Quantile Estimation Via a Combination of Conditional Monte Carlo and Randomized Quasi-Monte Carlo

Marvin K. Nakayama
Zachary T. Kaplan
Computer Science Department
New Jersey Institute of Technology
Newark, NJ 07102, USA

Bruno Tuffin
Inria-Rennes, Campus de Beaulieu
35042 Rennes, FRANCE

Pierre L'Ecuyer
DIRO, Pavillon Aisenstadt, Université de Montréal
Montréal (Québec), H3C 3J7, CANADA

ABSTRACT

We consider the problem of estimating the p-quantile of a distribution when observations from that distribution are generated from a simulation model. The standard estimator takes the p-quantile of the empirical distribution of independent observations obtained by Monte Carlo. As an improvement, we use conditional Monte Carlo to obtain a smoother estimate of the distribution function, and we combine this with randomized quasi-Monte Carlo to further reduce the variance. The result is a much more accurate quantile estimator, whose mean square error can converge even faster than the canonical rate of $O(1/n)$.

1 INTRODUCTION

The *p-quantile* of a random variable Y with cumulative distribution function (CDF) F, for $0 < p < 1$, is defined as $\xi = \xi_p = F^{-1}(p) = \inf\{y : F(y) \geq p\}$. The *value at risk* (VaR) in finance and actuarial science is defined as a quantile (Asmussen 2018). If Y represents a level of damage or loss, for example, then ξ_p is the level that is exceeded with probability $1-p$. A standard way of estimating ξ_p from a given sample of n independent and identically distributed (i.i.d.) observations of Y is to take the p-quantile of the empirical distribution, whose CDF $\hat{F}_n$ is a step function having a jump of size $1/n$ at each observation. In case these observations are generated from a simulation model by Monte Carlo (MC), one can do better by controlling the way the observations are generated and how the CDF is estimated from them.

In this paper, we do this by combining two techniques. The first one, *conditional Monte Carlo* (CMC), replaces the empirical CDF by an average of n realizations of the CDF of Y conditional on partial information that does not completely reveal Y, but from which we can easily compute its conditional CDF. It is often possible to do that in a way that the resulting (average) conditional CDF is continuous and smooth.

The second technique replaces the independent samples of Y by *randomized quasi-Monte Carlo* (RQMC) samples. More specifically, we assume that to simulate Y, we generate a vector $\mathbf{U}$ of d independent random numbers uniform over $(0,1)$ that drive the simulation, and Y can be computed as a function of these random numbers. Practically all MC simulations work that way, sometimes with a random d, but this is not a problem (L'Ecuyer and Lemieux 2000; L'Ecuyer 2009). RQMC replaces the n independent realizations of $\mathbf{U}$ by a set of n RQMC points in d dimensions. These points are not independent, but each of them has the uniform distribution over $(0,1)^d$, so each RQMC realization of Y has the correct distribution, and the n points together cover the unit cube more evenly (in some sense) than independent random points. If the

CMC is designed so that the CDF estimator becomes a smooth function of $\mathbf{U}$ in the neighborhood of ξ_p, then the combination with RQMC can be much more effective, because the variance of the CDF estimator can then be reduced significantly, and may converge at a faster rate than the canonical one.

Prior work on applying QMC and RQMC for quantile estimation include Papageorgiou and Paskov (1999), Jin and Zhang (2006), Lai and Tan (2006), He and Wang (2017), and Kaplan et al. (2019), although none of these papers combined it with CMC. Using CMC alone (without RQMC) for quantile estimation was proposed and studied in Nakayama (2014) and further discussed in Asmussen (2018). Combining CMC with RQMC for *density* estimation is studied in L'Ecuyer et al. (2020). These authors mention a possible application to quantile estimation, but leave its study for future work.

The paper's remainder is organized as follows. In Section 2, we define the mathematical setting and introduce an example that we carry all along the paper. In the five following sections, we recall the basic definitions and properties of MC, CMC, QMC, RQMC, and the CMC+RQMC combination, and we explain how each of them works to estimate a quantile. Section 8 reports the results of numerical experiments with our running example. A conclusion follows in Section 9.

2 MATHEMATICAL FRAMEWORK

Consider a random variable Y (also called the *response*) defined by

$$Y = b_Y(U_1, U_2, \ldots, U_d) = b_Y(\mathbf{U}) \tag{1}$$

for a given (deterministic) *response function* $b_Y : (0,1)^d \to \mathbb{R}$ with fixed integer $d \geq 1$, where $U_1, U_2, \ldots, U_d$ are i.i.d. $\mathscr{U}(0,1)$ (i.e., uniform on the interval $(0,1)$) random variables, so $\mathbf{U} = (U_1, U_2, \ldots, U_d) \sim \mathscr{U}(0,1)^d$. The function b_Y can be quite complicated, first transforming $\mathbf{U}$ into a random vector $\mathbf{V}$ having a specified joint distribution, and then using $\mathbf{V}$ in detailed computations to finally output a response Y; e.g., see Example 1 below. Let F be the *cumulative distribution function* (CDF) of Y:

$$F(y) = P(Y \leq y) = P(b_Y(\mathbf{U}) \leq y) = \int_{(0,1)^d} \mathbb{I}(b_Y(\mathbf{u}) \leq y)\, \mathrm{d}\mathbf{u} \qquad \text{for all } y \in \mathbb{R}, \tag{2}$$

where $\mathbb{I}$ denotes the indicator function. We assume that we know how to compute Y for any realization of $\mathbf{U}$, but we do not know how to compute F or the quantiles analytically or numerically. Our goal is to estimate the p-quantile $\xi = \xi_p$ for a fixed p. We assume that the density $f(y) = F'(y)$ is strictly positive in a neighborhood of ξ, so that $y = \xi$ is the unique root of $F(y) = p$. To estimate ξ, we will estimate F, replace it by its estimator in the equation $F(y) = p$, and take the root as our quantile estimator.

We illustrate the notation through the following example, motivated by a probabilistic safety assessment (PSA) of a nuclear power plant (NPP); see Section 3.2 of U.S. Nuclear Regulatory Commission (2005).

Example 1 A system experiencing a random load L has a random capacity C to withstand the load, and the system suffers damage when $L > C$. We define the system's *safety margin* as $Y = C - L$, and let F denote its CDF. This situation arises, e.g., in a NPP undergoing a hypothesized accident, where L denotes the peak cladding temperature (PCT), and C is the temperature at which the cladding is damaged; e.g., see Dube et al. (2014). Nuclear engineers model L as a random variable as its value depends on unforeseen aspects of how the event progresses. The capacity C is also considered as random because it involves uncertainties regarding the cladding's material properties.

Nuclear engineers employ detailed computer codes (Hess et al. 2009) to generate an observation of (L,C), which often entails numerically solving systems of differential equations. We express this through two deterministic functions $b_L : (0,1)^d \to \mathbb{R}$ and $b_C : (0,1)^d \to \mathbb{R}$ as

$$(L,C) = (b_L(\mathbf{U}), b_C(\mathbf{U})), \tag{3}$$

where $\mathbf{U} \sim \mathscr{U}(0,1)^d$. Note that we evaluate the functions b_L and b_C using the same uniform vector $\mathbf{U}$, so L and C can be dependent. The response function in (1) is $b_Y(\mathbf{u}) = b_C(\mathbf{u}) - b_L(\mathbf{u})$ for $\mathbf{u} \in (0,1)^d$, so we obtain an observation of the safety margin as $Y = b_Y(\mathbf{U})$.

