Quantile Estimation When Applying Conditional Monte Carlo

Marvin K. Nakayama
Computer Science Department, New Jersey Institute of Technology, Newark, NJ 07102, U.S.A.

Keywords: Quantile, Value-at-Risk, Variance Reduction, Conditional Monte Carlo, Confidence Interval.

Abstract: We describe how to use conditional Monte Carlo (CMC) to estimate a quantile. CMC is a variance-reduction technique that reduces variance by analytically integrating out some of the variability. We show that the CMC quantile estimator satisfies a central limit theorem and Bahadur representation. We also develop three asymptotically valid confidence intervals (CIs) for a quantile. One CI is based on a finite-difference estimator, another uses batching, and the third applies sectioning. We present numerical results demonstrating the effectiveness of CMC.

1 INTRODUCTION

For a continuous random variable X with a strictly increasing cumulative distribution function (CDF) F and fixed $0 < p < 1$, the p-quantile of X is defined as the constant ξ such that $P(X \leq \xi) = p$. A well-known example is the median, which is the 0.5-quantile. The p-quantile ξ also can equivalently be expressed as $\xi = F^{-1}(p)$.

Quantiles are often used in application areas to measure risk. For example, in finance, a quantile is known as a value-at-risk, and quantiles are widely employed to assess portfolio risk. For example, banking regulations specify capital requirements for a firm in terms of 0.99-quantiles of the random loss (Jorion, 2007). In nuclear engineering, safety and uncertainty analyses are often performed with a 0.95-quantile (U.S. Nuclear Regulatory Commission, 1989).

Suppose that we have a simulation model that outputs a random variable X. When applying simple random sampling (SRS), the typical approach to estimate the p-quantile ξ is to run independent and identically distributed (i.i.d.) replications of the model, and form an estimator of the CDF from the sample outputs. Inverting the CDF estimator yields a quantile estimator.

Because of the noise inherent in any stochastic simulation, the quantile estimator has some error, which should be measured. A standard way of assessing the error is by forming a confidence interval for the true quantile ξ. For example, the U.S. Nuclear Regulatory Commission requires nuclear plant licensees to satisfy a so-called 95/95 criterion, which entails establishing, with 95% confidence, that the 0.95-quantile lies below a mandated threshold; see Section 24.9 of (U.S. Nuclear Regulatory Commission, 2011). Thus, we need not only a point estimate of a quantile but also a confidence interval for it.

There are several approaches to construct a CI when applying SRS. One technique, which is sometimes called the nonparametric method, exploits a binomial property of the i.i.d. sample; see Section 2.6.1 of (Serfling, 1980). Another way first shows that the quantile estimator satisfies a central limit theorem (CLT), and then unfolds the CLT to obtain a CI. The key to applying this technique is consistently estimating the asymptotic variance constant appearing in the CLT; approaches for accomplishing this include using a finite difference (Bloch and Gastwirth, 1968; Bofinger, 1975) and kernel methods (Falk, 1986). Rather than consistently estimating the asymptotic variance, we can instead apply batching or sectioning, the latter of which was originally developed for SRS in Section III.5a of (Asmussen and Glynn, 2007) and extended in (Nakayama, 2014a) to work when applying the variance-reduction techniques control variates and importance sampling. Batching and sectioning divide the i.i.d. outputs into independent batches, computing a quantile estimator from each batch, and constructing a CI from the batch quantile estimators.

