Empirical Bernstein and betting confidence intervals for randomized quasi-Monte Carlo

Aadit Jain
Rancho Bernardo High School

Fred J. Hickernell
Illinois Institute of Technology

Art B. Owen
Stanford University

Aleksei G. Sorokin
Illinois Institute of Technology

April 2025

Abstract

Randomized quasi-Monte Carlo (RQMC) methods estimate the mean of a random variable by sampling an integrand at n equidistributed points. For scrambled digital nets, the resulting variance is typically $\tilde{O}(n^{-\theta})$ where $\theta \in [1, 3]$ depends on the smoothness of the integrand and $\tilde{O}$ neglects logarithmic factors. While RQMC can be far more accurate than plain Monte Carlo (MC) it remains difficult to get confidence intervals on RQMC estimates. We investigate some empirical Bernstein confidence intervals (EBCI) and hedged betting confidence intervals (HBCI), both from Waudby-Smith and Ramdas (2024), when the random variable of interest is subject to known bounds. When there are N integrand evaluations partitioned into R independent replicates of $n = N/R$ RQMC points, and the RQMC variance is $\Theta(n^{-\theta})$, then an oracle minimizing the width of a Bennett confidence interval would choose $n = \Theta(N^{1/(\theta+1)})$. The resulting intervals have a width that is $\Theta(N^{-\theta/(\theta+1)})$. Our empirical investigations had optimal values of n grow slowly with N, HBCI intervals that were usually narrower than the EBCI ones, and optimal values of n for HBCI that were equal to or smaller than the ones for the oracle.

1 Introduction

We study the combination of randomized quasi-Monte Carlo (RQMC) integration to estimate an expectation with some non-asymptotic (finite sample valid) methods to get a confidence interval for such an expectation. RQMC has significant accuracy benefits over plain Monte Carlo (that we outline below), but the usual confidence intervals based on RQMC estimates have justifications that are asymptotic in their sample size [27]. Empirical Bernstein and some related betting methods (described below) give finite sample assurances for bounded integrands. Specifically, if it is known that the integrand values must lie between 0 and 1, then those methods can construct a confidence interval that has at least $1-\alpha$ probability to contain the integral's value. This confidence level holds for any integrand satisfying the given bounds.

The context for our work is as follows. Many scientific problems require the numerical evaluation of multidimensional integrals. When the dimension is high enough, then classical tensor product integration rules like those in [8] become too expensive to use. It is then common to use plain Monte Carlo (MC) methods instead. MC methods based on random sampling typically have root mean squared errors (RMSEs) of $O(n^{-1/2})$ given n function evaluations. This rate is slow but it is the same in any dimension. MC also allows one to use statistical methods to quantify the uncertainty in the estimated integral. ~~Quasi-Monte~~ Ordinary quasi-Monte Carlo (QMC) methods (e.g., [10] and [28]) are deterministic sampling strategies that under a bounded variation assumption produce integral estimates with error $\tilde{O}(n^{-1})$ where $\tilde{O}(\cdot)$ means we neglect powers of $\log(n)$. ~~Plain~~ Because plain QMC methods are deterministic~~and so~~, they lose the uncertainty quantification advantage of MC. RQMC methods, surveyed in [22], allow one to make independent replicates of a statistically unbiased QMC method to support variance estimates and asymptotic confidence intervals. RQMC methods can also improve on the convergence rate of QMC methods obtaining an RMSE of $\tilde{O}(n^{-3/2})$ under a smoothness assumption [31, 33] on the integrand.

In addition to asymptotic confidence interval methods for RQMC, some other papers have provided finite sample uncertainty quantifications based on additional assumptions. For purely independent and identically distributed (IID) sampling, knowing a bound on the kurtosis facilitates a finite sample confidence interval for the mean of a random variable in terms of the sample mean and the sample standard deviation [13]. For deterministic QMC sampling of $\boldsymbol{x} \sim \mathbb{U}[0,1]^d$ with lattices or digital nets, assuming that the coefficients in a suitable orthogonal decomposition of $f : [0,1]^d \to \mathbb{R}$ decay reasonably, one may derive a deterministic error bound on $\mu = \mathbb{E}[f(\boldsymbol{x})]$ [14, 15, 21, 41]. A different approach is to assume that f is a realization of a Gaussian process and construct a credible interval for $\mathbb{E}[f(\boldsymbol{x})]$. If the covariance kernel and the sampling nodes are well-matched, the computational effort required is nearly $O(n)$ and not the $O(n^3)$ effort typically required for such credible intervals [19, 20]. The GAIL [5] and QMCPy [6] libraries implement these finite sample uncertainty quantifications under the assumptions outlined above.

Here, we study RQMC confidence intervals assuming that the integrand is bounded between zero and one. The advantage of this assumption is that there are more settings where we are certain that it holds.

The non-asymptotic confidence intervals that we emphasize are the recently developed hedged betting confidence intervals (HBCI) and some related predictable plug-in empirical Bernstein confidence intervals (EBCI), both from [44]. They are derived from some infinite confidence sequences and we believe that they provide the narrowest confidence intervals among currently available methods.

Section 2 gives our notation and some definitions and background on quasi-Monte Carlo, randomized quasi-Monte Carlo, and the EBCI and HBCI of [44]. In Section 3 we show how splitting N observations into $R = N/n$ independent replicates of n RQMC point sets can give narrower EBCI than MC does for an oracle that knows the RQMC variance at each n. The extent of the narrowing depends on the RQMC convergence rate which varies from problem to problem. Perhaps surprisingly, the more effective RQMC is, the smaller the optimal value of n becomes. For an RQMC variance proportional to $n^{-\theta}$, taking n proportional to $N^{1/(\theta+1)}$ gives the narrowest intervals. The resulting interval widths are $\Theta(N^{-\theta/(\theta+1)})$ which can be much wider than the assumed RQMC standard deviation of $\Theta(N^{-\theta/2})$. Section 4 makes some computational investigations. As predicted by the theory the best values of n to use ~~grew~~ grow slowly with N. The empirically best values of n for HBCI ~~were~~ are no larger than the oracle values, and sometimes smaller. Section 5 has some final comments and discussion.

2 Background and notation

We begin with some notation. We use $1\{E\}$ to denote a quantity that equals 1 when expression E holds and is zero otherwise. The vectors $\mathbf{0}$ and $\mathbf{1}$ have all components equal to 0 and 1 respectively with a dimension given by context. We write $a_n = O(b_n)$ as $n \to \infty$ if there exists $C > 0$ and $n_0 < \infty$ with $|a_n| \leqslant Cb_n$ whenever $n \geqslant n_0$. We write $a_n = \Omega(b_n)$ if there exists $C > 0$ and $n_0 < \infty$ with $a_n \geqslant Cb_n$ whenever $n \geqslant n_0$. We write $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $b_n = O(a_n)$.

2.1 Intervals and martingales

For $0 \leqslant \alpha \leqslant 1$, a $1-\alpha$ confidence interval for a parameter μ is a pair of random quantities A and B with

$$\Pr(A \leqslant \mu \leqslant B) \geqslant 1-\alpha. \tag{1}$$

A confidence interval is strict when the right hand side above is an equality. An asymptotic confidence interval is a sequence (A_n, B_n) of random quantities for which

$$\lim_{n\to\infty} \Pr(A_n \leqslant \mu \leqslant B_n) \geqslant 1-\alpha.$$

It is common for asymptotic confidence intervals to be strict, meaning that this limit equals $1-\alpha$. The best known example is Student's t confidence interval for the mean of a distribution with finite variance, based on a random sample from that distribution and justified by the central limit theorem.

The random sequence (A_n, B_n) is a $1-\alpha$ confidence sequence for μ if

$$\Pr(A_n \leqslant \mu \leqslant B_n,\ \forall n \geqslant 1) \geqslant 1-\alpha. \tag{2}$$

Confidence sequences allow us to select a stopping time ν based on $\{(A_n, B_n) \mid n \leqslant \nu\}$ and have at least $1-\alpha$ confidence that $A_\nu \leqslant \mu \leqslant B_\nu$. We can take the limits to be $\max_{n\leqslant\nu} A_n$ and $\min_{n\leqslant\nu} B_n$.

Martingale theory is one of the main tools for obtaining confidence sequences. If random variables

S_i for $i \geqslant 1$ satisfy $\mathbb{E}(S_n \,|\, S_1, \dots, S_{n-1}) = S_{n-1}$ then $(S_i)_{i \geqslant 1}$ is a martingale. If $\mathbb{E}(S_n \,|\, S_1, \dots, S_{n-1}) \leqslant S_{n-1}$ then $(S_i)_{i \geqslant 1}$ is a supermartingale. Ville's inequality is that for any $\eta > 0$ and any nonnegative supermartingale $(S_i)_{i \geqslant 1}$

$$\Pr\Bigl(\sup_{n \geqslant 1} S_n \geqslant \eta\Bigr) \leqslant \frac{\mathbb{E}(S_1)}{\eta}.$$

The confidence sequences from [44] are derived by applying Ville's inequality [42] to nonnegative supermartingales.