Let $\theta = P(Y < 0)$, which is the probability of sustaining damage during an event, and suppose that a regulator has specified that θ must not exceed a given threshold p, e.g., $p = 0.05$. We can restate the requirement that $\theta \leq p$ in terms of a quantile: the p-quantile ξ of Y must satisfy $\xi > 0$. A quantile provides information that can be more easily interpretable than the probability θ because ξ is expressed in the same units (temperature) as L and C. Indeed, nuclear risk studies are often performed using quantiles; e.g., see U.S. Nuclear Regulatory Commission (2005). □

3 MONTE CARLO

To estimate ξ via MC, we generate n i.i.d. random vectors $\mathbf{U}_i \sim \mathscr{U}(0,1)^d$, $i = 1,2,\ldots,n$, put $Y_i = b_Y(\mathbf{U}_i)$ for each i, and define the empirical CDF by

$$\widehat{F}_{\mathrm{MC},n}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(Y_i \leq y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(b_Y(\mathbf{U}_i) \leq y), \tag{4}$$

which is an unbiased MC estimator of $F(y)$. The MC estimator of the p-quantile ξ is then

$$\widehat{\xi}_{\mathrm{MC},n} = \widehat{F}^{-1}_{\mathrm{MC},n}(p). \tag{5}$$

It can be easily computed as $\widehat{\xi}_{\mathrm{MC},n} = Y_{\lceil np \rceil:n}$ where $\lceil \cdot \rceil$ denotes the ceiling function and $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$ are the n observations $Y_1, Y_2, \ldots, Y_n$ sorted by increasing order.

Under our assumption that $f(\xi) > 0$, the MC p-quantile estimator obeys a *central limit theorem* (CLT):

$$\sqrt{n}[\widehat{\xi}_{\mathrm{MC},n} - \xi] \Rightarrow N(0, \tau^2_{\mathrm{MC}}) \quad \text{as } n \to \infty, \tag{6}$$

where $\Rightarrow$ denotes convergence in distribution (Chapter 5 of Billingsley 1995),

$$\tau^2_{\mathrm{MC}} = \frac{\mathrm{Var}[\mathbb{I}(Y \leq \xi)]}{f^2(\xi)} = \frac{p(1-p)}{f^2(\xi)} \tag{7}$$

is the CLT's *asymptotic variance*, and $\mathrm{Var}[\cdot]$ is the variance operator; see Section 2.3.3 of Serfling (1980).

We can construct a confidence interval (CI) for ξ as follows. Let $\widehat{\tau}^2_{\mathrm{MC},n}$ be a consistent estimator of τ^2_{MC}, e.g., as in Bloch and Gastwirth (1968); i.e., $\widehat{\tau}^2_{\mathrm{MC},n} \Rightarrow \tau^2_{\mathrm{MC}}$ as $n \to \infty$. Then for a fixed confidence level $0 < \beta < 1$, we define a CI for ξ as $J_{\mathrm{MC},n} = [\widehat{\xi}_{\mathrm{MC},n} \pm z_\beta \widehat{\tau}_{\mathrm{MC},n}/\sqrt{n}]$, where $z_\beta = \Phi^{-1}(1-(1-\beta)/2)$ and Φ is the $N(0,1)$ CDF. The CI is asymptotically valid in the sense that $\lim_{n\to\infty} P(\xi \in J_{\mathrm{MC},n}) = \beta$. There are also other approaches for constructing a CI for ξ via MC; see Nagaraja and Nagaraja (2019).

As shown by Avramidis and Wilson (1998), the *root mean square error* (RMSE) of $\widehat{\xi}_{\mathrm{MC},n}$, defined as $(E[(\widehat{\xi}_{\mathrm{MC},n} - \xi)^2])^{1/2}$, decreases at the canonical MC rate of $O(n^{-1/2})$ as $n \to \infty$. This is a rather slow rate of convergence, as reducing the error by a factor of 10 requires a 100-fold increase in n.

4 CONDITIONAL MONTE CARLO

When estimating a mean, CMC (Section V.4 of Asmussen and Glynn 2007) reduces sampling error by analytically integrating out some of the variability. Nakayama (2014) applies CMC with MC for quantile estimation (also see Asmussen 2018). Dong and Nakayama (2017) combine CMC with Latin hypercube sampling (McKay et al. 1979), which can be viewed as a variant of stratified sampling in high dimensions, or as a primitive form of RQMC (Owen 1997a).

To apply CMC for quantile estimation, we apply CMC to estimate F, then we invert this CMC estimator of F to obtain the CMC quantile estimator. More specifically, let $\mathscr{G}$ be a sigma-field not containing all of the

information needed to compute Y (we discuss specific examples of $\mathscr{G}$ below). Using iterated expectations (p. 448 of Billingsley 1995), we can write

$$F(y) = E[\mathbb{I}(Y \le y)] = E[E[\mathbb{I}(Y \le y) \mid \mathscr{G}]] = E[q(y,\mathscr{G})], \tag{8}$$

where $q(y,\mathscr{G}) := P(Y \le y \mid \mathscr{G})$, which we assume is easily computable. Averaging n i.i.d. realizations of $q(y,\mathscr{G})$, say $q(y,\mathscr{G}_1),\ldots,q(y,\mathscr{G}_n)$, gives the following unbiased CMC estimator of $F(y)$:

$$\widehat{F}_{\mathrm{CMC},n}(y) = \frac{1}{n}\sum_{i=1}^{n} q(y,\mathscr{G}_i). \tag{9}$$

By inverting $\widehat{F}_{\mathrm{CMC},n}$, we get the CMC p-quantile estimator

$$\widehat{\xi}_{\mathrm{CMC},n} = \widehat{F}^{-1}_{\mathrm{CMC},n}(p). \tag{10}$$

Computing $\widehat{\xi}_{\mathrm{CMC},n}$ typically requires employing a numerical root-finding method, such as the bisection or Newton's method (Chapter 7 of Ortega and Rheinboldt 2000), which incurs extra computation cost.