In this paper, we use conditional Monte Carlo to estimate a quantile. CMC reduces variance (compared to SRS) by analytically integrating out the variability that remains after conditioning on an auxiliary random variable Y; e.g., see Section 8.3 of (Ross, 2006) or Section V.4 of (Asmussen and Glynn, 2007). We prove that the CMC quantile estimator satisfies a

K. Nakayama M..
Quantile Estimation When Applying Conditional Monte Carlo.
DOI: 10.5220/0005109702800285
In *Proceedings of the 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (SIMULTECH-2014), pages 280-285
ISBN: 978-989-758-038-3

CLT and a Bahadur representation (Bahadur, 1966). The latter shows that a quantile estimator can be approximated as the true quantile plus a linear transformation of the corresponding CDF estimator, with a remainder term that vanishes at some rate as the sample size grows. Since the CDF estimator is typically a sample average, it satisfies a CLT under appropriate conditions. Thus, the Bahadur representation provides insight into why a quantile estimator, which is *not* a sample average, satisfies a CLT. It also allows us to construct asymptotically valid CIs for ξ by using a finite difference or sectioning, and we develop those CIs in this paper.

CMC has previously been employed to derive an estimator of a sensitivity of a quantile with respect to a model parameter (Fu et al., 2009). For example, suppose a financial investor has a portfolio of loans, each of which may default. The investor may want to estimate the sensitivity of the 0.99-quantile of the loss of the portfolio, where the sensitivity is taken with respect to a parameter of the loss distribution of an individual obligor. While (Fu et al., 2009) apply CMC to estimate quantile sensitivities, the method has not been used (to the best of our knowledge) to estimate the quantile itself.

The rest of the paper develops as follows. Section 2 reviews how to apply SRS to estimate and construct CIs for a quantile ξ. Section 3 develops our CMC estimator of a quantile, shows that it satisfies a CLT and Bahadur representation, and uses these results to construct CIs for ξ. Section 4 presents numerical results from a simple model, and we provide concluding remarks in Section 5. Proofs of the results are given in (Nakayama, 2014b).

2 SIMPLE RANDOM SAMPLING

Let X be a random variable with CDF F. We first review how to estimate and construct confidence intervals for the p-quantile $\xi = F^{-1}(p) \equiv \inf\{x : F(x) \geq p\}$ of F (or equivalently of X) for a fixed $0 < p < 1$ when applying simple random sampling (SRS).

Let X_i, $i = 1, 2, \ldots, n$, be a sample of n i.i.d. observations from F. The SRS estimator of $F(x) = E[I(X \leq x)]$ is the empirical distribution function $F_n(x) = (1/n)\sum_{i=1}^{n} I(X_i \leq x)$. The SRS p-quantile estimator is $\xi_n = F_n^{-1}(p)$. We can alternatively compute ξ_n by first sorting the sample $X_1, X_2, \ldots, X_n$ into the order statistics $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, and then setting $\xi_n = X_{(\lceil np \rceil)}$, where $\lceil \cdot \rceil$ denotes the ceiling function.

Section 2.3 of (Serfling, 1980) provides an overview of ξ_n and its properties. For example, let f denote the derivative (when it exists) of F. If $f(\xi) > 0$, then ξ_n satisfies the following CLT:

$$\sqrt{n}(\xi_n - \xi) \Rightarrow N(0, p(1-p)/f^2(\xi)) \qquad (1)$$

as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution (e.g., see Chapter 5 of (Billingsley, 1995)), and $N(a, b^2)$ is a normal random variable with mean a and variance b^2. Moreover, ξ_n also satisfies a so-called (weak) Bahadur representation (Bahadur, 1966; Ghosh, 1971):

$$\xi_n = \xi - \frac{F_n(\xi) - p}{f(\xi)} + R'_n, \quad \text{with} \quad \sqrt{n}R'_n \Rightarrow 0 \qquad (2)$$

as $n \to \infty$. By (2), the left side of (1) equals $-\sqrt{n}(F_n(\xi) - p)/f(\xi) + \sqrt{n}R'_n$, where the first term converges weakly to the right side of (1), and the second weakly vanishes by (2). Thus, the Bahadur representation provides insight into why ξ_n, which is *not* a sample average, satisfies a CLT, as it can be approximated in terms of the empirical distribution, which is a sample mean.

