A strong law of large numbers for scrambled net integration

Art B. Owen
Stanford University
February 2020

Abstract
This article provides a strong law of large numbers for integration on digital nets randomized by a nested uniform scramble. This strong law requires a square integrable integrand and it holds along a sequence of sample sizes of the form \( n = mb^k \) for \( m = 1, \ldots, M \) and \( k \geq 0 \) and a base \( b \geq 2 \). Previously known results implied a strong law only for Riemann integrable functions with a weak law for square integrable ones.

1 Introduction
Randomized quasi-Monte Carlo (RQMC) methods support uncertainty quantification of quasi-Monte Carlo (QMC) estimates of an integral. When the integrand of interest has bounded variation in the sense of Hardy and Krause [14, 21], then these RQMC methods satisfy both a strong and a weak law of large numbers, respectively SLLN and WLLN. There are however many important cases where the integrand is not of bounded variation. These include integrable singularities [43, 3, 15, 37] and integrands with kinks and/or jumps [11, 12, 13, 16]. In those cases, we can easily get a WLLN, if the integrand is in \( L^2 \). This article proves an SLLN for scrambled net quadrature with integrands in \( L^2 \).

To keep this paper at a manageable length, the relevant properties of QMC and RQMC methods are presented but the details of their constructions are omitted. For QMC, see the survey [7] or the monographs [29, 8]. For a survey of RQMC methods see [24].

An outline of this paper is as follows. Section 2 introduces notation, reviews the SLLN and WLLN for Monte Carlo and some quasi-Monte Carlo counterparts. It includes a lemma to show that functions of bounded variation in the sense of Hardy and Krause must also be Riemann integrable. That is either a new result or one hard to find in the literature. Section 3 defines digital nets and describes some properties of randomly scrambled digital nets. Section 4 has the main result. It is a strong law of large numbers for scrambled net sampling. The integrand is assumed to be in \( L^2 \). The strong law holds along a set of sample sizes of the form \( n = mb^k \) for \( m = 1, \ldots, M \) and \( k \geq 0 \) and a base \( b \geq 2 \). Previously known results implied a strong law only for Riemann integrable functions with a weak law for square integrable ones.
sizes with a geometric spacing. More precisely, the given sample sizes are of the form \( mb^k \) for all \( m = 1, \ldots, M \) and all integers \( k \geq 0 \) where \( b \geq 2 \) is the base used to describe the underlying digital nets. This is not a practical hindrance as those include the best sample sizes for digital nets and sequences. For instance, digital nets are designed for \( n \) equal to a power of \( b \geq 2 \). Section 5 provides some additional context and discussion including randomly shifted lattice versions of RQMC.

2 Notation and background

We are given an extended real function \( f \in L^1[0,1]^d \) for some dimension \( d \geq 1 \). The Monte Carlo (MC) method takes independent \( x_i \sim U[0,1]^d \) and estimates 
\[
\mu = \int_{[0,1]^d} f(x) \, dx \text{ by } \hat{\mu}_n = \hat{\mu}_{n}^{MC} = \frac{1}{n} \sum_{i=1}^{n} f(x_i).
\]
Many problems that do not originate as integrals over \([0,1]^d\) have such a representation using transformations to generate non-uniformly distributed random variables over the cube and other spaces [6]. We suppose that those transformations are subsumed into \( f \).

The WLLN is that for any \( \epsilon > 0 \),
\[
\lim_{n \to \infty} \Pr(|\hat{\mu}_n^{MC} - \mu| > \epsilon) = 0.
\]
The SLLN is that
\[
\Pr\left( \lim_{n \to \infty} \hat{\mu}_n^{MC} = \mu \right) = 1
\]
which we may write as \( \Pr(\lim \sup_{n \to \infty} |\hat{\mu}_n^{MC} - \mu| > \epsilon) = 0 \) to parallel the WLLN.

In QMC sampling, the \( x_i \) are constructed so that the discrete distribution placing probability \( 1/n \) on each of \( x_1, \ldots, x_n \) (with repeated points counted multiple times) is close to the continuous uniform distribution on \([0,1]^d\). There are various ways, called discrepancies [4], to quantify the distance between these discrete and continuous measures. The most widely used discrepancy is
\[
D_n^* = D_n^*(x_1, \ldots, x_n) = \sup_{a \in [0,1]^d} \left| \frac{1}{n} \sum_{i=1}^{n} 1\{x_i \in [0,a)\} - \prod_{j=1}^{d} a_j \right|
\]
where \( [0,a) = \{x \in [0,1]^d \mid 0 \leq x_j < a_j, j = 1, \ldots, d\} \).