In our setting where $Y = b_Y(\mathbf{U})$ with $\mathbf{U} \sim \mathscr{U}(0,1)^d$, we may for example define $\mathscr{G}$ as (the sigma-field generated by) a random vector

$$\mathbf{Z} \equiv (Z_1, Z_2, \ldots, Z_l) = b_{\mathbf{Z}}(\mathbf{U}) \tag{11}$$

for a deterministic function $b_{\mathbf{Z}} : (0,1)^d \to \mathbb{R}^l$ for some $l \ge 1$, using the *same* $\mathscr{U}(0,1)^d$ vector $\mathbf{U}$ from (1), making Y and $\mathbf{Z}$ generally dependent. Thus, $b_{\mathbf{Z}}(\mathbf{U})$ has all of the details needed to produce one realization of the information contained in the sigma-field $\mathscr{G}$. Although the function $b_{\mathbf{Z}}$ is defined to take as input all d coordinates from $\mathbf{U}$ in (1), it may not use all of them. Writing $\mathbf{Z}_i = b_{\mathbf{Z}}(\mathbf{U}_i)$, (9) becomes

$$\widehat{F}_{\mathrm{CMC},n}(y) = \frac{1}{n}\sum_{i=1}^{n} q(y,\mathbf{Z}_i) = \frac{1}{n}\sum_{i=1}^{n} q(y,b_{\mathbf{Z}}(\mathbf{U}_i)). \tag{12}$$

Example 1 (continued) To apply CMC to our example, we define the conditioning vector in (11) as $\mathbf{Z} = L$, so $b_{\mathbf{Z}}(\mathbf{U}) = b_L(\mathbf{U})$, where b_L is from (3). As in Dube et al. (2014), we now assume that the load L and the capacity C are independent. From a modeling perspective, it is reasonable to have L and C independent as L is determined by random variables that affect how the hypothesized accident unfolds, whereas C depends on uncertainties about the material properties of the cladding. Let G_C be the marginal CDF of C. We then write the CDF F of the safety margin $Y = C - L$ as

$$F(y) = P(C - L \le y) = E[P(C \le y + L \mid L)] = E[G_C(y+L)] \tag{13}$$

by the independence of L and C. Thus, we have that $q(y,\mathbf{Z}) = G_C(y+L) = G_C(y + b_L(\mathbf{U}))$, and (12) gives $\widehat{F}_{\mathrm{CMC},n}(y) = (1/n)\sum_{i=1}^{n} G_C(y + b_L(\mathbf{U}_i))$ as the CMC estimator of the CDF.

The independence of L and C holds, e.g., when the two functions $b_L(\mathbf{u})$ and $b_C(\mathbf{u})$ in (3) depend only on disjoint subsets of the components of $\mathbf{u} = (u_1, u_2, \ldots, u_d) \in (0,1)^d$. Specifically, consider a subset $\mathscr{S} \subset \{1,2,\ldots,d\}$ of the coordinate indices, and let $\bar{\mathscr{S}} = \{1,2,\ldots,d\} - \mathscr{S}$ be its complement. For a vector $\mathbf{u} \in (0,1)^d$, let $\mathbf{u}_{\mathscr{S}} \in (0,1)^{|\mathscr{S}|}$ (resp., $\mathbf{u}_{\bar{\mathscr{S}}} \in (0,1)^{|\bar{\mathscr{S}}|}$) be the projection of $\mathbf{u}$ onto the coordinates in $\mathscr{S}$ (resp., $\bar{\mathscr{S}}$). Now assume that there exist functions $b^*_L : (0,1)^{|\mathscr{S}|} \to \mathbb{R}$ and $b^*_C : (0,1)^{|\bar{\mathscr{S}}|} \to \mathbb{R}$ such that $b_L(\mathbf{u}) = b^*_L(\mathbf{u}_{\mathscr{S}})$ and $b_C(\mathbf{u}) = b^*_C(\mathbf{u}_{\bar{\mathscr{S}}})$ for every $\mathbf{u} \in (0,1)^d$. Then $L = b_L(\mathbf{U}) = b^*_L(\mathbf{U}_{\mathscr{S}})$ and $C = b_C(\mathbf{U}) = b^*_C(\mathbf{U}_{\bar{\mathscr{S}}})$ are independent because $\mathbf{U}_{\mathscr{S}}$ and $\mathbf{U}_{\bar{\mathscr{S}}}$ are as $\mathbf{U} \sim \mathscr{U}(0,1)^d$ has independent components. □

As noted by Nakayama (2014), when $f(\xi) > 0$, the CMC p-quantile estimator obeys the CLT $\sqrt{n}[\widehat{\xi}_{\mathrm{CMC},n} - \xi] \Rightarrow N(0, \tau^2_{\mathrm{CMC}})$ as $n \to \infty$, where

$$\tau^2_{\mathrm{CMC}} = \mathrm{Var}[q(\xi,\mathscr{G})]/f^2(\xi). \tag{14}$$

A variance decomposition (e.g., p. 456 of Billingsley 1995) implies that for each $y \in \mathbb{R}$,

$$\begin{aligned}\mathrm{Var}[\mathbb{I}(Y \le y)] &= \mathrm{Var}[E[\mathbb{I}(Y \le y) \mid \mathscr{G}]] + E[\mathrm{Var}[\mathbb{I}(Y \le y) \mid \mathscr{G}]] \\ &= \mathrm{Var}[q(y,\mathscr{G})] + E[\mathrm{Var}[\mathbb{I}(Y \le y) \mid \mathscr{G}]] \;\ge\; \mathrm{Var}[q(y,\mathscr{G})]. \end{aligned} \tag{15}$$

As a consequence, $\mathrm{Var}[\widehat{F}_{\mathrm{CMC},n}(y)] = \mathrm{Var}[q(y,\mathscr{G})]/n \le \mathrm{Var}[\mathbb{I}(Y \le y)]/n = \mathrm{Var}[\widehat{F}_{\mathrm{MC},n}(y)]$ for each $y \in \mathbb{R}$, and by taking $y = \xi$, this also gives $\tau^2_{\mathrm{CMC}} \le \tau^2_{\mathrm{MC}}$ by (7) and (14). This shows that CMC reduces the asymptotic variance of the p-quantile estimator compared to MC, but the RMSE of $\widehat{\xi}_{\mathrm{CMC},n}$ still decreases at the canonical MC rate of $O(n^{-1/2})$, although with a smaller hidden constant.

5 QUASI-MONTE CARLO

QMC replaces the n independent random points $\mathbf{U}_i$ by a set $\mathscr{P}_n$ of n *deterministic* points $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,d}) \in [0,1)^d$ which cover the unit hypercube more evenly than typical independent random points, in the sense that their empirical distribution has a *low discrepancy* with respect to the uniform distribution over $[0,1)^d$, lower than for typical independent random points. Here we take the interval $[0,1)$ closed on the left because most constructions have points with coordinates equal to 0, but this will disappear when the points are randomized with RQMC. (In practice, we want to avoid 0 and 1 because it causes a problem when generating random variates by inversion from distributions having an infinite tail.) The most common constructions of low-discrepancy point sets are *integration lattices* (Sloan and Joe 1994) and *digital nets* (Niederreiter 1992; Dick and Pillichshammer 2010).