(Ghosh, 1971) also establishes a version of (2) for the p_n-quantile with perturbed p_n that converges to p, rather than the p-quantile for fixed p. This variation can be useful for constructing a consistent estimator of $\lambda \equiv 1/f(\xi)$, which appears in the asymptotic variance in (1) and can be used to construct a confidence interval for ξ. If $f(\xi) > 0$, then for any $p_n = p + O(n^{-1/2})$, the SRS estimator $F_n^{-1}(p_n)$ of the p_n-quantile $F^{-1}(p_n)$ satisfies

$$F_n^{-1}(p_n) = \xi'_{p_n} - \frac{F_n(\xi) - p}{f(\xi)} + R'_n, \quad \text{with} \quad \sqrt{n}R_n \Rightarrow 0 \qquad (3)$$

as $n \to \infty$, where

$$\xi'_{p_n} = \xi + (p_n - p)/f(\xi). \qquad (4)$$

To see how to use these results to consistently estimate λ, first note that $\lambda = \frac{d}{dp}F^{-1}(p) = \lim_{h \to 0}[F^{-1}(p+h) - F^{-1}(p-h)]/2h$. This suggests estimating λ with the *finite difference*

$$\lambda_n = \frac{F_n^{-1}(p + h_n) - F_n^{-1}(p - h_n)}{2h_n}, \qquad (5)$$

where $h_n > 0$ is known as the *bandwidth*. The terms in the numerator of the finite difference are precisely in the form of (3) with $p_n = p \pm h_n$, which allows proving $\lambda_n \Rightarrow \lambda$ as $n \to \infty$ when $h_n = cn^{-1/2}$ for any constant $c > 0$; see Section 2.6.3 of (Serfling, 1980). Using a different proof technique, (Bloch and Gastwirth, 1968) and (Bofinger, 1975) also show $\lambda_n \Rightarrow \lambda$ when f is continuous in a neighborhood of ξ, and $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Thus, unfolding the CLT in (1) leads to the following two-sided $(1-\alpha)$-level $(0 < \alpha < 1)$ confidence interval for ξ:

$$I_n = [\xi_n \pm z_\alpha \sqrt{p(1-p)}\lambda_n/\sqrt{n}], \qquad (6)$$

where $z_\alpha = \Phi^{-1}(1-\alpha/2)$ and Φ is the CDF of a standard (mean 0 and unit variance) normal. The CI I_n is asymptotically valid in the sense that $P(\xi \in I_n) \to 1-\alpha$ as $n \to \infty$.

Determining an appropriate value for the bandwidth h_n in the finite difference can be difficult in practice. Alternatively, we can avoid trying to consistently estimate λ by instead applying batching or sectioning. In batching, we divide the n outputs $X_1, X_2, \ldots, X_n$ into $b \geq 2$ equal-sized batches, where the jth batch, $j = 1, 2, \ldots, b$, consists of the $m = n/b$ outputs $X_{(j-1)m+i}$, $i = 1, 2, \ldots, m$. A reasonable choice for the number of batches is $b = 10$. For each batch j, we define the CDF estimator $F_{j,m}(x) = (1/m)\sum_{i=1}^{m} I(X_{(j-1)m+i} \leq x)$ and corresponding p-quantile estimator $\xi_{j,m} = F_{j,m}^{-1}(p)$. Since the n outputs are i.i.d., the b batches are i.i.d., so $\xi_{j,m}$, $j = 1, 2, \ldots, b$, are i.i.d. We compute their sample average $\bar{\xi}_{b,m} = (1/b)\sum_{j=1}^{b} \xi_{j,m}$ and their sample variance $S_{b,m}^2 = (1/(b-1))\sum_{j=1}^{b}(\xi_{j,m} - \bar{\xi}_{b,m})^2$. An asymptotically valid (as $m \to \infty$ with $b \geq 2$ fixed) $(1-\alpha)$-level CI for ξ using batching is then

$$J_n = [\bar{\xi}_{b,m} \pm t_\alpha S_{b,m}/\sqrt{b}],$$

where $t_\alpha = T_{b-1}^{-1}(1-\alpha/2)$ and T_{b-1} is the CDF of a Student t distribution with $b-1$ degrees of freedom.