The problem we consider is to approximate the finite-dimensional integral

$$\mu = \int_{[0,1]^d} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \mathbb{E}(f(\boldsymbol{x})), \quad \text{for } \boldsymbol{x} \sim \mathbb{U}[0,1]^d.$$

In MC sampling we take $\boldsymbol{x}_i \overset{\text{iid}}{\sim} \mathbb{U}[0,1]^d$ and estimate μ by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}_i).$$

Let $\sigma^2 = \mathrm{var}(f(\boldsymbol{x}))$ for $\boldsymbol{x} \sim \mathbb{U}[0,1]^d$. We assume that $0 < \sigma^2 < \infty$. In that case the RMSE of $\hat{\mu}$ is $\sigma/\sqrt{n}$. For $n \geqslant 2$,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (f(\boldsymbol{x}_i) - \hat{\mu})^2$$

is an unbiased estimate of σ^2. Furthermore

$$\lim_{n \to \infty} \Pr\Bigl(\sqrt{n}\frac{\hat{\mu} - \mu}{s} \leqslant t_{(n-1)}^{1-\alpha}\Bigr) = 1 - \alpha$$

where $t_{(n-1)}^{1-\alpha}$ is the $1-\alpha$ quantile of Student's t distribution on $n-1$ degrees of freedom. Then $\hat{\mu} \pm s t^{1-\alpha/2}/\sqrt{n}$ is an asymptotic $1-\alpha$ confidence interval for μ.

2.2 QMC and RQMC

QMC sampling replaces random points $\boldsymbol{x}_i$ by deterministic points that more evenly sample the unit cube. There are many ways to quantify that property. The most elementary one is the star discrepancy:

$$D_n^* = D_n^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \sup_{\boldsymbol{a}\in[0,1]^d} \left| \frac{1}{n}\sum_{i=1}^n 1\{\boldsymbol{x}_i \in [\boldsymbol{0}, \boldsymbol{a})\} - \prod_{j=1}^d a_j \right|.$$

Small values of D_n^* show that all anchored boxes $[\boldsymbol{0}, \boldsymbol{a})$ have very nearly the desired proportion of the n points, which is the volume of $[\boldsymbol{0}, \boldsymbol{a})$. It is possible to choose points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ so that $D_n^* = \tilde{O}(n^{-1})$ while plain MC points have $D_n^* = \tilde{O}(n^{-1/2})$ [28].

The improvement in star discrepancy yields an improvement in integration error via the Koksma-Hlawka inequality [12]:

$$|\hat{\mu} - \mu| \leqslant D_n^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \times V_{\mathrm{HK}}(f). \tag{3}$$

As described above D_n^* is a measure of how non-uniform the points $\boldsymbol{x}_i$ are. The new factor $V_{\mathrm{HK}}(f)$ is the total variation of f in the sense of Hardy and Krause. See [32]. It follows from (3) that we can get $|\hat{\mu} - \mu| = \tilde{O}(n^{-1})$ using QMC. We will refer to this as the BVHK rate.

If we knew D_n^* and $V_{\mathrm{HK}}(f)$, then (3) would provide a perfect quantification of our uncertainty about μ. It would be non-asymptotic and even non-probabilistic. Unfortunately, D_n^* is generally very expensive to obtain and $V_{\mathrm{HK}}(f)$ is ordinarily far harder to compute than μ. As a result, equation (3) shows us that MC can be outperformed but it does not provide a generally usable error estimate.

RQMC methods generate random points $\boldsymbol{x}_i$ that simultaneously satisfy two properties:

1) $\boldsymbol{x}_i \sim \mathbb{U}[0,1]^d$, for all $i = 1, \dots, n$, and

2) $D_n^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \tilde{O}(n^{-1})$ almost surely.

From the second property, $|\hat{\mu} - \mu| = \tilde{O}(n^{-1})$. As a result the RMSE of RQMC is $\tilde{O}(n^{-1})$ in the BVHK case.

To get an approximate confidence interval from RQMC we can form R statistically independent RQMC estimates $\hat{\mu}_1, \dots, \hat{\mu}_R$. Using the first property above these IID estimates are unbiased and we can get an asymptotic confidence interval for μ of the form

$$\hat{\mu} \pm S t_{(R-1)}^{1-\alpha/2} / \sqrt{R}$$

for

$$\hat{\mu} = \frac{1}{R}\sum_{i=1}^{R}\hat{\mu}_i \quad \text{and} \quad S^2 = \frac{1}{R-1}\sum_{i=1}^{R}(\hat{\mu}_i - \hat{\mu})^2.$$

Here S^2/R is an unbiased estimate of the MSE of the pooled estimate $\hat{\mu}$. The confidence level $1-\alpha$ is obtained in the limit as $R \to \infty$ [27] for fixed n. That convergence is generally quicker for more nearly Gaussian $\hat{\mu}_i$.

In addition to getting an unbiased estimate of the RQMC variance, randomization can provide some other benefits. First, some RQMC methods can attain better accuracy than plain QMC. Under sufficient smoothness, the scrambled digital sequences in [30] attain RMSEs of $\tilde{O}(n^{-3/2})$ [31, 33]. The same holds for the scramblings of [25]. We will refer to this as the 'smooth case'. Higher order digital nets [9] can attain even better convergence rates, though the rates are not necessarily evident empirically when d is moderately large, which [29] attribute to numerical precision.

Those RQMC methods have further useful properties. First, $\mathrm{var}(\hat{\mu}) = o(1/n)$ as $n \to \infty$ for any integrand with $\sigma^2 < \infty$, and so their efficiency with respect to MC becomes unbounded as $n \to \infty$. Furthermore, the condition of finite variation in the sense of Hardy and Krause can be quite strict. For example, when $d \geqslant 2$, step discontinuities in the integrand f ordinarily make $V_{\mathrm{HK}}(f)$ infinite unless those discontinuities are axis parallel [32]. This makes QMC unattractive for estimating integrals of binary valued functions, while RQMC will still attain an RMSE that is $o(n^{-1/2})$.

While there are numerous RQMC methods in use, we will restrict ourselves here to scramblings of digital nets and sequences, such as the ones of Sobol' [39]. Our variance formulas assume that the scramblings are either those of [30] or [25]. In our computations we used the linear matrix scramble of [25] with a digital shift.

While RQMC methods yield very good accuracy, most of the known confidence interval methods for them supply only asymptotic confidence as $R \to \infty$ [27]. In some very limited settings there are central limit theorems for $\hat{\mu}$ as $n \to \infty$ [24]. Having the confidence statements be asymptotic in R is unfortunate since the RMSE of $\hat{\mu}$ is proportional to $R^{-1/2}$ while the RMSE vanishes at a faster rate in n.

2.3 Hoeffding and Empirical Bernstein intervals

We would like a confidence interval that is non-asymptotic, meaning that the coverage guarantee (1) holds for limits A and B computed from a sample of size n. This can often be obtained under parametric statistical models in which the distribution of $f(\boldsymbol{x})$ belongs to a known finite dimensional family such as the Gaussian distributions. It is not desirable to make such assumptions and so we prefer a nonparametric method with non-asymptotic confidence.

There is a theorem of Bahadur and Savage [2] that reveals sharp constraints on our ability to construct nonparametric and non-asymptotic confidence intervals. Here is the description based on [35]:

> They consider a set $\mathcal{F}$ of distributions on $\mathbb{R}$. Letting $\mu(F) = \mathbb{E}(Y)$ for $Y \sim F$, their conditions are:
>
> (i) For all $F \in \mathcal{F}$, $\mu(F)$ exists and is finite.
>
> (ii) For all $m \in \mathbb{R}$ there is $F \in \mathcal{F}$ with $\mu(F) = m$.
>
> (iii) $\mathcal{F}$ is convex: if $F, G \in \mathcal{F}$ and $0 < \pi < 1$, then $\pi F + (1-\pi)G \in \mathcal{F}$.
>
> Then their Corollary 2 shows that a Borel set constructed based on $Y_1, \dots, Y_N \overset{\text{iid}}{\sim} F$ that contains $\mu(F)$ with probability at least $1-\alpha$ also contains any other $m \in \mathbb{R}$ with probability at least $1-\alpha$. More precisely: we can get a confidence set, but not a useful one. They allow N to be random so long as $\Pr(N < \infty) = 1$.

We can escape the restriction from Bahadur and Savage by stipulating that every distribution in $\mathcal{F}$ has support in a known interval $[a, b]$ of finite length. This violates their clause (ii). After a linear transformation we take that interval to be $[0, 1]$.

We will use $Y_i = \hat{\mu}_i$ for $i = 1, \dots, R$ to represent the values of the R replicated RQMC points that go into our confidence interval calculation. We set $\bar{Y}_R = (1/R)\sum_{i=1}^R Y_i$. We assume that Y_i are IID with the same distribution as Y and that $0 \leqslant Y \leqslant 1$ holds with probability one. Then

$$\bar{Y}_R \pm \sqrt{\frac{\log(2/\alpha)}{2R}} \tag{4}$$

for $0 < \alpha < 1$ is a $1-\alpha$ confidence interval for $\mu = \mathbb{E}(Y_i)$. Equation (4) gives the confidence interval of Hoeffding [16]. The interval is not always nested within $[0,1]$, but we can simply take its intersection with $[0,1]$. That will preserve the confidence level.

While Hoeffding's interval is a valid confidence interval for the mean of any distribution supported on $[0,1]$, it can be quite conservative for some of them, giving confidence over $1-\alpha$. This raises the possibility of getting coverage $1-\alpha$ from narrower intervals. One approach is to make use of $\sigma^2 = \mathrm{var}(Y)$ if it is known. Hoeffding's confidence interval is a consequence of Hoeffding's exponential probability bound in [16]. Other exponential probability bounds take advantage of a known value $\sigma^2 = \mathrm{var}(Y)$ or even a sample estimate of that variance. Those bounds can be used to get improved confidence intervals.