QMC theory was developed mainly for when we want to approximate an integral of the form $\gamma := \int_{[0,1)^d} h(\mathbf{u})\,\mathrm{d}\mathbf{u}$ for some function $h : [0,1)^d \to \mathbb{R}$ by the average $\widehat{\gamma}_{\mathrm{QMC},n} = (1/n)\sum_{i=1}^n h(\mathbf{u}_i)$ over the points $\mathbf{u}_i \in \mathscr{P}_n$. There is a large variety of Cauchy-Schwarz-type inequalities of the form

$$|\widehat{\gamma}_{\mathrm{QMC},n} - \gamma| \le D(\mathscr{P}_n) \cdot V(h), \tag{16}$$

where $D(\mathscr{P}_n)$ measures the discrepancy of $\mathscr{P}_n$ and $V(h)$ measures the variation of h. One special case is the classical *Koksma-Hlawka inequality* in which $D(\mathscr{P}_n)$ is the star discrepancy $D^*(\mathscr{P}_n)$ and V is the Hardy-Krause variation V_{HK} (Niederreiter 1992, Section 2.2), but it is impractical because these quantities are too hard to compute. However, other more easily computable discrepancies, together with matching definitions of V, also exist (Dick and Pillichshammer 2010; L'Ecuyer 2009; L'Ecuyer and Munger 2016). Explicit point set constructions can achieve $D(\mathscr{P}_n) = O(n^{-\alpha+\varepsilon})$ for any $\varepsilon > 0$, often for $\alpha = 1$ (e.g., for the star discrepancy) and for $\alpha > 1$ for some discrepancies. Then, for any h for which $V(h) < \infty$ for the corresponding V, the integration error converges to 0 at the same rate (at worst) as $D(\mathscr{P}_n)$. It is true that the hidden constant in the convergence order can increase very quickly with the dimension d, but there are nevertheless many large-dimensional integrands for which QMC is much more accurate than MC. Typically, this occurs when h can be decomposed approximately as a sum of low-dimensional functions; see Paskov and Traub (1995), L'Ecuyer (2009), Dick and Pillichshammer (2010) and L'Ecuyer and Munger (2016).

Papageorgiou and Paskov (1999) approximate F and ξ as in (4) and (5), with the $\mathbf{U}_i$'s replaced by the QMC points $\mathbf{u}_i$. This gives

$$\widehat{F}_{\mathrm{QMC},n}(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(b_Y(\mathbf{u}_i) \le y) \tag{17}$$

and $\widehat{\xi}_{\mathrm{QMC},n} = \widehat{F}^{-1}_{\mathrm{QMC},n}(p)$. One could think of applying QMC theory to show that $\widehat{F}_{\mathrm{QMC},n}(y)$ converges to $F(y)$ faster than $\widehat{F}_{\mathrm{MC},n}(y)$. But this CDF estimator corresponds to applying QMC to the integrand $h(\mathbf{u}) = h_y(\mathbf{u}) := \mathbb{I}(b_Y(\mathbf{u}) \le y)$, which is a discontinuous function of $\mathbf{u}$. This discontinuity implies that the variation $V(h_y)$ is typically infinite, so there is no error bound, and the gain from QMC is usually very small in that situation; see, e.g., Sections 15.12 and 16.5 of Owen (2019) and L'Ecuyer (2018).

We saw earlier that CMC reduces the variance of a CDF estimator, but much more importantly, it also provides a very powerful opportunity to make $q(y, b_{\mathbf{Z}}(\mathbf{u}))$ continuous and smooth in $\mathbf{u}$. Then, its combination with QMC could give a much better approximation of F than the QMC approximation in (17), and also a much better quantile estimator. Assuming that the conditioning sigma-field $\mathscr{G}$ in (8) is defined by $\mathbf{Z} = b_{\mathbf{Z}}(\mathbf{U})$ in (11), the idea is simply to replace each $\mathbf{U}_i$ in (12) with $\mathbf{u}_i \in \mathscr{P}_n$ to get a CDF approximation $\widehat{F}_{\mathrm{CQ},n}(y)$, whose inversion gives an approximation $\widehat{\xi}_{\mathrm{CQ},n} = \widehat{F}^{-1}_{\mathrm{CQ},n}(p)$ of the p-quantile. (Here, the subscript "CQ" is an abbreviation for the combination CMC+QMC.)

CMC has already been used in the setting of estimating the derivative of an expectation or of a quantile with respect to some model parameter (L'Ecuyer and Perron 1994; Fu and Hu 1997; Fu et al. 2009). It was also used in combination with QMC or RQMC to estimate a mean (L'Ecuyer and Lemieux 2000 and He 2019). L'Ecuyer et al. (2020) obtain large gains by combining CMC with RQMC for density estimation.

A major limitation of QMC, however, is that it does not provide an easily computable bound or estimate for the error on the CDF and for the errors $|\widehat{\xi}_{\mathrm{QMC},n} - \xi|$ and $|\widehat{\xi}_{\mathrm{CQ},n} - \xi|$ on the quantile approximations. This motivates RQMC, described next.

6 RANDOMIZED QUASI-MONTE CARLO

RQMC turns QMC into a variance reduction method by randomizing the points in a way that they retain their low discrepancy as a group, while each individual point has the uniform distribution over the unit hypercube (Owen 1995; Owen 1997b; L'Ecuyer and Lemieux 2000; L'Ecuyer and Lemieux 2002; Lemieux 2009; L'Ecuyer 2018). By doing $r \geq 2$ independent randomizations, one can obtain an unbiased estimator for the variance of the CDF estimator by using the sample variance across randomizations, and eventually an estimator for the variance of the quantile estimator and a confidence interval for the quantile. But as noted by Kaplan et al. (2019), one has to be careful in how to adapt the RQMC approach for estimating a mean to estimating the quantile ξ instead, because the quantile estimator is generally biased.

We now provide more details on RQMC. To have a fair comparison with the MC and CMC estimators in (5) and (10), which are based on n evaluations of an integrand, we want to implement RQMC with the same total number of integrand evaluations. Thus, we start with a low-discrepancy point set of size $m < n$, which we randomize $r = n/m$ times, where we assume that $r \geq 2$ is integer-valued. Specifically, let $\mathscr{P}_m = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\} \subset [0,1)^d$ be a low-discrepancy point set of size m. We randomize this point set $r = n/m$ independent times to obtain r independent randomized point sets $\mathscr{P}_m^{(k)}$, $k = 1, 2, \ldots, r$, with $\mathscr{P}_m^{(k)} = \{\mathbf{X}_1^{(k)}, \mathbf{X}_2^{(k)}, \ldots, \mathbf{X}_m^{(k)}\}$ and each $\mathbf{X}_i^{(k)} = (X_{i,1}^{(k)}, X_{i,2}^{(k)}, \ldots X_{i,d}^{(k)})$ has the uniform distribution over $(0,1)^d$.

Appropriate randomizations of QMC point sets, which preserve the low-discrepancy properties, depend on the type of point set construction. For integration lattices, a random shift modulo 1 is the most common and appropriate approach (Cranley and Patterson 1976; Tuffin 1996; L'Ecuyer and Lemieux 2000). It works as follows. For the kth randomization, we generate a single random vector $\mathbf{S}_k = (S_{k,1}, S_{k,2}, \ldots, S_{k,d}) \sim \mathscr{U}(0,1)^d$ and we add it to all the points $\mathbf{u}_i$ of the original set $\mathscr{P}_m$, modulo 1, component-wise. This gives a randomized point set $\mathscr{P}_m^{(k)} = \{\mathbf{X}_1^{(k)}, \mathbf{X}_2^{(k)}, \ldots, \mathbf{X}_m^{(k)}\}$, where each $\mathbf{X}_i^{(k)} = ((u_{i,1} + S_{k,1}) \bmod 1, (u_{i,2} + S_{k,2}) \bmod 1, \ldots, (u_{i,d} + S_{k,d}) \bmod 1)$. Clearly, each $\mathbf{X}_i^{(k)}$ has the uniform distribution over $(0,1)^d$, and the point set keeps its lattice structure; it becomes a shifted lattice. We get independent randomizations simply by taking r independent random shifts $\mathbf{S}_k$, $k = 1, 2, \ldots, r$. When estimating a mean, the resulting estimator is called a *randomly-shifted lattice rule*, and its variance properties have been studied extensively (Cranley and Patterson 1976; Tuffin 1996; L'Ecuyer and Lemieux 2000; L'Ecuyer and Munger 2016).