Similar to batching, sectioning was originally developed in Section III.5a of (Asmussen and Glynn, 2007) for SRS, and it replaces the batching point estimator $\bar{\xi}_{b,m}$ with the overall quantile estimator ξ_n. Specifically, let $S'^2_{b,m} = (1/(b-1))\sum_{j=1}^{b}(\xi_{j,m} - \xi_n)^2$, and the sectioning two-sided $(1-\alpha)$-level CI for ξ when applying SRS is

$$J'_n = [\xi_n \pm t_\alpha S'_{b,m}/\sqrt{b}].$$

The asymptotic validity of J'_n can be established by exploiting the Bahadur representation in (2) for fixed $p_n = p$. An advantage of sectioning over batching arises from the fact that quantile estimators are generally biased. While the bias decreases (nonmonotonically) to zero as the sample size n increases, it can be significant for small sample sizes. The bias of the batching point estimator $\bar{\xi}_{b,m}$ is determined by the batch size $m = n/b < n$, so $\bar{\xi}_{b,m}$ can be considerably more biased than the overall quantile estimator ξ_n, which has bias governed by the overall sample size n. Since the sectioning CI is centered at a less-biased point ξ_n, whereas the batching CI is centered at $\bar{\xi}_{b,m}$, the sectioning CI typically has better coverage than the batching CI for a fixed overall sample size $n = bm$; see the numerical results in Section 4.

3 CONDITIONAL MONTE CARLO

Now suppose that (X,Y) is a random vector with joint distribution H. As before, we want to estimate $\xi = F^{-1}(p)$, where F again denotes the (marginal) distribution of X. Let (X_i, Y_i), $i = 1, 2, \ldots, n$, be a sample of n i.i.d. pairs from H.

Since $F(x) = E[E[I(X \leq x)\,|\,Y]] = E[P(X \leq x\,|\,Y)]$, a conditional Monte Carlo estimator of $F(x)$ is

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} E[I(X_i \leq x)\,|\,Y_i] = \frac{1}{n}\sum_{i=1}^{n} G(Y_i, x), \quad (7)$$

where $G(Y,x) = P(X \leq x\,|\,Y)$. The CMC p-quantile estimator is $\hat{\xi}_n = \hat{F}_n^{-1}(p)$. Applying CMC relies critically on being able to compute G and invert $\hat{F}_n$.

Computing $\hat{F}_n^{-1}(p)$ for CMC appears to be more involved than for SRS or the other variance-reduction techniques examined in (Chu and Nakayama, 2012). For example, consider the simple case when (X,Y) has a bivariate normal distribution with zero marginal means, unit marginal variances, and correlation ρ. The conditional distribution of X given $Y = y$ is $N(\rho y, 1-\rho^2)$ (e.g., see pp. 167–168 of (Mood et al., 1974)), so

$$G(Y,x) = P(X \leq x\,|\,Y) = \Phi\left(\frac{x - \rho Y}{\sqrt{1-\rho^2}}\right), \quad (8)$$

and

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x - \rho Y_i}{\sqrt{1-\rho^2}}\right).$$

Identifying $\hat{\xi}_n$ such that $\hat{F}_n(\hat{\xi}_n) = p$, i.e., $\hat{\xi}_n = \hat{F}_n^{-1}(p)$, can be accomplished using a root-finding algorithm, e.g., Newton's method, the secant method or the false-position method; e.g., see Sections 7.1 and 7.2 of (Ortega and Rheinboldt, 1987). In contrast to the secant and false-position methods, Newton's method requires computing the derivative of $\hat{F}_n$. Given $Y_1, Y_2, \ldots, Y_n$, note that $\hat{F}_n(x)$ is strictly increasing and differentiable in x, with sample-path derivative

$$\frac{d}{dx}\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} (1-\rho^2)^{-1/2}\phi\left(\frac{x - \rho Y_i}{\sqrt{1-\rho^2}}\right),$$

where ϕ is the density of a standard normal.