Theorem 3 of Maurer and Pontil [26] is that

$$\mathbb{E}(Y) - \frac{1}{R}\sum_{i=1}^{R} Y_i \leqslant \sqrt{\frac{2\sigma^2\log(1/\delta)}{R}} + \frac{\log(1/\delta)}{3R}.$$

holds with probability at least $1-\delta$. They call this Bennett's inequality, because it derives from an inequality in [3]. They do not supply a proof, but it can be proved from Bernstein's inequality given as Theorem 3 of [4]. Taking $\delta = \alpha/2$ we get

$$\bar{Y}_R \pm \sqrt{\frac{2\sigma^2\log(2/\alpha)}{R}} + \frac{\log(2/\alpha)}{3R} \tag{5}$$

as a $1-\alpha$ confidence interval for μ. We will make use of the Bennett confidence intervals (5) for some theoretical investigations.

In most practical settings where μ is unknown, σ^2 will also be unknown. Replacing σ^2 by s^2 in (5) would not always give a valid confidence interval. Theorem 4 of [26] shows that for $R \geqslant 2$

$$\bar{Y}_R \pm \sqrt{\frac{2s^2\log(4/\alpha)}{R}} + \frac{7\log(4/\alpha)}{3(R-1)}. \tag{6}$$

is a $1-\alpha$ confidence interval for μ. This is called an empirical Bernstein confidence interval named after an inequality of Bernstein that is similar to Bennett's. We see that both terms on the right hand side of (6) are larger than the ones in the Bennett interval (5). Their interval has improved asymptotic constants compared to the original empirical Bernstein inequality from [1]. The EBCI and HBCI from [44] provide a further asymptotic improvement.

2.4 Betting methods

Our interest in this problem was sparked in part by some recent work by [44] on confidence intervals generated by betting arguments. They consider a more general setting of confidence sequences, as in equation (2), that can then be specialized to confidence intervals as needed. For random $Y_i \in [0,1]$ they assume that $\mathbb{E}(Y_i \,|\, Y_1, \dots, Y_{i-1}) = \mu$ for all $i \geqslant 1$ without assuming that the Y_i are identically distributed or even that they are independent. For $i = 1$, the conditional expectation above is simply $\mathbb{E}(Y_1) = \mu$.

Suppose that the conditional mean is really μ and there is a null hypothesis $H_0(m)$ that $\mathbb{E}(Y_i \,|\, Y_1, \dots, Y_{i-1}) = m$. Then if $m \neq \mu$ somebody starting with a stake of \$1.00 can make a series of bets against $H_0(m)$ as the Y_i are revealed in order and have an expected fortune that grows without bound. Just prior to time i, the bettor picks a value ~~$\lambda_i = \lambda_i(m) \in (-1/(1-m), 1/m)$~~ $\lambda_i = \lambda_i(m) \in (-1/(1-m), 1/m)$. The bettor's capital at time $t \geqslant 1$ is

$$\mathcal{K}_t(m) = \prod_{i=1}^{t} \bigl(1 + \lambda_i(m)(Y_i - m)\bigr).$$

Taking $\lambda_i(m) > 0$ amounts to betting that $Y_i > m$. Then their capital is multiplied by $1 + \lambda_i(Y_i - m)$. If they are right then their capital grows, but if $Y_i < m$, then it shrinks. To take $\lambda_i(m) < 0$ is to bet that $Y_i < m$.

Section 4 of [44] notes that $\mathcal{K}_t(\mu)$ is a nonnegative martingale, so that $\mathbb{E}(\mathcal{K}_t(\mu)) = 1$ for all $t \geqslant 1$. There is no way for the bettor to pick bet sizes $\lambda_i(\mu)$ to make an expected profit. Then from Ville's Theorem

$$\Pr(\mathcal{K}_t(\mu) \leqslant 1/\alpha,\ \forall t \geqslant 1) \geqslant 1 - \alpha.$$

If we watch $\mathcal{K}_t(\mu)$ indefinitely, then there is at most probability α that it will ever go above $1/\alpha$.

As a result we can get a confidence sequence by retaining at time t, all the values m for which $\max_{1\leqslant\tau\leqslant t}\mathcal{K}_\tau(m)\leqslant 1/\alpha$. This idiom uses a hypothetically infinite ensemble of bettors to do this, but [44] give implementable algorithms for it.

The other side of the coin is that there are good betting strategies when $m\neq\mu$. Under those, the bettor's expected fortune will grow steadily when $m\neq\mu$. That is what one needs to make a confidence sequence not just valid but also useful in having a width that converges to zero with increased sampling.

One version of the betting strategy above produces confidence sequences that are analogous to Hoeffding intervals and another produces empirical Bernstein intervals. We present the predictable plug-in EBCI followed by a hedged betting interval that combines two betting strategies. We do not include the predictable plug-in Hoeffding intervals. Those do not use a sample variance, and they are wider than the others for larger t.

The confidence intervals we study make use of running sample moments

$$\hat\mu_t=\frac{1/2+\sum_{i=1}^t Y_i}{t+1}\quad\text{and}\quad\hat\sigma_t^2=\frac{1/4+\sum_{i=1}^t(Y_i-\hat\mu_i)^2}{t+1}$$

with the means and standard deviations both biased slightly towards 1/2. In their 'predictable plug-in empirical Bernstein' intervals, the betting amounts are

$$\lambda_t^{\text{PrPl-EB}}=\sqrt{\frac{2\log(2/\alpha)}{\hat\sigma_{t-1}^2 t\log(1+t)}}\wedge c$$

for some $c\in(0,1)$ with $c=1/2$ or $3/4$ given as reasonable defaults. Note that this method only bets that $Y_i>m$. Theorem 2 of [44] gives a $1-\alpha$ confidence sequence

$$C_t^{\text{PrPl-EB}}=\frac{\sum_{i=1}^t\lambda_iY_i}{\sum_{i=1}^t\lambda_i}\pm\frac{\log(2/\alpha)+\sum_{i=1}^t\nu_i\psi_{\mathrm{E}}(\lambda_i)}{\sum_{i=1}^t\lambda_i},\tag{7}$$

where

$$\nu_i=4(Y_i-\hat\mu_{i-1})^2\quad\text{and}\quad\psi_{\mathrm{E}}(\lambda)=(-\log(1-\lambda)-\lambda)/4.$$

The running intersection $\cap_{1\leqslant\tau\leqslant t} C_\tau^{\text{PrPl-EB}}$ is also a $1-\alpha$ confidence sequence for μ and of course we can intersect any of these intervals with $[0,1]$.

These confidence sequences can be specialized to confidence intervals when we have a fixed target sample size R in mind. For that case [44] recommends

$$\lambda_i^{\text{PrPl-EB}(R)} = \sqrt{\frac{2\log(2/\alpha)}{R\hat\sigma_{i-1}^2}} \wedge c, \quad i = 1,\dots,R. \tag{8}$$

The above formula for λ_i uses the ordering of the data values and puts unequal weight on the R different Y_i. There is a permutation strategy in [44] to treat the Y_i values more symmetrically, but they report that it makes little difference.

The EBCI we study take the form

$$\bigcap_{1\leqslant\tau\leqslant t} C_t^{\text{PrPl-EB}(R)}$$

where $C_t^{\text{PrPl-EB}(R)}$ is the interval in (7) using λ_i from (8). Equation (17) of [44] gives the scaled half width of these EBCI as

$$\sqrt{R}\biggl(\frac{\log(2/\alpha)+\sum_{i=1}^R \nu_i\psi_{\text{E}}(\lambda_i)}{\sum_{i=1}^R \lambda_i}\biggr) \stackrel{\text{a.s.}}{\to} \sigma\sqrt{2\log(2/\alpha)}$$

as $R\to\infty$. The corresponding limit for the empirical Bernstein CIs of [26] is $\sigma\sqrt{2\log(4/\alpha)}$.

The hedged betting process of [43] works as follows. For positive predictable sequences $\lambda_i^+(m)$ and $\lambda_i^-(m)$ defined below, they use capital processes

$$K_t^+(m) = \prod_{i=1}^t \bigl(1+\lambda_i^+(m)(Y_i-m)\bigr) \quad \text{and} \quad K_t^-(m) = \prod_{i=1}^t \bigl(1-\lambda_i^+(m)(Y_i-m)\bigr)$$

that bet, respectively, on $Y_i > m$ and $Y_i < m$. The hedged process is

$$K_t^{\pm}(m) = \max\bigl(\theta K_t^+(m), (1-\theta)K_t^-(m)\bigr)$$

for $0 \leqslant \theta \leqslant 1$. For any θ, $K_t^{\pm}(\mu)$ is a nonnegative martingale hence subject to Ville's theorem. The supplement to [44] suggests $\theta = 1/2$ as a default. If $m \neq \mu$ then $K_t^{\pm}(m)$ will tend to grow with t. Of course the hedging θ has nothing to do with the RQMC variance rate θ and the quantities are distinct enough that no confusion should arise.