For digital nets, the uniformity is usually measured in terms of equidistribution of the points for certain families of rectangular boxes, and a random shift does not preserve these properties. Other types of randomizations such as *random digital shifts* (Lemieux and L'Ecuyer 2001; Dick and Pillichshammer 2010; L'Ecuyer 2018), and various scrambles (Owen 1995; Owen 1997b; Owen 2019), do preserve them.

To construct an asymptotically valid CI based on RQMC, it is desirable for the RQMC estimator to obey a CLT with a Gaussian limit. When estimating a mean, Loh (2003) proves such a CLT as the size

m of the point set grows large for RQMC using a digital net with full nested scrambling (Owen 1995). However, this scrambling is computationally more expensive than random shifts (digital or modulo 1).

For mean estimators based on randomly-shifted lattice rules, L'Ecuyer et al. (2010) determine that the limiting distribution (as $m \to \infty$) is generally not Gaussian, with histograms from numerical studies exhibiting asymmetry and multiple modes. Also, to specify the number r of randomizations, practitioners often choose r as not too large, e.g., $r = 10$ or 20 at most, to be able take a larger m to exploit the power of RQMC. But this can result in a poor variance estimator and a CI with poor coverage, especially when considering biased estimators (as for a quantile) or when studying tail behavior (e.g., a p-quantile with $p \approx 0$ or $p \approx 1$). Thus, as the computation budget increases, to obtain an RQMC CI that is asymptotically valid, it can be useful to have the number r of randomizations increase to infinity to ensure a CLT with a Gaussian limit and to allow for consistent estimation of the CLT's asymptotic variance.

We now describe one possible approach, which we call RQMC1, for quantile estimation based on the typical RQMC method for estimating a mean, as described at the beginning of this section. For each randomization $k = 1, 2, \ldots, r$, compute

$$\widehat{F}_{\mathrm{RQMC1},m,k}(y) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{I}(b_Y(\mathbf{X}_i^{(k)}) \le y) \tag{18}$$

as an unbiased CDF estimator from the randomized point set $\mathscr{P}_m^{(k)}$. Inverting $\widehat{F}_{\mathrm{RQMC1},m,k}$ yields $\widehat{\xi}_{\mathrm{RQMC1},m,k} = \widehat{F}^{-1}_{\mathrm{RQMC1},m,k}(p)$, and the $\widehat{\xi}_{\mathrm{RQMC1},m,k}$, $k = 1, 2, \ldots, r$, are i.i.d. Then the *RQMC1 quantile estimator* is

$$\overline{\xi}_{\mathrm{RQMC1},m,r} = \frac{1}{r}\sum_{k=1}^{r} \widehat{\xi}_{\mathrm{RQMC1},m,k}. \tag{19}$$

Even if (18) is unbiased for $F(y)$, the nonlinearity of the inverse operation leads to $\widehat{\xi}_{\mathrm{RQMC1},m,k}$ having bias in general, so under the assumption that $E[|\widehat{\xi}_{\mathrm{RQMC1},m,k}|] < \infty$, the law of large numbers implies that

$$\overline{\xi}_{\mathrm{RQMC1},m,r} \overset{\text{w.p.1}}{\to} E[\widehat{\xi}_{\mathrm{RQMC1},m,k}] \neq \xi, \quad \text{as } r \to \infty \text{ with } m \text{ fixed;} \tag{20}$$

i.e., $\overline{\xi}_{\mathrm{RQMC1},m,r}$ *converges to the wrong value*. While the bias of $\overline{\xi}_{\mathrm{RQMC1},m,r}$ decreases to zero as m grows, (20) considers the asymptotic regime in which r increases with m fixed, causing the problems. If $\tau'^2_{\mathrm{RQMC1},m} \equiv \mathrm{Var}[\widehat{\xi}_{\mathrm{RQMC1},m,k}] < \infty$, then $\overline{\xi}_{\mathrm{RQMC1},m,r}$ obeys a CLT $\sqrt{r}(\overline{\xi}_{\mathrm{RQMC1},m,r} - E[\widehat{\xi}_{\mathrm{RQMC1},m,k}]) \Rightarrow \mathcal{N}(0, \tau'^2_{\mathrm{RQMC1},m})$ as $r \to \infty$ with m fixed, where $\tau'^2_{\mathrm{RQMC1},m}$ depends on m. However, the CLT uses $E[\widehat{\xi}_{\mathrm{RQMC1},m,k}]$ as the centering constant rather than the true quantile ξ, so a CI based on the CLT will have poor coverage, as the CI is anchored at $\overline{\xi}_{\mathrm{RQMC1},m,r}$, which on average does not equal the true value ξ.

To address this, Kaplan et al. (2019) propose a second RQMC quantile estimator that *does converge* to ξ in the asymptotic setting of (20). Rather than use each randomization to compute a quantile estimator, as in (19), we instead employ all r randomizations to compute a *single overall* CDF estimator, which is inverted to obtain a single overall quantile estimator. Specifically, first define the CDF estimator

$$\widetilde{F}_{\mathrm{RQMC2},m,r}(y) = \frac{1}{r}\sum_{k=1}^{r} \widehat{F}_{\mathrm{RQMC1},m,k}(y) = \frac{1}{rm}\sum_{k=1}^{r}\sum_{i=1}^{m} \mathbb{I}(b_Y(\mathbf{X}_i^{(k)}) \le y)$$

based on all rm evaluations of b_Y. Inverting this leads to the *RQMC2 quantile estimator*

$$\widetilde{\xi}_{\mathrm{RQMC2},m,r} = \widetilde{F}^{-1}_{\mathrm{RQMC2},m,r}(p). \tag{21}$$

As with the RQMC1 quantile estimator, the RQMC2 estimator $\widetilde{\xi}_{\mathrm{RQMC2},m,r}$ is also biased. But unlike the bias of $\overline{\xi}_{\mathrm{RQMC1},m,r}$, which does not shrink as $r \to \infty$ with m fixed, leading to the convergence to the

wrong value in (20), the bias of $\widetilde{\xi}_{\mathrm{RQMC2},m,r}$ shrinks to 0. Indeed, it does so fast enough to ensure that under our assumption that $f(\xi) > 0$, the RQMC2 quantile estimator obeys a CLT

$$\sqrt{r}[\widetilde{\xi}_{\mathrm{RQMC2},m,r} - \xi] \Rightarrow \mathcal{N}(0, \tau^2_{\mathrm{RQMC2},m}), \quad \text{as } r \to \infty \text{ with } m \text{ fixed}, \tag{22}$$

as noted in Kaplan et al. (2019), and the asymptotic variance is $\tau^2_{\mathrm{RQMC2},m} = \mathrm{Var}[\widehat{F}_{\mathrm{RQMC1},m,k}(\xi)]/f^2(\xi)$, which depends on m. Thus, the CLT (22) has the desired centering constant ξ, in contrast to the CLT for the RQMC1 estimator $\overline{\xi}_{\mathrm{RQMC1},m,r}$.