The following shows that the CMC p_n-quantile estimator $\hat{F}_n^{-1}(p_n)$ satisfies a Bahadur representation.

Theorem 1. *If $f(\xi) > 0$, then for any $p_n = p + O(n^{-1/2})$, the CMC p_n-quantile estimator $\hat{F}_n^{-1}(p_n)$ satisfies*

$$\hat{F}_n^{-1}(p_n) = \xi_{p_n} - \frac{\hat{F}_n(\xi) - p}{f(\xi)} + R_n, \text{ with } \sqrt{n}R_n \Rightarrow 0 \quad (9)$$

as $n \to \infty$, where ξ'_{p_n} is given in (4).

In particular, the results of Theorem 1 hold for $p_n = p$ fixed, in which case $\xi'_{p_n} = \xi$. It then follows from (9) that the CMC p-quantile estimator $\hat{\xi}_n = \hat{F}_n^{-1}(p)$ satisfies the following CLT:

$$\sqrt{n}(\hat{\xi}_n - \xi) = -\frac{\sqrt{n}}{f(\xi)}(\hat{F}_n(\xi) - p) + \sqrt{n}R_n \Rightarrow N(0, \tau^2) \tag{10}$$

as $n \to \infty$, where

$$\tau^2 = \frac{\mathrm{Var}[G(Y,\xi)]}{f^2(\xi)}. \tag{11}$$

The numerator in the right side of (11) arises since $\hat{F}_n(\xi)$ is the sample average of the i.i.d. $G(Y_i,\xi)$, $i = 1,2,\ldots,n$. By the well-known variance-decomposition formula (e.g., see Section 2.10 of (Ross, 2006)), we have that

$$\begin{aligned} p(1-p) &= \mathrm{Var}[I(X \le \xi)] \\ &= E[\mathrm{Var}[I(X \le \xi)|Y]] + \mathrm{Var}[E[I(X \le \xi)|Y]] \\ &\ge \mathrm{Var}[E[I(X \le \xi)|Y]] = \mathrm{Var}[G(Y,\xi)], \end{aligned} \tag{12}$$

where the inequality follows from the nonnegativity of (conditional) variance. Thus, comparing (1) with (10) and (11), we see that the CMC p-quantile estimator $\hat{\xi}_n$ has smaller asymptotic variance than the SRS p-quantile estimator ξ_n.

The CLT in (10) can be unfolded to construct an asymptotically valid confidence interval for ξ if we can consistently estimate τ^2 in (11). Theorem 1 can be used to prove that the finite difference

$$\hat{\lambda}_n = \frac{\hat{F}_n^{-1}(p+h_n) - \hat{F}_n^{-1}(p-h_n)}{2h_n} \tag{13}$$

satisfies $\hat{\lambda}_n \Rightarrow \lambda = 1/f(\xi)$ as $n \to \infty$ when the bandwidth $h_n = c/\sqrt{n}$ for any constant $c > 0$. We can consistently estimate the numerator $\psi^2 \equiv \mathrm{Var}[G(Y,\xi)]$ in (11) via

$$\hat{\psi}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[G(Y_i,\hat{\xi}_n) - \bar{G}_n\right]^2,$$

where $\bar{G}_n = (1/n)\sum_{i=1}^{n} G(Y_i,\hat{\xi}_n)$. If $G(y,x)$ is continuous in x for each y, then $\bar{G}_n = p$ since $\hat{F}_n(\hat{\xi}_n) = p$. The proof of the consistency of $\hat{\psi}_n^2$ is complicated by the fact that $G(Y_i,\hat{\xi}_n)$, $i = 1,2,\ldots,n$, are not i.i.d. because they all depend on $\hat{\xi}_n$, which in turn is a function of all Y_i, $i = 1,2,\ldots,n$. But this can be handled employing arguments developed in (Chu and Nakayama, 2012). Now we can consistently estimate τ using $\hat{\tau}_n = \hat{\psi}_n\hat{\lambda}_n$, so an asymptotically valid two-sided CI for ξ is