The hedged betting process produces confidence sequences and, as they did for the empirical Bernstein processes, they recommend a way to produce a confidence interval based on R sample values. For that they take

$$\lambda_i^+(m) = |\tilde{\lambda}_i^{\pm}| \wedge \frac{c}{m} \quad \text{and} \quad \lambda_i^-(m) = |\tilde{\lambda}_i^{\pm}| \wedge \frac{c}{1-m}$$

for $i = 1, \dots, R$ where

$$\tilde{\lambda}_i^{\pm} = \sqrt{\frac{2\log(2/\alpha)}{R\hat{\sigma}_{i-1}^2}}$$

for $c \in [0, 1)$, with suggested defaults of $c = 1/2$ or $c = 3/4$. The confidence interval is then

$$\bigl\{m \in [0,1] \mid \max_{1 \leqslant i \leqslant R} K_i^{\pm}(m) \leqslant 1/\alpha\bigr\}.$$

3 Asymptotic confidence interval comparisons

Our confidence intervals will use R IID replicates that each use n RQMC points in $[0,1]^d$. That takes $N = nR$ evaluations of f. Here we study how to get the narrowest confidence intervals with that budget of N function evaluations.

We take

$$Y_i = \frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{x}_{i,k})$$

where for each i, $\boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,n}$ is an RQMC point set. The Y_i are IID when we use independent scrambles for all R point sets. Let $\sigma_n^2 = \text{var}(Y_i)$. We know that $\sigma_n^2 = o(n^{-1})$. For the BVHK case, $\sigma_n^2 = \tilde{O}(n^{-2})$. For the smooth case, $\sigma_n^2 = \tilde{O}(n^{-3})$.

We know from [44] that the widths of their predictable plug-in EBCI approaches that of some

Bennett intervals described below, as $R \to \infty$ for fixed n. The same is true of their HBCI. Here we study the value of n that optimizes the Bennett intervals.

As mentioned above, the tilde in $\tilde{O}$ hides powers of $\log(n)$. Those powers are present in error bounds for adversarially chosen integrands but do not seem to come up in real problems [38, 36]. We will consider working models of the form

$$\sigma_n^2 = \sigma_0^2 n^{-\theta}, \tag{9}$$

even though the actual variance is not necessarily any power of n. This approach is less cumbersome than bounding the variance between $\Omega(n^{-\theta})$ and $O(n^{-\theta+\epsilon})$, and we believe it provides clearer insights.

We know that taking $\theta = 1$ underestimates the quality of RQMC because the variance is $o(1/n)$ for the RQMC methods we use. The smooth case with $\theta = 3$ is optimistic while the BVHK case with $\theta = 2$ is intermediate.

If $R \to \infty$ for fixed n, then the predictable plug-in EBCI have half widths $H^{\text{PrPl-EB}}$ that satisfy

$$\sqrt{R} H^{\text{PrPl-EB}(R)} \to \sigma_n \sqrt{2\log(2/\alpha)}. \tag{10}$$

The HBCI have half widths H^{PrPl} that satisfy the same rate. As a result, both methods give interval widths that are $O(n^{-\theta/2} R^{-1/2})$ as $R \to \infty$. For a fixed product $N = nR$, this width is narrowest at $R = 1$ and $n = N$, but of course that is infeasible and also that argument ignores the fact that (10) is based on a limiting argument as $R \to \infty$.

To study the tradeoff between n and R, we use Bennett's inequality which gives a half width of

$$H^{\text{Ben}}(n, R) = \sqrt{\frac{2\sigma_n^2 \log(2/\alpha)}{R}} + \frac{\log(2/\alpha)}{3R} \tag{11}$$

when using R IID observations that each have variance σ_n^2. An oracle that knew σ_n^2 could use this formula to select the best n and R for a confidence interval subject to a constraint $nR = N$. The predictable plug-in EBCI and the HBCI from [44] do not assume a known variance for the Y_i.

However those intervals attain the same asymptotic limiting width as $R \to \infty$ that an oracle would get from Bennett's formula.

Next, we investigate the oracle's choice of n and R to minimize the half width subject to $nR = N$. While n and R both have to be positive integers with product N, we will first relax the problem to a continuous variable n with $R = N/n$. After that, we discuss integer solutions.

Theorem 1. *Let σ_n^2 follow* (9) *and choose $N > 0$. If $\theta = 1$, then the minimizer of* (11) *over $n \in [1, \infty)$ is $n_* = 1$. If $\theta > 1$, then the minimizer of* (11) *over $n \in (0, \infty)$ is*

$$n_* = \left(\frac{9(\theta - 1)^2\, \sigma_0^2 N}{2 \log(2/\alpha)} \right)^{1/(\theta+1)}. \tag{12}$$

Proof. Under the model (9), with $R = N/n$ the oracle's half width is

$$H(n) = H^{\mathrm{Ben}}(n, N/n) = \sigma_0 n^{(1-\theta)/2} N^{-1/2} \sqrt{2 \log(2/\alpha)} + \frac{\log(2/\alpha)}{3} \frac{n}{N}. \tag{13}$$

If $\theta = 1$, then $H(n)$ is minimized over $[1, \infty)$ at $n = 1$.

For $\theta > 1$, the function $H(n)$ is a convex function of n. It has a unique minimum over $(0, \infty)$ at some n_* where $H'(n_*) = 0$. Now

$$H'(n) = \frac{1-\theta}{2} \frac{\sigma_0}{\sqrt{N}} n^{-(\theta+1)/2} \sqrt{2 \log(2/\alpha)} + \frac{\log(2/\alpha)}{3N}.$$

This vanishes at

$$n^{-(\theta+1)/2} = \frac{2 \log(2/\alpha)/(3N)}{(\theta - 1)\sigma_0 \sqrt{2 \log(2/\alpha)/N}} = \frac{\sqrt{2 \log(2/\alpha)}}{3(\theta - 1)\sigma_0 \sqrt{N}}$$

from which (12) follows. □

If $n_* < 1$, then $n = 1$ is best. This includes the degenerate case with $\sigma_0 = 0$. If $n_* > N$, then $n = N$ is best. If $1 < n_* < N$ and n_* is not an integer then the best integer n is either $\lfloor n_* \rfloor$ or $\lceil n_* \rceil$ by the convexity noted in the proof.

For scrambled Sobol' points it is generally best to take n to be a power of 2 [34], which is not reflected in Theorem 1. We incorporate that restriction into some numerical examples in Section 4.

As noted above, when $\theta = 1$, the best n is $n = 1$. In that case RQMC is the same as MC, using one uniformly distributed point in $[0,1]^d$. This result holds for any sampler that gets the ordinary MC rate, even if it has a favorable constant. In particular, any MC variance reduction techniques that do not change the convergence rate will not lead to using $n > 1$. They will improve the asymptotic confidence interval width by reducing σ_0.

When $\theta > 1$, we see that the shortest intervals come from taking $n = \Theta(N^{1/(\theta+1)})$. The BVHK rate ($\theta = 2$) is then $n = \Theta(N^{1/3})$ while the rate for smooth integrands ($\theta = 3$) is $n = \Theta(N^{1/4})$. We see that when the RQMC convergence rate becomes better, the optimal RQMC sample size to use grows at a slower rate in N. This may seem like a paradox, but the explanation can be seen in equation (11). If σ_n drops very quickly then the $O(1/\sqrt{R})$ term it is in does not dominate the other $O(1/R)$ term as much and a larger R is better.

Because the oracle takes n to be a smaller power of N for larger θ, we might wonder why it always picks $n = 1$ for the smallest value, $\theta = 1$. This stems from the factor $\theta - 1$ in the constant of proportionality in (12). For fixed σ_0, N and α we could maximize (12) over θ to see which rate gives the largest n.

What we see from Theorem 1 is that the optimal n grows slowly with N. The variance of the average of R RQMC samples using $n = N/R$ is

$$\Theta(R^{-1}n^{-\theta}) = \Theta(N^{-1}n^{1-\theta}) = \Theta(N^{-1}N^{(1-\theta)/(\theta+1)}) = \Theta(N^{-2\theta/(\theta+1)}).$$

By requiring a guaranteed confidence level, rather than an asymptotic one, we get a higher variance for $\hat{\mu}$ than $\Theta(N^{-\theta})$ that we could have had with $R = O(1)$. For $\theta = 2$ the variance rises from $\Theta(N^{-2})$ to $\Theta(N^{-4/3})$ and for $\theta = 3$, it rises from $\Theta(N^{-3})$ to $\Theta(N^{-3/2})$. Put another way, obtaining a nonasymptotic confidence level raises the variance by $\Theta(N^{2/3})$ and $\Theta(N^{3/2})$ in those two cases.

While we cannot always know the best value of n to use in a given setting, Theorem 1 gives us some guidance. The asymptotic value of θ may be known from theory. We may also have experience with similar integrands to make a judgment about what value of θ is reasonable in a given setting, or we can do some preliminary sampling to get a working value of θ before constructing a betting confidence interval. For any choice of θ, that interval will be valid, though possibly wider than an interval using the true optimal n. Because $0 \leqslant Y_i \leqslant 1$, we know that $\sigma_0^2 \leqslant 1/4$. Therefore under the model (9), Theorem 1 provides the guidance to take

$$n \leqslant \Bigl(\frac{9(\theta-1)^2 N}{8\log(2/\alpha)}\Bigr)^{1/(\theta+1)}. \tag{14}$$

For example, with $N = 2^{10}$ and $\alpha = 0.05$, the bound in (14) is always below 7 whenever $1 \leqslant \theta \leqslant 3$. The best integer value is 7 for some values of θ. For higher confidence levels (smaller α), the optimal n is smaller still.

