 (0.206)	0.893 (0.218)	0.900 (0.218)	0.882 (0.235)	0.905 (0.067)
AV							
100	0.951 (1.688)	0.984 (2.518)	0.717 (0.837)	0.951 (1.688)	0.902 (1.217)	0.509 (1.062)	0.890 (0.336)
400	0.914 (0.658)	0.954 (0.819)	0.802 (0.497)	0.850 (0.566)	0.901 (0.609)	0.778 (0.624)	0.897 (0.169)
1600	0.895 (0.311)	0.923 (0.348)	0.847 (0.274)	0.884 (0.304)	0.894 (0.305)	0.865 (0.328)	0.893 (0.084)
6400	0.896 (0.153)	0.912 (0.163)	0.873 (0.144)	0.894 (0.153)	0.898 (0.152)	0.893 (0.166)	0.890 (0.042)
CV							
100	0.928 (2.445)	0.954 (3.746)	0.689 (1.137)	0.928 (2.445)	0.873 (1.615)	0.588 (1.571)	0.895 (0.441)
400	0.907 (0.902)	0.938 (1.126)	0.799 (0.678)	0.829 (0.765)	0.896 (0.821)	0.783 (0.899)	0.901 (0.222)
1600	0.893 (0.422)	0.921 (0.472)	0.838 (0.372)	0.881 (0.412)	0.899 (0.412)	0.898 (0.462)	0.901 (0.111)
6400	0.897 (0.208)	0.909 (0.220)	0.869 (0.196)	0.893 (0.207)	0.900 (0.206)	0.898 (0.225)	0.903 (0.056)
IS+SS							
100	0.984 (1.213)	0.997 (1.874)	0.715 (0.539)	0.984 (1.213)	0.881 (0.766)	0.848 (0.956)	0.882 (1.461)
400	0.920 (0.435)	0.962 (0.542)	0.804 (0.327)	0.855 (0.368)	0.894 (0.399)	0.886 (0.450)	0.898 (0.747)
1600	0.902 (0.206)	0.932 (0.230)	0.847 (0.181)	0.892 (0.201)	0.894 (0.202)	0.891 (0.220)	0.898 (0.376)
6400	0.904 (0.102)	0.918 (0.108)	0.875 (0.096)	0.901 (0.101)	0.903 (0.101)	0.901 (0.110)	0.904 (0.188)

Table VII. Coverage levels (average half widths) for larger SAN wtih $h_n = 0.5n^{-1/2}$ and $p = 0.99$

n	CFD	FFD	BFD	Comb. CFD	Exact ϕ_p	Batch	Est. Mean
	CMC						
100	0.508 (2.063)	0.693 (4.126)	0.410 (1.134)	0.508 (2.063)	0.941 (3.686)	0.042 (1.676)	0.894 (0.530)
400	0.925 (2.649)	0.964 (4.217)	0.579 (0.898)	0.925 (2.649)	0.912 (1.843)	0.740 (1.614)	0.903 (0.266)
1600	0.980 (1.541)	0.998 (2.437)	0.707 (0.601)	0.980 (1.541)	0.898 (0.922)	0.898 (1.020)	0.898 (0.133)
6400	0.939 (0.540)	0.980 (0.718)	0.799 (0.363)	0.892 (0.470)	0.897 (0.461)	0.894 (0.499)	0.905 (0.067)
	AV						
100	0.790 (2.642)	0.870 (4.315)	0.429 (0.876)	0.790 (2.642)	0.918 (2.596)	0.260 (1.632)	0.890 (0.336)
400	0.977 (2.381)	0.994 (3.873)	0.589 (0.652)	0.977 (2.381)	0.904 (1.300)	0.839 (1.588)	0.897 (0.169)
1600	0.982 (1.030)	0.998 (1.600)	0.707 (0.429)	0.982 (1.030)	0.893 (0.650)	0.834 (0.683)	0.893 (0.084)
6400	0.940 (0.384)	0.985 (0.510)	0.796 (0.257)	0.897 (0.333)	0.893 (0.325)	0.890 (0.361)	0.890 (0.042)
	CV						
100	0.434 (1.793)	0.480 (2.722)	0.323 (1.000)	0.434 (1.793)	0.752 (2.953)	0.016 (1.533)	0.895 (0.441)
400	0.902 (2.469)	0.925 (3.666)	0.573 (0.915)	0.902 (2.469)	0.874 (1.732)	0.693 (1.565)	0.901 (0.222)
1600	0.980 (1.460)	0.995 (2.281)	0.713 (0.584)	0.980 (1.460)	0.892 (0.875)	0.868 (0.970)	0.901 (0.111)
6400	0.943 (0.521)	0.982 (0.694)	0.798 (0.348)	0.898 (0.455)	0.896 (0.439)	0.902 (0.498)	0.903 (0.056)
	IS+SS						
100	0.977 (1.410)	0.994 (2.199)	0.424 (0.320)	0.977 (1.410)	0.852 (0.863)	0.796 (1.248)	0.886 (1.993)
400	0.990 (0.748)	0.999 (1.164)	0.582 (0.236)	0.990 (0.748)	0.883 (0.457)	0.875 (0.539)	0.896 (1.019)
1600	0.993 (0.379)	1.000 (0.589)	0.710 (0.154)	0.993 (0.379)	0.891 (0.231)	0.895 (0.260)	0.900 (0.512)
6400	0.943 (0.138)	0.988 (0.183)	0.797 (0.092)	0.904 (0.120)	0.893 (0.116)	0.901 (0.129)	0.903 (0.256)