$$\hat{I}_n = [\hat{\xi}_n \pm z_\alpha\hat{\tau}_n/\sqrt{n}]. \tag{14}$$

As with SRS, choosing an appropriate value for the bandwidth in the finite-difference $\hat{\lambda}_n$ can be difficult when applying CMC, and we may instead apply batching or sectioning to construct a CI for ξ with CMC. For batching, we divide the $G(Y_i,\cdot)$, $i = 1,2,\ldots,n$, into $b \ge 2$ nonoverlapping batches, each of size $m = n/b$. As with SRS, a reasonable choice for the number of batches is $b = 10$. For each $j = 1,2,\ldots,b$, the jth batch consists of $G(Y_{(j-1)m+i},\cdot)$, $i = 1,2,\ldots,m$, which we use to compute a CDF estimator $\hat{F}_{j,m}$, with

$$\hat{F}_{j,m}(x) = \frac{1}{m}\sum_{i=1}^{m} G(Y_{(j-1)m+i},x),$$

and p-quantile estimator $\hat{\xi}_{j,m} = \hat{F}_{j,m}^{-1}(p)$. The b batch quantile estimators $\hat{\xi}_{j,m}$, $j = 1,2,\ldots,b$, are i.i.d., and we compute their sample average $\tilde{\xi}_{b,m} = (1/b)\sum_{j=1}^{b}\hat{\xi}_{j,m}$ and sample variance $\tilde{S}_{b,m}^2 = (1/(b-1))\sum_{j=1}^{b}(\hat{\xi}_{j,m} - \tilde{\xi}_{b,m})^2$. The batching CI for ξ when applying CMC is then

$$\tilde{J}_{b,m} = [\tilde{\xi}_{b,m} \pm t_\alpha\tilde{S}_{b,m}/\sqrt{b}].$$

Because of the bias of quantile estimators, it is often better to apply sectioning instead of batching when n is small, where we again replace the batching point estimator $\tilde{\xi}_{b,m}$ with the overall quantile estimator $\hat{\xi}_n$. Define $\hat{S}_{b,m}^2 = (1/(b-1))\sum_{j=1}^{b}(\hat{\xi}_{j,m} - \hat{\xi}_n)^2$, and the sectioning CI for ξ when applying CMC is then

$$\hat{J}_{b,m} = [\hat{\xi}_n \pm t_\alpha\hat{S}_{b,m}/\sqrt{b}].$$

As with SRS, when applying CMC, the sectioning CI $\hat{J}_{b,m}$ should have better coverage than the batching CI $\tilde{J}_{b,m}$ for fixed overall sample size $n = bm$.

The following result establishes the asymptotic validity of the CMC CIs.

Theorem 2. *Suppose $f(\xi) > 0$. Then the following hold:*

(i) $P(\xi \in \tilde{J}_{b,m}) \to 1-\alpha$ and $P(\xi \in \hat{J}_{b,m}) \to 1-\alpha$ as $m \to \infty$ with $b \ge 2$ fixed.

(ii) If the bandwidth $h_n = cn^{-1/2}$ in (13) *for any constant $c > 0$, then $P(\xi \in \hat{I}_n) \to 1-\alpha$ as $n \to \infty$.*

4 NUMERICAL RESULTS

We next present numerical results from simulation experiments on the bivariate normal discussed in Section 3. Recall (X,Y) is bivariate normal, with marginal means 0, unit marginal variances, and correlation $\rho = 0.5$. Our goal is to estimate and construct

CIs for the p-quantile ξ of X for different values of p and different sample sizes n. Tables 1 and 2 contain the results when applying simple random sampling and conditional Monte Carlo, respectively, giving the estimated coverage of nominal 90% CIs for ξ and the average half widths (AHWs) of the CIs from 10^4 independent replications, where we use different methods to construct the CIs.