The values of N and α enter the Bennett half width of (13) only through the expression

$$N_\alpha \equiv \frac{N}{\log(2/\alpha)}.$$

Thus we can think of smaller α as providing a smaller effective budget N_α.

Theorem 2. *Under the conditions of Theorem 1 with $\theta > 1$ and $\sigma_0 > 0$, let H_{MC} be the half width $H^{\mathrm{Ben}}(1, N)$ using $n = 1$ and H_{RQMC} be the half width $H^{\mathrm{Ben}}(n, N/n)$ using $n = n_*$ from (12). Let $N_\alpha = N/\log(2/\alpha)$. Then for $\theta > 1$*

$$\begin{aligned}\frac{H_{\mathrm{RQMC}}}{H_{\mathrm{MC}}} &= \frac{\theta+1}{\sqrt{2\sigma_0^2 N_\alpha + 1/3}}\Bigl(\frac{1}{2}[3(\theta-1)]^{1-\theta} N_\alpha \sigma_0^2\Bigr)^{1/(\theta+1)} \\ &= \Theta(N^{\frac{1-\theta}{2\theta+2}})\end{aligned}$$

as $N \to \infty$.

Proof. Using equation (13) in the numerator, the ratio of these widths is

$$\begin{aligned}\rho = \frac{H_{\mathrm{RQMC}}}{H_{\mathrm{MC}}} &= \frac{n^{(1-\theta)/2}\sqrt{2\sigma_0^2\log(2/\alpha)/N} + n\log(2/\alpha)/(3N)}{\sqrt{2\sigma_0^2\log(2/\alpha)/N} + \log(2/\alpha)/(3N)}\\ &= \frac{n^{(1-\theta)/2}\sqrt{2\sigma_0^2 N_\alpha} + n/3}{\sqrt{2\sigma_0^2 N_\alpha} + 1/3}.\end{aligned}$$

From equation (12)

$$n^{(1-\theta)/2} = (3(\theta-1)\sigma_0\sqrt{N_\alpha/2})^{(1-\theta)/(\theta+1)}$$

and we may also write $n = (3(\theta-1)\sigma_0(N_\alpha/2)^{1/2})^{2/(\theta+1)}$.

So for $\theta > 1$, the width ratio is

$$\begin{aligned}\rho &= \frac{(3(\theta-1)\sigma_0\sqrt{N_\alpha/2})^{(1-\theta)/(\theta+1)}\sqrt{2\sigma_0^2 N_\alpha} + (3(\theta-1)\sigma_0(N_\alpha/2)^{1/2})^{2/(\theta+1)}/3}{\sqrt{2\sigma_0^2 N_\alpha} + 1/3}\\ &= \frac{(3(\theta-1))^{(1-\theta)/(\theta+1)}2^{\theta/(\theta+1)} + (3(\theta-1))^{2/(\theta+1)}2^{-1/(\theta+1)}/3}{\sqrt{2\sigma_0^2 N_\alpha} + 1/3}(N_\alpha\sigma_0^2)^{1/(\theta+1)}\\ &= \frac{\theta+1}{\sqrt{2\sigma_0^2 N_\alpha} + 1/3}\left(\frac{1}{2}[3(\theta-1)]^{1-\theta}N_\alpha\sigma_0^2\right)^{1/(\theta+1)}\end{aligned}$$

as required. □

For the BVHK case with $\theta = 2$, we find that RQMC narrows the widths by $\Theta(N^{-1/6})$. For the smooth case with $\theta = 3$, we get a more favorable width ratio of $\Theta(N^{-1/4})$. The MC widths are $\Theta(N^{-1/2})$, and so for the BVHK case the RQMC widths are $\Theta(N^{-2/3})$ while for the smooth case, they are $\Theta(N^{-3/4})$.

4 Finite sample evaluations

Here we present some finite sample evaluations of the EBCI (specifically the predictable plug-in one) and the HBCI, both using RQMC, for 95% confidence. We begin with finite N and for some integrands with known RQMC variance, we find the n that minimizes the interval widths from

Bennett's inequality. They are then the exact best n for Bennett's inequality which the HBCIs and EBCIs of [44] match asymptotically. After that we present finite sample results on the EBCIs and HBCIs for a variety of integrands differing in smoothness and dimension. The HBCIs are usually narrower. Finally, we compare the empirically optimal n for the HBCIs to the values that optimize Bennett's inequality. The HCBI widths approach those from Bennett's inequality for large R and we find that the optimal n for HBCI tend to be the same or smaller than the ones we get from Bennett's inequality.

For our HBCI computations, we used $\theta = 1/2$. ~~The~~ The simulations presented here were run at `https://github.com/aaditj1962161/Betting-Paper-Simulations-for-QMC`.The software we used comes from [45] which has code to reproduce the figures in [44]. It calls the HBCI and predictable plug-in EBCI algorithms from [17]. The scrambled Sobol' points we used were from QMCPy [6].

4.1 Semi-empirical evaluations

We can compute the width of intervals based on Bennett's inequality for some specific cases. We saw in Section 3 that the RQMC sample sizes n grow at slow rates $\Theta(N^{1/3})$ or $\Theta(N^{1/4})$ as the budget N increases. Here we get a more detailed view of that phenomenon using the asymptotic formula for some specific variance rates. We can compute these confidence interval widths for very large values of N that would be impractical to simulate.

The digital nets of Sobol' [39] under the nested uniform scramble of [30] have a stratification property. The n values of x_{ij} for $i = 1, \ldots, n$ and any component $j = 1, \ldots, d$ have one value uniformly distributed in each interval $[(\ell-1)/n, \ell/n)$ for $\ell = 1, \ldots, n$ and those n values are mutually independent. This allows us to work out the finite sample RQMC variances for some illustrative examples. The random linear scramble of [25] combined with a digital shift attains the same variance as the nested uniform scramble, but it has some dependencies among the stratified sample values. In both cases, the d vectors $(x_{1j}, \ldots, x_{nj}) \in \mathbb{R}^n$ are mutually independent.

We only consider values of n that are powers of 2 which is a best practice when using Sobol' sequences [34]. The smooth error rate $\tilde{O}(n^{-3/2})$ for the RMSE is attainable along powers of 2 but not for general n. An elegant argument due to Sobol' [40] explains why. He notes that

$$|f(\boldsymbol{x}_{n+1}) - \mu| \leqslant (n+1)|\hat{\mu}_{n+1} - \mu| + n|\hat{\mu}_n - \mu|.$$

A rate of $o(n^{-1})$ would imply that $f(\boldsymbol{x}_{n+1})$ itself consistently estimates μ. For any sampling algorithm, that could only be true for a very minimal set of functions f. We also take N to be a power of 2.

For $d = 1$ and f of bounded variation in the usual sense, we get $\mathrm{var}(\hat{\mu}) \leqslant V_{\mathrm{HK}}(f)^2/n^2$. This holds because D_n^* for the stratified sample cannot be larger than $1/n$. For illustration we choose a function with $V_{\mathrm{HK}}(f) = 1$ for which we know exactly what the RQMC variance is for n a power of 2. That integrand is $f(x) = 1\{x < 1/3\}$. It is constant within $n-1$ of the intervals $[(\ell-1)/n, \ell/n)$. All of this function's variation happens within just one of those intervals. By inspecting the base 2 expansion $1/3 = 0.010101\cdots$ we can see that the remaining value of $f(x_\ell)$ is 1 or 0 and it equals 1 with a probability that is either $1/3$ if $\log_2(n)$ is an odd integer or is $2/3$ if $\log_2(n)$ is an even integer. It follows that $\mathrm{var}(\hat{\mu}_n) = 2/(9n^2)$.

Table 1 shows how the width minimizing value of n increases with N in the BVHK case with $f(x) = 1\{x < 1/3\}$. These are widths $W = 2H^{\mathrm{Ben}}(n, N/n)$ using equation (11), not half widths. Some are larger than 1, because the Bennett width does not reflect that we would intersect the intervals with $[0, 1]$. The increase in n is quite slow. By $N = 1024$ the best RQMC sample size is only $n = 8$ and this remains the optimal sample size up to $N = 4096$. The value $N^{2/3}W$ very quickly becomes constant. For non-integer θ (not shown here) the scaled widths fluctuate within a finite range instead of approaching a constant.

The result in Table 1 for $N = 1024$ is to choose $n = 8$ which violates the guidance that the best n is at most 7 when $N = 1024$ and $\alpha = 0.05$. The reason is that this table restricts to powers of 2, and $n = 8$ is better than $n = 4$.