7 COMBINING CMC WITH RQMC FOR QUANTILE ESTIMATION

We can combine CMC with RQMC for quantile estimation as follows. As in Section 5, we assume that $\mathscr{G}$ in (8) is defined by $\mathbf{Z} = b_{\mathbf{Z}}(\mathbf{U})$ in (11). Then just as we went from the QMC CDF estimator in (17) to the CQ CDF estimator by substituting $\mathbb{I}(b_Y(\mathbf{u}_i) \le y)$ with $q(y, b_{\mathbf{Z}}(\mathbf{u}_i))$, we similarly replace $\mathbb{I}(b_Y(\mathbf{X}_i^{(k)}) \le y)$ in (18) for the RQMC1 CDF estimator with $q(y, b_{\mathbf{Z}}(\mathbf{X}_i^{(k)}))$ to obtain the combined CMC+RQMC1 (abbreviated CR1) estimator of the CDF from randomization k as

$$\widehat{F}_{\mathrm{CR1},m,k}(y) = \frac{1}{m}\sum_{i=1}^{m} q(y, b_{\mathbf{Z}}(\mathbf{X}_i^{(k)})).$$

Inverting $\widehat{F}_{\mathrm{CR1},m,k}$ yields $\widehat{\xi}_{\mathrm{CR1},m,k} = \widehat{F}^{-1}_{\mathrm{CR1},m,k}(p)$, where $\widehat{\xi}_{\mathrm{CR1},m,k}$, $k = 1, 2, \ldots, r$, are i.i.d. Then our *CR1 quantile estimator* is

$$\overline{\xi}_{\mathrm{CR1},m,r} = \frac{1}{r}\sum_{k=1}^{r} \widehat{\xi}_{\mathrm{CR1},m,k}. \tag{23}$$

But the CR1 quantile estimator also suffers from the problems in (20) that the RQMC1 quantile estimator has caused by bias (but perhaps less so, as we will see in the numerical results of Section 8).

For the combined CMC+RQMC2 (shortened to CR2) quantile estimator, we first compute the CR2 CDF estimator by

$$\widetilde{F}_{\mathrm{CR2},m,r}(y) = \frac{1}{r}\sum_{k=1}^{r} \widehat{F}_{\mathrm{CR1},m,k}(y) = \frac{1}{rm}\sum_{k=1}^{r}\sum_{i=1}^{m} q(y, b_{\mathbf{Z}}(\mathbf{X}_i^{(k)}))$$

based on all rm evaluations of function q. Inverting this leads to our *CR2 quantile estimator*

$$\widetilde{\xi}_{\mathrm{CR2},m,r} = \widetilde{F}^{-1}_{\mathrm{CR2},m,r}(p). \tag{24}$$

Using the framework in Chu and Nakayama (2012), we can prove that under our assumption that $f(\xi) > 0$, the CR2 quantile estimator obeys a CLT

$$\sqrt{r}[\widetilde{\xi}_{\mathrm{CR2},m,r} - \xi] \Rightarrow \mathcal{N}(0, \tau^2_{\mathrm{CR2},m}), \quad \text{as } r \to \infty \text{ with } m \text{ fixed}, \tag{25}$$

with asymptotic variance $\tau^2_{\mathrm{CR2},m} = \mathrm{Var}[\widehat{F}_{\mathrm{CR1},m,k}(\xi)]/f^2(\xi)$, which depends on m.

8 NUMERICAL RESULTS

We now present numerical results from a stylized version of the model in Example 1, where the goal is to estimate the 0.05-quantile ξ of the system's safety margin $Y = C - L$. We first describe the model, which is also considered in Nakayama (2015), Alban et al. (2017), and Kaplan et al. (2019). As in some actual NPP PSA studies (e.g., Dube et al. 2014), the CDF G_C of the capacity C is assumed triangular with support $[1800, 2600]$ and mode 2200, with L and C independent. The marginal CDF G_L of the

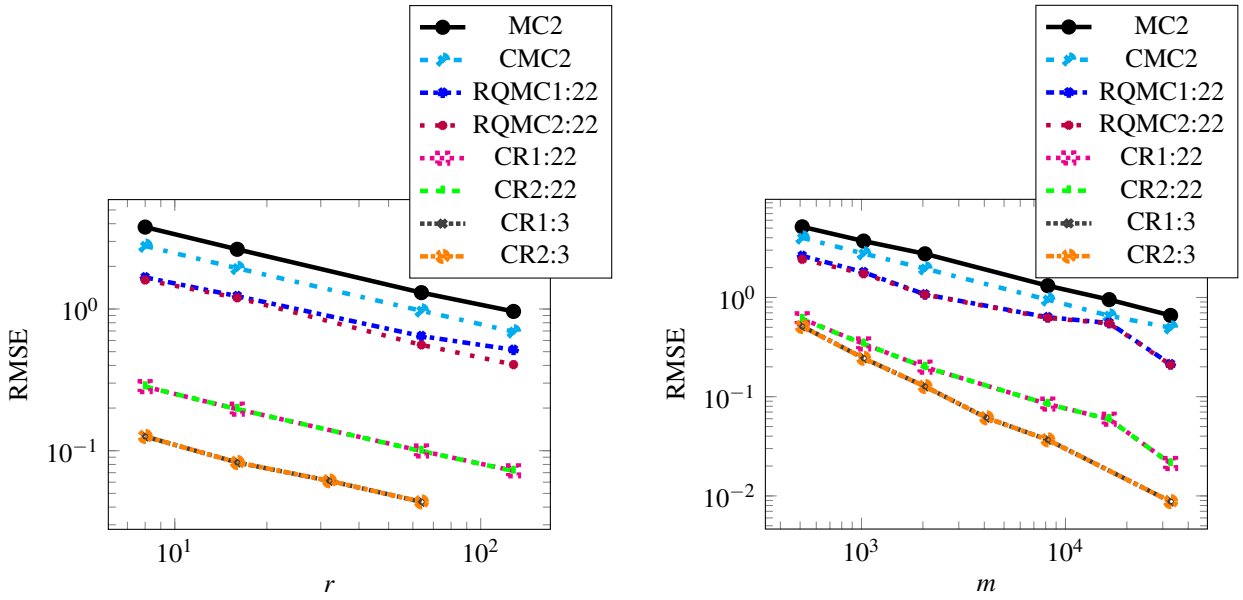


Figure 1: The left panel shows the RMSE for fixed $m = 4096$ as r increases, and the right panel displays the RMSE for fixed $r = 32$ as m increases. Both plots have log-log scale. The notation "$:d$" in the legend specifies the problem dimension without applying CMC. For CRv:d for $d = 3$ and 22, the plots for $v = 1$ and 2 lie on top of each other.