Table 1: Coverage (and average half width) of nominal 90% confidence intervals for the p-quantile of a standard normal X when applying simple random sampling (SRS) with bandwidth $h_n = 0.2n^{-1/2}$ and $b = 10$ batches.

$p = 0.8$				
n	Exact	FD	Batch	Section
100	0.900 (0.235)	0.808 (0.230)	0.620 (0.235)	0.903 (0.260)
400	0.897 (0.118)	0.851 (0.117)	0.821 (0.125)	0.909 (0.129)
1600	0.898 (0.059)	0.875 (0.059)	0.876 (0.063)	0.901 (0.064)
6400	0.902 (0.029)	0.877 (0.028)	0.898 (0.032)	0.905 (0.032)
$p = 0.95$				
n	Exact	FD	Batch	Section
100	0.902 (0.348)	0.799 (0.330)	0.825 (0.330)	0.861 (0.340)
400	0.902 (0.174)	0.848 (0.172)	0.646 (0.171)	0.900 (0.188)
1600	0.900 (0.087)	0.878 (0.087)	0.830 (0.092)	0.900 (0.095)
6400	0.901 (0.043)	0.891 (0.044)	0.883 (0.047)	0.902 (0.047)
$p = 0.99$				
n	Exact	FD	Batch	Section
100	0.926 (0.614)	0.497 (0.325)	0.024 (0.330)	0.762 (0.502)
400	0.905 (0.307)	0.913 (0.405)	0.696 (0.267)	0.841 (0.284)
1600	0.907 (0.154)	0.891 (0.160)	0.907 (0.164)	0.907 (0.168)
6400	0.902 (0.077)	0.893 (0.077)	0.887 (0.082)	0.902 (0.083)

In each table, the column labeled "Exact" contains the results for the CIs in (6) and (14) but where we replace the finite difference estimator of λ with its exact value. This method is typically not implementable in practice since λ is usually unknown, but we include results for it as a benchmark to which we compare the others. For the finite difference (FD), we use the bandwidth $h_n = 0.2/\sqrt{n}$ in (5) and (13). When $p \approx 1$ and n is small, we can have $p + h_n > 1$, so the finite differences (5) and (13) become undefined since the inverse of the estimated CDF is evaluated outside of its domain. In these cases, we replace $p + h_n$ and $p - h_n$ with $q_1 \equiv 1 - (1-p)/10$ and $q_2 \equiv 2p - 1 + (1-p)/10$, respectively, where q_2 is chosen so that q_1 and q_2 are symmetric around p; the denominator in the finite difference is then $q_1 - q_2$. The columns labeled "Batch" and "Section" are for batching and sectioning, respectively, with $b = 10$ batches. Numerical results in (Nakayama, 2014a) with $b = 10$ and $b = 20$ reveal that $b = 20$ often leads to poorer coverage than $b = 10$ for small n.

Table 2: Coverage (and average half width) of nominal 90% confidence intervals for the p-quantile of X of a bivariate normal (X,Y) with $\rho = 0.5$ when applying conditional Monte Carlo (CMC) with bandwidth $h_n = 0.2n^{-1/2}$ and $b = 10$ batches.