Our second example is $f(x) = xe^{x-1}$. This function is bounded between 0 and 1. We chose it because it is smooth enough to have $\theta = 3$ while also not a polynomial or an antisymmetric function which would make the problem artificially easy for some methods. Because of the stratification, we

$\log_2(N)$	N	n	W	$N^{2/3}W$
0	1	1	5.02e+00	5.020114
4	16	2	7.60e−01	4.826379
7	128	4	1.90e−01	4.826379
10	1,024	8	4.75e−02	4.826379
13	8,192	16	1.19e−02	4.826379
16	65,536	32	2.97e−03	4.826379
19	524,288	64	7.42e−04	4.826379
22	4,194,304	128	1.86e−04	4.826379
25	33,554,432	256	4.64e−05	4.826379
28	268,435,456	512	1.16e−05	4.826379

Table 1: As N varies for $f(x) = 1\{x < 1/3\}$, this table shows the value of n that minimizes the width of Bennett's interval, along with the minimizing width W and $N^{2/3}W$. Each row corresponds to a value of N for which n has increased.

know that

$$\mathrm{var}(\hat{\mu}) = \frac{1}{n^2}\sum_{\ell=1}^{n}\mathrm{var}(f(x_\ell)) \tag{15}$$

where $x_\ell \sim \mathbb{U}[(\ell-1)/n, \ell/n]$ are independent. Also

$$\mathrm{var}(f(x_\ell)) = n\int_{(\ell-1)/n}^{\ell/n} f(x)^2\,\mathrm{d}x - \Bigl(n\int_{(\ell-1)/n}^{\ell/n} f(x)\,\mathrm{d}x\Bigr)^2 \tag{16}$$

which can be computed for finite n using the closed form antiderivatives of f and f^2. We know that as n increases through powers of 2

$$\mathrm{var}(\hat{\mu}) \asymp \frac{1}{12n^3}\int_0^1 f'(x)^2\,\mathrm{d}x = \frac{5-e^{-2}}{48n^3}$$

but the exact variances are somewhat different for $n \leqslant 8$.

Table 2 shows how the optimal sample sizes n grows with N for our smooth example $f(x) = xe^{x-1}$. An RQMC sample size of $n = 8$ becomes optimal for $N = 4{,}096$ and remains optimal up to $N = 32{,}768$. The values $N^{3/4}W$ settle down rapidly as N increases. Consistent with the results of Section 3, the values n grow at a slower rate with N than in the BVHK case while the widths W

$\log_2(N)$	N	n	W	$N^{3/4}W$
0	1	1	4.00e+00	4.003728
4	16	2	5.17e−01	4.138086
8	256	4	6.52e−02	4.175717
12	4,096	8	8.17e−03	4.185407
16	65,536	16	1.02e−03	4.187848
20	1,048,576	32	1.28e−04	4.188459
24	16,777,216	64	1.60e−05	4.188612
28	268,435,456	128	2.00e−06	4.188651

Table 2: As N varies for $f(x) = xe^{x-1}$, this table shows the value of n that minimizes the width of Bennett's interval, along with the minimizing width W and $N^{3/4}W$. Each row corresponds to a value of N for which n has increased.

decrease at a faster rate.

4.2 Ridge functions

While the betting intervals are asymptotically of the same width as we would get from Bennett's formula, here we investigate what choices of n and R give the narrowest interval widths for some finite N. These computations produce actual intervals that we then intersect with $[0,1]$. It is well known that smoothness and dimension affect the performance of RQMC methods. We choose to use ridge functions because that makes it straightforward to vary both the dimension and smoothness of our test functions.

Let Φ be the $\mathcal{N}(0,1)$ cumulative distribution function. Then for $\boldsymbol{x} \sim \mathbb{U}[0,1]^d$ the vector $\boldsymbol{z} = \Phi^{-1}(\boldsymbol{x})$ componentwise, has the $\mathcal{N}(0,I)$ distribution over $\mathbb{R}^d$. Then $\mathbf{1}^\mathsf{T}\boldsymbol{z}/\sqrt{d} \sim \mathcal{N}(0,1)$. Our ridge function

$$f(\boldsymbol{x}) = g\Bigl(\frac{\mathbf{1}^\mathsf{T}\Phi^{-1}(\boldsymbol{x})}{\sqrt{d}}\Bigr).$$

then has the same mean and variance and differentiability properties in any dimension d. We choose

$$
\begin{aligned}
g_{\mathrm{jump}}(v) &= 1\{v \geqslant 1\},\\
g_{\mathrm{kink}}(v) &= \frac{\min(\max(-2,v),1)+2}{3},\\
g_{\mathrm{smooth}}(v) &= \Phi(v+1), \quad \text{and}\\
g_{\mathrm{finance}}(v) &= \min\Bigl(1, \frac{\sqrt{\max(v+2,0)}}{2}\Bigr).
\end{aligned}
$$

All of the above choices give ridge functions bounded between 0 and 1. They span a very wide range of QMC regularity conditions. Using g_{jump} makes f discontinuous and it has infinite variation in the sense of Hardy and Krause for $d \geqslant 2$. Using g_{kink} makes f have infinite variation in the sense of Hardy and Krause for $d \geqslant 3$ and yet it has been seen to be a more 'QMC friendly' integrand than g_{jump} in high dimensions because it has an asymptotically bounded mean dimension [18], meaning that it is dominated by low dimensional interactions. It is continuous, has one weak derivative but is not differentiable everywhere. Using g_{smooth} makes f infinitely differentiable which is generally a very QMC friendly property. Using g_{finance} gives f a property similar to some integrands in financial option valuation, namely an unbounded derivative where $v = -2$.

For each of these four ridge functions, we considered budgets $N = 2^K$ for $K \in \{8, 10, 12, 14, 16\}$ with RQMC sample sizes $n = 2^k$ for $k \in \{0, 1, 2, 3, 4, 5, 6\}$ in dimensions $d \in \{1, 2, 4, 16\}$. All of those conditions were repeated 20 times independently. In each repetition we computed the width of the HBCI, the width of the EBCI, and the width of the CLT window.

The HBCI were narrower than the corresponding EBCI about 92.1% of the time. The CLT intervals were narrower than the corresponding HBCI 99.6% of the time. The few times that the CLT intervals were wider were predominantly from cases with $n = 1$. The EBCI were never narrower than the CLT intervals. In some cases, the EBCI had width 1. This only happened for $N = 256$ and $n = 32$ or 64. Those cases have small replicate numbers $R = 8$ and $R = 4$, respectively.

The HBCI are most interesting because they have non-asymptotic coverage and are usually narrower than the EBCI. We made an extensive graphical exploration of their width versus n (for every combination of dimension, ridge function and N). We similarly explored width versus budget N and width versus dimension d, in each case at all levels of all of the other variables.

It was evident that the dimension d made hardly any difference to the width of the confidence intervals. An additive model using factors equal to the dimension d, the RQMC sample size n, the function f and the budget N explained 94.09% of the variance in the logarithm of the HBCI width. That same model without dimension explained 94.05% of this variance. It was surprising to see no dimension effect for the HBCI width in our ridge functions. We knew that the mean and variance of $f(\boldsymbol{x})$ for any of those ridge functions does not depend on d but that does not imply that averages over n RQMC points would be invariant to d. Because d made no material difference to the widths, we pooled information from all dimensions d.

It was clear graphically, and not surprising, that increasing N by a factor of 4 always narrows the HBCI width by a noticeable amount. The interesting non-monotone patterns were from the way that the width depends on n with the other variables fixed. Figure 1 shows the mean widths of HBCI versus n for all four ridge functions and all five budgets N.

We see in Figure 1 some things that align very well with our analysis and some that were not anticipated. The optimal samples sizes n are very small and they grow slowly with the budget N as expected. The least smooth ridge function g_{jump} has the fastest growing n as expected. The interval widths for g_{finance} and g_{smooth} were very close to each other despite having very different smoothness and then they had identical optimal n at each N. For $N = 2^{10}$ the optimal n for g_{kink} is larger than the one for g_{smooth} as expected but at $N = 2^{16}$, it is g_{smooth} that has the larger optimal n.

Our expectations were based on asymptotic considerations. First, we used the fact that the HBCI widths asymptotically approach the optimal EBCI widths from Bennett's inequality. Second, we supposed that the RQMC variance is asymptotic to $\sigma_0^2 n^{-\theta}$ apart from logarithmic factors. We also used the finite sample Bennett widths from (13) because the asymptotic expression $\sigma_n\sqrt{2\log(2/\alpha)/R}$ does not have a meaningful optimum. The empirical results strongly confirm our asymptotic findings that small n is best overall, but among such small n, it is less clear how to choose n based on smoothness of f.