load L is a mixture of $t = 4$ lognormals, as in Nakayama (2015). Specifically, for each $s = 1, 2, \ldots, t$, let $G_{L,\langle s\rangle}$ be the CDF of $L_{\langle s\rangle} = \exp(\mu_{\langle s\rangle} + \sigma_{\langle s\rangle} Z_{\langle s\rangle})$, where $Z_{\langle s\rangle} \sim \mathcal{N}(0,1)$, and $\mu_{\langle s\rangle}$ and $\sigma_{\langle s\rangle} > 0$ are given constants, so $L_{\langle s\rangle}$ has a lognormal distribution. Our experiments set $\mu_{\langle s\rangle} = 7.4 + 0.1s$ and $\sigma_{\langle s\rangle} = 0.01 + 0.01s$. Then G_L is defined as $G_L(y) = \sum_{s=1}^{t} \lambda_{\langle s\rangle} G_{L,\langle s\rangle}(y)$ for given positive constants $\lambda_{\langle s\rangle}$, $1 \le s \le t$, summing to 1. We set $\lambda_{\langle 1\rangle} = 0.99938 \times 0.9981 \times 0.919$, $\lambda_{\langle 2\rangle} = 0.00062$, $\lambda_{\langle 3\rangle} = 0.99938 \times 0.9981 \times 0.081$, and $\lambda_{\langle 4\rangle} = 0.99938 \times 0.0019$, where the factors in each product match branching probabilities given in an event tree used in an NPP PSA study by Dube et al. (2014).

For this model, we can define the response function b_Y in (1) to generate the safety margin $Y = b_Y(U_1, U_2, U_3)$ using just $d = 3$ i.i.d. uniform inputs: one to choose the component in the load mixture G_L, another to generate the appropriate lognormal, and the third to sample the capacity C. For CMC, we condition on L, as in (13), so (12) uses $q(y, \mathbf{Z}) = G_C(y + L) = G_C(y + b_L(\mathbf{U}))$, as in Example 1 of Section 4. Thus, CMC requires only 2 uniforms, as the one for sampling C is integrated out by G_C. The analytical tractability of the model allows us to numerically compute the 0.05-quantile as $\xi = 11.79948572$.

Given that the effectiveness of RQMC typically degrades as the problem dimension d increases (Section 15.9 of Owen 2019), we also consider for comparison a stochastically equivalent artificial model with larger d, defined as follows. We generate the lognormal $L_{\langle s\rangle}$ by exponentiating the sum of 20 independent normals, with parameters chosen so that the sum still has mean $\mu_{\langle s\rangle}$ and variance $\sigma^2_{\langle s\rangle}$. Thus, b_Y in (1) now has $d = 22$.

We consider quantile estimation using two versions ($v = 1$ and 2) each of RQMC and CMC+RQMC. RQMCv denotes version v of RQMC, so the RQMC1 estimator is $\overline{\xi}_{\mathrm{RQMC1},m,r}$ in (19), and the RQMC2 estimator is $\widetilde{\xi}_{\mathrm{RQMC2},m,r}$ in (21). We have two versions of the CMC+RQMC quantile estimator: the CR1 estimator $\overline{\xi}_{\mathrm{CR1},m,r}$ in (23), and the CR2 estimator $\widetilde{\xi}_{\mathrm{CR2},m,r}$ in (24). In each case, our RQMC point set is a randomly-shifted lattice rule of size m, and we make r independent random shifts. For MC and CMC, we consider only a single version $v = 2$, in which we compute a single CDF estimator based on all $n = rm$ outputs, and invert this to get the MC2 and CMC2 quantile estimators.

Figure 1 presents two sets of log-log plots of the RMSE, estimated from 10^3 independent replications, of the various quantile estimators. In the left plots, the number r of randomizations increases, with a point set of fixed size $m = 4096$. On the right side, m grows with fixed $r = 32$. The figures show the RQMC results for different problem dimensions, where d ($= 3$ or $= 22$) in the notation "$:d$" represents the problem dimension without applying CMC. For MC and CMC, changing d does not affect the RMSE, so Figure 1 presents results for only $d = 22$.

We now compare the mean-square error (MSE) of each $v = 2$ method for $r = 64$ and $m = 4096$. With MC2 as the baseline, CMC2 (resp., RQMC2:22 and CR2:22) reduces MSE by a factor of 1.8 (resp., 5.5 and 172). Thus, while RQMC2 improves on MC2 and CMC2, the combination of CR2 performs substantially better, illustrating the benefits of a smoother integrand for RQMC. Moreover, RQMC2:3 (not shown in Figure 1) and CR2:3 reduce MSE (compared to MC2) by factors of 21.5 and 897, respectively, thus demonstrating the impact of the problem dimension d on RQMC's effectiveness.

As previously explained in Kaplan et al. (2019), RQMC1 has issues as r increases with m fixed, as seen in (20), which we next explain. To understand the left plots in Figure 1, recall that the MSE decomposes as the sum of bias squared and variance. The variance of the RQMC1 quantile estimator shrinks to 0 as $r \to \infty$ with fixed m. But because quantile estimators are generally biased, the RQMC1 bias does not decrease because m is fixed. Thus, the RMSE of RQMC1 will converge to a *strictly positive value* as $r \to \infty$ with m fixed. We can see this start to happen in the left plots in Figure 1, where the RMSE of the RQMC1 estimator is leveling off as r increases. (The other $v = 1$ estimators also eventually suffer from the same problem, although it may not be clear for the range of r considered.) In contrast, the RQMC2 quantile estimator obeys a CLT (as $r \to \infty$ with m fixed) with centering constant ξ, so its RMSE shrinks to 0. The left plots of Figure 1 demonstrate the steady decrease in RMSE of the $v = 2$ estimators.

The right plots of Figure 1 show that as m increases, the rate (i.e., slope) at which the RMSE decreases for RQMC (especially for CRv:3) can be better than for MCv. This demonstrates that RQMC can not only reduce variance but also improve convergence rates.

9 CONCLUSION

We have shown how combining CMC with RQMC can have a synergistic effect in improving the accuracy of quantile estimators in a setting where observations are obtained from simulation. The synergy comes from CMC replacing the empirical CDF (a step function) by a smooth, RQMC-friendly, CDF estimator. This can make the subsequent RQMC improvement more substantial than when RQMC is applied alone. As $f(\xi)$ appears in the denominator of the asymptotic variance in (25), constructing a confidence interval for ξ may also benefit from applying CMC+RQMC for density estimation. In follow-up work, we intend to provide sufficient conditions under which we can prove that the MSE of the quantile estimator converges at a faster rate than the canonical $O(1/n)$. We also want to experiment with a larger variety of examples.

ACKNOWLEDGMENTS

This work has been supported in part by the National Science Foundation under Grant No. CMMI-1537322 to the first author. The work of P. L'Ecuyer was supported by a Discovery Grant RGPIN-2018-05795 from NSERC-Canada and a Grant from the IVADO Fundamental Research Program.

REFERENCES

Alban, A., H. Darji, A. Imamura, and M. K. Nakayama. 2017, September. "Efficient Monte Carlo methods for estimating failure probabilities". *Reliability Engineering and System Safety* 165:376–394.