$p = 0.8$				
n	Exact	FD	Batch	Section
100	0.896 (0.087)	0.895 (0.087)	0.892 (0.093)	0.898 (0.093)
400	0.899 (0.043)	0.899 (0.043)	0.899 (0.047)	0.898 (0.047)
1600	0.900 (0.021)	0.899 (0.021)	0.897 (0.023)	0.897 (0.023)
6400	0.896 (0.011)	0.896 (0.011)	0.896 (0.012)	0.896 (0.012)
$p = 0.95$				
n	Exact	FD	Batch	Section
100	0.889 (0.105)	0.898 (0.109)	0.873 (0.102)	0.895 (0.103)
400	0.897 (0.049)	0.898 (0.050)	0.893 (0.052)	0.898 (0.052)
1600	0.892 (0.024)	0.893 (0.024)	0.895 (0.026)	0.897 (0.026)
6400	0.895 (0.012)	0.895 (0.012)	0.896 (0.013)	0.897 (0.013)
$p = 0.99$				
n	Exact	FD	Batch	Section
100	0.888 (0.172)	0.944 (0.259)	0.822 (0.114)	0.886 (0.116)
400	0.893 (0.065)	0.967 (0.097)	0.882 (0.060)	0.894 (0.061)
1600	0.889 (0.029)	0.913 (0.032)	0.893 (0.031)	0.896 (0.031)
6400	0.893 (0.014)	0.898 (0.015)	0.897 (0.015)	0.897 (0.015)

In general, we see that in both tables, the coverages converge to the nominal level as n gets large, demonstrating the CIs' asymptotic validity. When $p \approx 1$ and n is small, sectioning generally gives better coverage than batching because the former centers its

CI at a less-biased point. Sectioning also outperforms FD in terms of coverage. Comparing FD and Exact for $p = 0.99$ and small n, we see that the AHW of FD typically is quite different from AHW for Exact, which indicates that in these cases, FD does a poor job estimating λ, resulting in FD's poor coverage.

Relative to SRS, CMC reduces the AHW about 60% (resp., 70% and 80%) for $p = 0.8$ (resp., $p = 0.95$ and $p = 0.99$). Thus, the variance reduction from CMC improves as we consider more extreme quantiles. For each of the smaller values of n, the coverage for each SRS CI (except Exact) worsens as p increases. While CMC coverage also degrades somewhat as p approaches 1, the impact is much less pronounced. Also, for large n, the slightly wider AHW for batching and sectioning compared to Exact and FD arises because the former two methods are based on a Student t limit, whereas the latter two rely on a normal limit, which has lighter tails.

5 CONCLUSIONS

We developed an estimator of a quantile ξ using conditional Monte Carlo, which is guaranteed to reduce asymptotic variance compared to simple random sampling. We established that the CMC quantile estimator satisfies a weak Bahadur representation, which implies a CLT holds. We used these results to produce three asymptotically valid confidence intervals for ξ as the sample size $n \to \infty$. The CIs are based on batching, sectioning and a finite difference. Our numerical results seem to indicate that compared to SRS, CMC not only reduces variance, but it also leads to CIs with better coverage. For both SRS and CMC, the sectioning CI has better coverage than the batching and finite-difference intervals for small n, especially when $p \approx 1$. Thus, of the three CIs we proposed, we recommend using sectioning.

ACKNOWLEDGEMENTS

This work has been supported in part by the National Science Foundation under Grants No. CMMI-0926949, CMMI-1200065, and DMS-1331010. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

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

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

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, New York, third edition.

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

Falk, M. (1986). On the estimation of the quantile density function. *Statistics & Probability Letters*, 4:69–73.

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.

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition*. McGraw-Hill.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 3rd edition.

Nakayama, M. K. (2014a). Confidence intervals using sectioning for quantiles when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation*. To appear.

Nakayama, M. K. (2014b). Efficient quantile estimation using conditional Monte Carlo. In preparation.

Ortega, J. M. and Rheinboldt, W. C. (1987). *terative Solution of Nonlinear Equations in Several Variables*. SIAM.

Ross, S. M. (2006). *Simulation*. Academic Press, San Diego, CA, fourth edition.

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

U.S. Nuclear Regulatory Commission (1989). Best-estimate calculations of emergency core cooling performance. Nuclear Regulatory Commission Regulatory Guide 1.157, U.S. Nuclear Regulatory Commission, Washington, DC.

U.S. Nuclear Regulatory Commission (2011). Applying statistics. U.S. Nuclear Regulatory Commission Report NUREG-1475, Rev 1, U.S. Nuclear Regulatory Commission, Washington, DC.