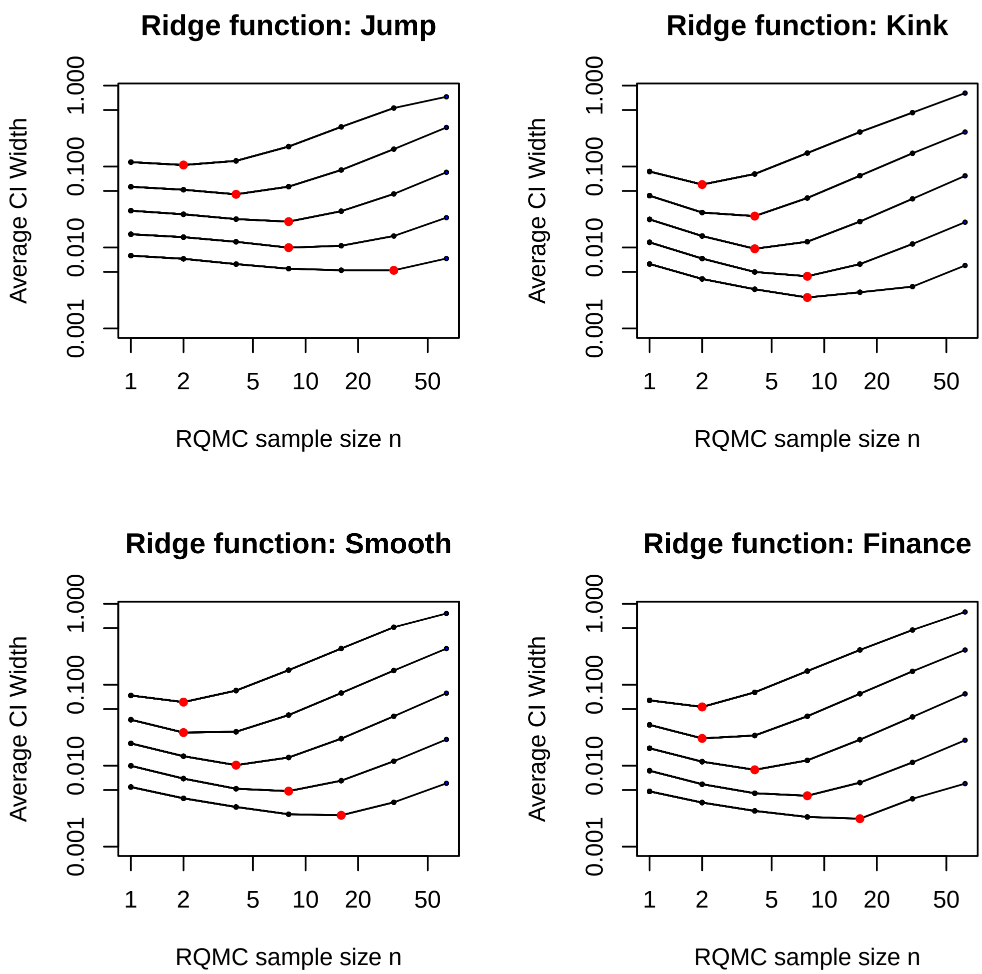

Figure 1: Mean confidence interval width for the HBCI as a function of the RQMC sample size n. There is one panel for each of the 4 ridge functions. Within a panel the curves are for $N = 2^{2r}$ for $r = 4, 5, 6, 7, 8$ from top to bottom. The smallest mean width for each N is marked with a larger circle. The means are taken over all 20 replicates and all 4 dimensions used.

4.3 Functions with known RQMC variance

Here we return to the functions from Section 4.1 with known non-asymptotic RQMC variance. They allow us to directly compare the widths of the HBCI to the value we get from Bennett's formula. We know that the HBCI have lengths that asymptotically match Bennett's formula, but we also

	$1\{x < 1/3\}$		xe^{x-1}	
N	n_*	n_{opt}	n_*	n_{opt}
256	4	4	4	2
1,024	8	4	4	4
4,096	8	8	8	4
16,384	16	16	8	8
65,536	32	16	16	8

Table 3: For the given budgets N, we show n_* as the minimizer of the Bennett width (13) over $n = 2^k$ for integer k and n_{opt} is the value of n that minimized the average width of the HBCI for the given N. These are shown for functions $f(x) = 1\{x < 1/3\}$ and $f(x) = xe^{x-1}$.

want to study how those lengths approach Bennett's formula in some examples.

Using the known variances, we can compute the half width from Bennett's inequality (11) for our σ_n^2 and $R = N/n$~~double it~~, double it, and then divide the average HBCI width of 20 independent trials by this value. The results are shown in Figure 2. For our two example functions we see that as $R = N/n$ increases (so n decreases for fixed N) the ratio of widths decreases and approaches the theoretically expected value of 1. At the smallest R (largest n) the HBCI were two or three times as wide as the Bennett intervals for the discontinuous function. They were somewhat wider for the smooth function.

The HBCI are computed without knowledge of the exact variance σ_n^2 that the Bennett interval uses. They are instead based on estimates of that variance and they must therefore allow for uncertainty in the variance estimate. It is then not surprising that the resulting intervals are wider than the Bennett ones. We see in Table 3 that the HBCI's minimum widths were not found at larger n than the minimizers of the Bennett intervals. They were either equal to or half as large as n_*. We also see that empirical best values of n are no larger for the smooth function than for the discontinuous one, as expected.

5 Discussion

In favorable cases the RQMC variance is $\tilde{O}(n^{-3})$ and then using a fixed number R of replicates of point sets with $n \to \infty$ we can estimate the integral μ with a standard deviation of $\tilde{O}(N^{-3/2})$ for

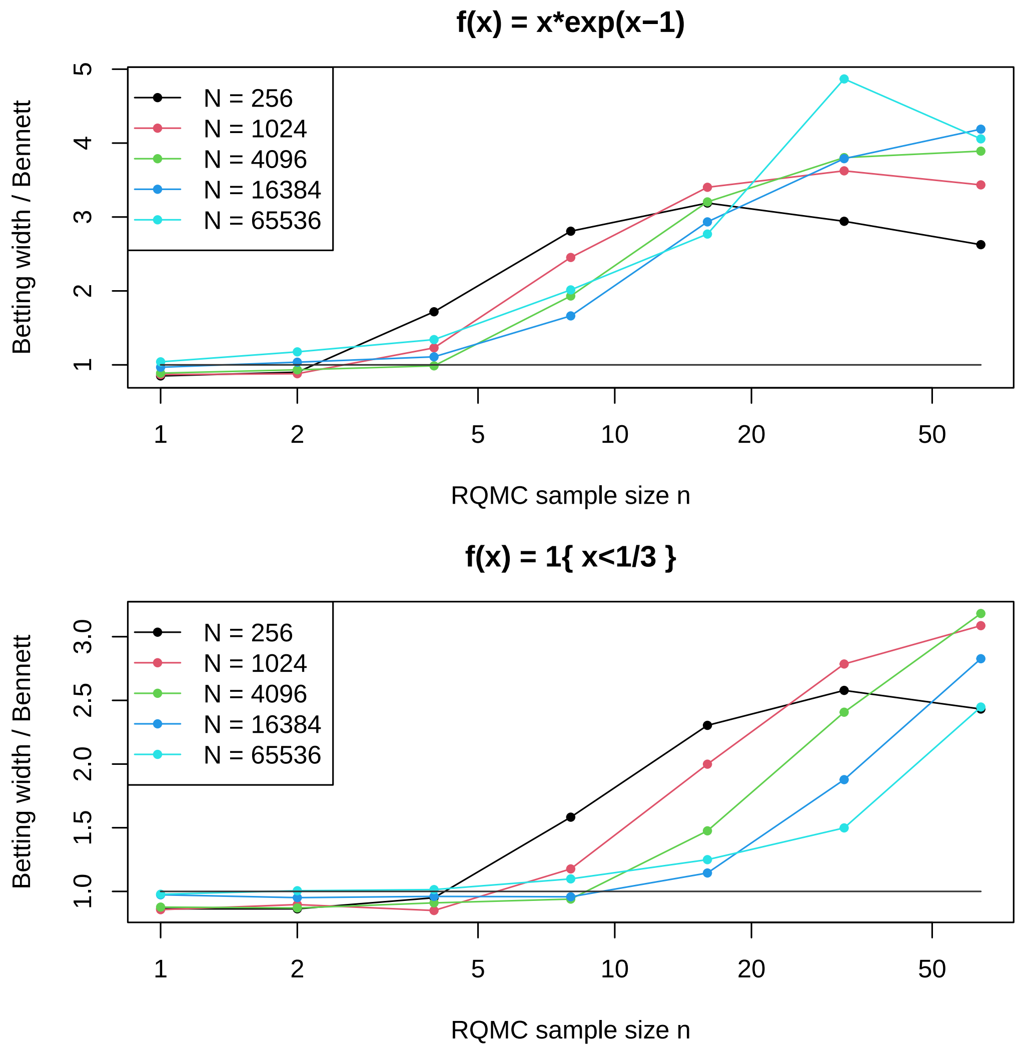

Figure 2: Ratio of mean HBCI width to width from Bennett's inequality as a function of the RQMC sample size n for the given budgets N. The means for the HBCI are taken over the 20 independent replicates.

$N = nR$ along with an unbiased estimate of the variance of that estimate. It remains very difficult to get a confidence interval of width $\tilde{O}(N^{-3/2})$ [35]. Using Student's t with a small R gave good results in an extensive simulation by [23] but that success is not yet theoretically understood. For an integrand subject to known bounds, we can get a non-asymptotic confidence interval by using

R replicates of n RQMC points using a predictable plug-in empirical Berstein confidence interval. There, if $\sigma_n^2 = \Theta(n^{-\theta})$ then the optimal n is $\Theta(N^{1/(\theta+1)})$ and the resulting confidence intervals have width $\Theta(N^{-\theta/(\theta+1)})$. While it remains to find a way to choose the optimal n empirically, we have strong theoretical guidance from equation (14) that we can interpret as an upper bound on n for the oracle which is then a reasonable upper bound for HBCIs which have to estimate the RQMC variance. When n must be a small power of 2, we can rule out many suboptimal choices.

We have noted several times that there is an asymptotic equivalence between the Bennett interval widths and those of the EBCI and HBCI intervals. In our setting that equivalence is as $R \to \infty$ for fixed n. In the RQMC context, the best n is quite small. Then for large N it is not surprising that an asymptote as $R \to \infty$ is very predictive of our results.