Asmussen, S. 2018, September. "Conditional Monte Carlo for Sums, with Applications to insurance and Finance". *Annals of Actuarial Science* 12(2):455–478.

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

Avramidis, A. N., and J. R. Wilson. 1998. "Correlation-Induction Techniques for Estimating Quantiles in Simulation". *Operations Research* 46:574–591.

Billingsley, P. 1995. *Probability and Measure*. 3rd ed. New York: John Wiley and Sons.

Bloch, D. A., and J. L. Gastwirth. 1968. "On a Simple Estimate of the Reciprocal of the Density Function". *Annals of Mathematical Statistics* 39:1083–1085.

Chu, F., and M. K. Nakayama. 2012. "Confidence Intervals for Quantiles When Applying Variance-Reduction Techniques". *ACM Transactions On Modeling and Computer Simulation* 22(2):10:1–10:25.

Cranley, R., and T. N. L. Patterson. 1976. "Randomization of Number Theoretic Methods for Multiple Integration". *SIAM Journal on Numerical Analysis* 13(6):904–914.

Dick, J., and F. Pillichshammer. 2010. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge, U.K.: Cambridge University Press.

Dong, H., and M. K. Nakayama. 2017. "Quantile Estimation Using Conditional Monte Carlo and Latin Hypercube Sampling". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A.D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 1986–1997. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

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

Fu, M. C., L. J. Hong, and J.-Q. Hu. 2009. "Conditional Monte Carlo Estimation of Quantile Sensitivities". *Management Science* 55:2019–2027.

Fu, M. C., and J. Q. Hu. 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. New York: Kluwer Academic.

He, Z. 2019. "On the Error Rate of Conditional Quasi-Monte Carlo for Discontinuous Functions". *SIAM Journal on Numerical Analysis* 57(2):854–874.

He, Z., and X. Wang. 2017. "Convergence of Randomized Quasi-Monte Carlo Sampling for Value-at-Risk and Conditional Value-at-Risk". arXiv:1706.00540.

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

Jin, X., and A. X. Zhang. 2006. "Reclaiming Quasi-Monte Carlo Efficiency in Portfolio Value-at-Risk Simulation Through Fourier Transform". *Management Science* 52(6):925–938.

Kaplan, Z. T., Y. Li, M. K. Nakayama, and B. Tuffin. 2019. "Randomized Quasi-Monte Carlo for Quantile Estimation". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 428–439. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Lai, Y., and K. Tan. 2006. "Simulation of Nonlinear Portfolio Value-at-Risk by Monte Carlo and Quasi-Monte Carlo Methods". In *Financial Engineering and Applications*, edited by M. Holder: ACTA Academic Press.

L'Ecuyer, P. 2009. "Quasi-Monte Carlo Methods with Applications in Finance". *Finance and Stochastics* 13(3):307–349.

L'Ecuyer, P. 2018. "Randomized Quasi-Monte Carlo: An Introduction for Practitioners". In *Monte Carlo and Quasi Monte Carlo Methods: MCQMC 2016*, edited by P. W. Glynn and A. B. Owen, 29–52. Berlin: Springer-Verlag.

L'Ecuyer, P., and C. Lemieux. 2000. "Variance Reduction via Lattice Rules". *Management Science* 46(9):1214–1235.

L'Ecuyer, P., and C. Lemieux. 2002. "Recent Advances in Randomized Quasi-Monte Carlo Methods". In *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, edited by M. Dror, P. L'Ecuyer, and F. Szidarovszki, 419–474. Netherlands: Kluwer Academic Publishers.

L'Ecuyer, P., and D. Munger. 2016. "Algorithm 958: Lattice Builder: A General Software Tool for Constructing Rank-1 Lattice Rules". *ACM Transactions on Mathematical Software* 42(2):Article 15.

L'Ecuyer, P., D. Munger, and B. Tuffin. 2010. "On the Distribution of Integration Error by Randomly-Shifted Lattice Rules". *Electronic Journal of Statistics* 4:950–993.

L'Ecuyer, P., and G. Perron. 1994. "On the Convergence Rates of IPA and FDC Derivative Estimators for Finite-Horizon Stochastic Simulations". *Operations Research* 42(4):643–656.

L'Ecuyer, P., F. Puchhammer, and A. Ben Abdellah. 2020. "Monte Carlo and Quasi-Monte Carlo Density Estimation via Conditioning". Preprint, http://arxiv.org/abs/1906.04607.

Lemieux, C. 2009. *Monte Carlo and Quasi-Monte Carlo Sampling*. Series in Statistics. New York: Springer.

Lemieux, C., and P. L'Ecuyer. 2001. "Selection Criteria for Lattice Rules and Other Low-Discrepancy Point Sets". *Mathematics and Computers in Simulation* 55(1–3):139–148.

Loh, W.-L. 2003. "On the Asymptotic Distribution of Scrambled Net Quadrature". *Annals of Statistics* 31(4):1282–1324.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. "A Comparison of Three Methods for Selecting Input Variables in the Analysis of Output from a Computer Code". *Technometrics* 21:239–245.

Nagaraja, C. H., and H. N. Nagaraja. 2019. "Distribution-free Approximate Methods for Constructing Confidence Intervals for Quantiles". *International Statistical Review*. To appear.

Nakayama, M. K. 2014. "Quantile Estimation When Applying Conditional Monte Carlo". In *SIMULTECH 2014 Proceedings*, 280–285.

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

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*, Volume 63. Philadelphia: SIAM.

Ortega, J. M., and W. C. Rheinboldt. 2000. *terative Solution of Nonlinear Equations in Several Variables*. Philadelphia: SIAM.

Owen, A. B. 1995. “Randomly Permuted (t,m,s)-nets and (t,s)-sequences”. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing: Lecture Notes in Statistics*, Volume 106, 299–317. Springer.

Owen, A. B. 1997a. “Monte Carlo variance of scrambled net quadrature”. *SIAM Journal of Numerical Analysis* 34:1884–1910.

Owen, A. B. 1997b. “Scrambled net variance for integrals of smooth functions”. *Annals of Statistics* 25(4):1541–1562.

Owen, A. B. 2019. *Monte Carlo Theory, Methods and Examples*. Draft. In preparation.

Papageorgiou, A., and S. H. Paskov. 1999. “Deterministic simulation for risk management”. *Journal of Portfolio Management* 25(5):122–127.

Paskov, S., and J. Traub. 1995. “Faster Valuation of Financial Derivatives”. *The Journal of Portfolio Management* 22:113–120.

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

Sloan, I. H., and S. Joe. 1994. *Lattice Methods for Multiple Integration*. Oxford, UK: Carendon Press.

Tuffin, B. 1996. “On the Use of Low Discrepancy Sequences in Monte Carlo Methods”. *Monte Carlo Methods and Applications* 2(4):295–320.

U.S. Nuclear Regulatory Commission 2005. “Final Safety Evaluation For WCAP-16009-P, Revision 0, “Realistic Large Break LOCA Evaluation Methodology Using Automated Statistical Treatment Of Uncertainty Method (ASTRUM)” (TAC No. MB9483)”. Technical report, Washington, DC. https://www.nrc.gov/docs/ML0509/ML050910159.pdf.