It also remains to find a good way to use RQMC in confidence sequences as opposed to confidence intervals. We might set up R independent infinite sequences of RQMC points and then stop them at the first n where they provide a $1 - \alpha$ interval narrower than some ϵ. In a task like that it makes sense to only use $n = 2^k$ for integers k. That is partly because powers of two are good for scrambled Sobol' points, but also because in RQMC it is generally advisable to study sample sizes n that grow geometrically not arithmetically [40]. The idea is that if n is not large enough to get a good answer, then $n + 1$ is unlikely to be much of an improvement. A confidence sequence based on geometrically growing n would not have to be as wide as one for arithmetically growing n.

For scrambled Sobol' points it is strongly advisable to take $n = 2^k$ because those are the best sample sizes and they can even give a better convergence rate. The Halton sequence [11] does not have strongly superior sample sizes and so n need not be a power of two for it. The two scrambles we considered for Sobol' points can also be applied to the Halton sequence getting $\mathrm{var}(\hat{\mu}) = o(1/n)$ without requiring special sample sizes n [37]. There are however no published non-trivial settings where scrambled Halton points can obtain $\theta > 2$.

Acknowledgments

We thank Aaditya Ramdas and Ian Waudby-Smith for helpful discussions. ABO was supported by the U.S. National Science Foundation grant DMS-2152780. FJH and AGS were supported by U.S. National Science Foundation grant DMS-2316011. AGS was supported by the U.S. DOE Office of Science Graduate Student Research Program.

References

[1] J.-Y. Audibert, R. Munos, and C. Szepesvári. Tuning bandit algorithms in stochastic environments. In *International conference on algorithmic learning theory*, pages 150–165. Springer, 2007.

[2] R. R. Bahadur and L. J. Savage. The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Stat.*, 27:1115–1122, 1956.

[3] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

[4] S. Boucheron, G. Lugosi, and O. Bousquet. Concentration inequalities. In *Summer school on machine learning*, pages 208–240. Springer, 2003.

[5] S.-C. T. Choi, Y. Ding, F. J. Hickernell, L. Jiang, Ll. A. Jiménez Rugama, X. Tong, Y. Zhang, and X. Zhou. GAIL—Guaranteed Automatic Integration Library in MATLAB: Documentation for version 2.1. *arXiv:1503.06544*, 2015.

[6] S.-C. T. Choi, F. J. Hickernell, R. Jagadeeswaran, M. McCourt, and A. Sorokin. QMCPy: A quasi-Monte Carlo Python library (versions 1–1.6), 2025.

[7] R. Cools and D. Nuyens, editors. *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014*, volume 163 of *Springer Proceedings in Mathematics and Statistics*. Springer-Verlag, Berlin, 2016.

[8] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration.* Academic Press, Orlando, Florida, 1984.

[9] J. Dick. Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. *Ann. Statist.*, 39:1372–1398, 2011.

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

[11] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, 2:84–90, 1960.

[12] F. J. Hickernell. Koksma-Hlawka inequality. In S. Kotz, N. L. Johnson, C. B. Read, N. Balakrishnan, and B. Vidakovic, editors, *Encyclopedia of Statistical Sciences*, volume 6, pages 3862–3867. John Wiley & Sons, Hoboken, NJ, second edition, 2006.

[13] F. J. Hickernell, L. Jiang, Y. Liu, and A. B. Owen. Guaranteed conservative fixed width confidence intervals via Monte Carlo sampling. In J. Dick, F. Y. Kuo, G. W. Peters, and I. H. Sloan, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2012*, volume 65 of *Springer Proceedings in Mathematics and Statistics*, pages 105–128. Springer-Verlag, Berlin, 2013.

[14] F. J. Hickernell and Ll. A. Jiménez Rugama. Reliable adaptive cubature using digital sequences. In Cools and Nuyens [7], pages 367–383. arXiv:1410.8615 [math.NA].

[15] F. J. Hickernell, Ll. A. Jiménez Rugama, and D. Li. Adaptive quasi-Monte Carlo methods for cubature. In J. Dick, F. Y. Kuo, and H. Woźniakowski, editors, *Contemporary Computational Mathematics — a celebration of the 80th birthday of Ian Sloan*, pages 597–619. Springer-Verlag, 2018.

[16] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[17] S. R. Howard, I. Waudby-Smith, and A. Ramdas. ConfSeq: Software for confidence sequences and uniform boundaries. `https://github.com/gostevehoward/confseq`, 2021.

[18] C. R. Hoyt and A. B. Owen. Mean dimension of ridge functions. *SIAM Journal on Numerical Analysis*, 58(2):1195–1216, 2020.

[19] R. Jagadeeswaran and F. J. Hickernell. Fast automatic Bayesian cubature using lattice sampling. *Stat. Comput.*, 29:1215–1229, 2019.

[20] R. Jagadeeswaran and F. J. Hickernell. Fast automatic Bayesian cubature using Sobol' sampling. In Z. Botev, A. Keller, C. Lemieux, and B. Tuffin, editors, *Advances in Modeling and Simulation: Festschrift in Honour of Pierre L'Ecuyer*, pages 301–318. Springer, Cham, 2022.

[21] Ll. A. Jiménez Rugama and F. J. Hickernell. Adaptive multidimensional integration based on rank-1 lattices. In Cools and Nuyens [7], pages 407–422. arXiv:1411.1966.

[22] P. L'Ecuyer and C. Lemieux. Recent advances in randomized quasi-Monte Carlo methods. In M. Dror, P. L'Ecuyer, and F. Szidarovszki, editors, *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, pages 419–474. Kluwer Academic Publishers, Boston, 2002.

[23] P. L'Ecuyer, M. K. Nakayama, A. B. Owen, and B. Tuffin. Confidence intervals for randomized quasi-Monte Carlo estimators. In *WSC '23: Proceedings of the Winter Simulation Conference*, pages 445–456, 2024.

[24] W.-L. Loh. On the asymptotic distribution of scrambled net quadrature. *Ann. Stat.*, 31:1282–1324, 2003.

[25] J. Matoušek. On the L_2-discrepancy for anchored boxes. *J. Complexity*, 14:527–556, 1998.

[26] A. Maurer and M. Pontil. Empirical Bernstein bounds and sample variance penalization. In *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)*, pages 1–9, 2009.

[27] M. K. Nakayama and B. Tuffin. Sufficient conditions for central limit theorems and confidence intervals for randomized quasi-Monte Carlo methods. *ACM Transactions on Modeling and Computer Simulation*, 34(3):1–38, 2024.

[28] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 1992.

[29] D. Nuyens and R. Cools. Higher order quasi-Monte Carlo methods: A comparison. In *AIP Conference Proceedings*, volume 1281, pages 553–557. American Institute of Physics, 2010.

[30] A. B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pages 299–317. Springer-Verlag, New York, 1995.

[31] A. B. Owen. Scrambled net variance for integrals of smooth functions. *Ann. Stat.*, 25:1541–1562, 1997.

[32] A. B. Owen. Multidimensional variation for quasi-Monte Carlo. In J. Fan and G. Li, editors, *Contemporary Multivariate Analysis and Experimental Design*, volume 2 of *Series in Biostatistics*, pages 49–74, Singapore, 2005. World Scientific.

[33] A. B. Owen. Local antithetic sampling with scrambled nets. *Ann. Statist.*, 36:2319–2343, 2008.

[34] A. B. Owen. On dropping the first Sobol' point. In A. Keller, editor, *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Oxford, England, August 2020*, Springer Proceedings in Mathematics and Statistics, pages 71–86. Springer, Cham, 2022.

[35] A. B. Owen. Error estimation for quasi-Monte Carlo. In B. Feng and C. Lemieux, editors, *Monte Carlo and Quasi-Monte Carlo 2024*. Springer, 2026. To appear.

[36] A. B. Owen and Z. Pan. Where are the logs? In Z. Botev, A. Keller, C. Lemieux, and B. Tuffin, editors, *Advances in Modeling and Simulation: festschrift for Pierre L'Ecuyer*. Springer, Cham, 2022.

[37] A. B. Owen and Z. Pan. Gain coefficients for scrambled Halton points. *SIAM Journal on Numerical Analysis*, 62(3):1021–1038, 2024.

[38] Ch. Schlier. A practitioner's view on QMC integration. Technical report, Universität Freiburg, Fakultät für Physik, 2002.

[39] I. M. Sobol'. The distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Comput. Math. and Math. Phys.*, 7:86–112, 1967.

[40] I. M. Sobol. On quasi-Monte Carlo integrations. *Mathematics and computers in simulation*, 47(2-5):103–112, 1998.

[41] A. G. Sorokin and R. Jagadeeswaran. On bounding and approximating functions of multiple expectations using quasi-Monte Carlo. In A. Hinrichs, P. Kritzer, and F. Pillichshammer, editors, *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Linz, Austria, July 2022*, Springer Proceedings in Mathematics and Statistics, pages 583–589. Springer, Cham, 2024.

[42] J. Ville. *Étude critique de la notion de collectif.* PhD thesis, Université de Paris, 1939.

[43] D. D. Wackerly, W. Mendenhall III, and R. L. Scheaffer. *Mathematical Statistics with Applications.* Thomson Brooks/Cole, Belmont, CA, seventh edition, 2008.

[44] I. Waubdy-Smith and A. Ramdas. Estimating means of bounded random variables by betting. *J. Roy. Statist. Soc. B*, 86:1–27, 2024.

[45] I. Waudby-Smith. Betting paper simulations. `https://github.com/WannabeSmith/betting-paper-simulations`, 2025.

List of Figures

List of Tables