Because QMC is deterministic, it has no analogue to the WLLN (1). There is an SLLN analogue as follows. Let \( \hat{\mu}_n^{QMC} = (1/n) \sum_{i=1}^{n} f(x_i) \) where now the points \( x_i \) have been chosen to have small discrepancy. If \( f \) is Riemann integrable and \( D_n^* \to 0 \) then [22, p. 3]
\[
\lim_{n \to \infty} \hat{\mu}_n^{QMC} = \mu
\]
providing the QMC version of the SLLN (2). There is a converse, where if \( |\hat{\mu}_n - \mu| \to 0 \) whenever \( D_n^* \to 0 \), then \( f \) must be Riemann integrable. See
the references and discussion in [27]. Riemann integrable \( f \) must also be in \( L^\infty[0,1]^d \).

A better known result about QMC is the Koksma-Hlawka inequality

\[
|\hat{\mu}_n^{\text{QMC}} - \mu| \leq D_n^* \times V_{\text{HK}}(f) \tag{4}
\]

(see [17]) where \( V_{\text{HK}}(f) \) is the total variation of \( f \) in the sense of Hardy and Krause. If \( V_{\text{HK}}(f) < \infty \), then we write \( f \in \text{BVHK}(0,1]^d \). We don’t need bounded variation to get an SLLN for QMC. It does however give some information on the rate of convergence because some QMC constructions provide infinite sequences \( x_i \) whose initial subsequences satisfy

\[
D_n^*(x_1, \ldots, x_n) = O\left(\frac{\log(n)^d}{n}\right),
\]

and then \( |\hat{\mu}_n^{\text{QMC}} - \mu| = O(n^{-1+\epsilon}) \) for any \( \epsilon > 0 \).

The counterpart in MC to the Koksma-Hlawka inequality is that

\[
\mathbb{E}((\hat{\mu}_n^{\text{MC}} - \mu)^2)^{1/2} = n^{-1/2} \sigma(f) \tag{5}
\]

when, for \( x \sim \mathcal{U}[0,1]^d \) we have \( \sigma^2 = \sigma^2(f) = \mathbb{E}((f(x) - \mu)^2) < \infty \). Where the rate for QMC comes after restricting from Riemann integrability to bounded variation, the rate for MC comes about after restricting from \( L^1 \) to \( L^2 \). The MC counterpart (5) is exact while the QMC version (4) is an upper bound.

A Riemann integrable function is not necessarily in BVHK. For instance \( f(x) = 1\{\sum_{j=1}^d x_j \leq 1\} \) is Riemann integrable but, for \( d \geq 2 \), it is not in \( \text{BVHK} \) [36]. A function in BVHK is necessarily Riemann integrable. This result is hard to find in the literature. It must have been known to Hardy, Krause, Hobson and others over a century ago, at least for \( d = 2 \), which earlier work emphasized. Here is a proof.

**Lemma 1.** If \( f \in \text{BVHK}[0,1]^d \), then \( f \) is also Riemann integrable.

**Proof.** If \( f \) is in \( \text{BVHK}[0,1]^d \) then \( f(x) = f(0) + f_+(x) - f_-(x) \) where \( f_\pm \) are uniquely determined completely monotone functions on \([0,1]^d\) with \( f_\pm(0) = 0 \) [1, Theorem 2]. Completely monotone functions are, a fortiori, monotone. Now both \( f_\pm \) are bounded monotone functions on \([0,1]^d\). They are then Riemann integrable by the corollary in [23]. \(\square\)

While QMC has a superior convergence rate to MC for \( f \in \text{BVHK} \), MC has an advantage over QMC in that \( \sigma^2/n \) is simple to estimate from independent replicates, while \( D_n^* \) is extremely difficult to estimate [9] and \( V_{\text{HK}}(f) \) much harder to estimate than \( \mu \). In a setting where attaining accuracy is important it will also be important to estimate the attained accuracy. Independent replication of RQMC estimates allows variance estimation for them [32, 24].

In RQMC, one starts with points \( a_1, \ldots, a_n \in [0,1]^d \) having a small discrepancy and randomizes them producing points \( x_1, \ldots, x_n \). These points satisfy the following conditions: individually \( x_i \sim \mathcal{U}[0,1]^d \), and collectively, \( x_1, \ldots, x_n \) have low discrepancy. The RQMC estimate of \( \mu \) is \( \hat{\mu}_n^{\text{RQMC}} = (1/n) \sum_{i=1}^n f(x_i) \). From the uniformity of the points \( x_i \) we find that \( \mathbb{E}(\hat{\mu}_n^{\text{RQMC}}) = \mu \). Theorem 1 in Section 4 gives an SLLN for RQMC when \( f \) is Riemann integrable.
3 Scrambled nets and sequences

In this section, we focus on scrambled versions of digital nets and sequences. Let $b \geq 2$ be an integer base. Let $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$ and $c = (c_1, \ldots, c_d)$ where $c_j \in \{0, 1, \ldots, b^{k_j} - 1\}$. Then

$$E(k, c) = \prod_{j=1}^{d} \left\lfloor \frac{c_j}{b_j}, \frac{c_j + 1}{b_j} \right\rfloor$$

is an elementary interval in base $b$. It has volume $b^{-|k|}$ where $|k| = \|k\|_1$.

**Definition 1.** For integers $m \geq t \geq 0$, $b \geq 2$ and $d \geq 1$, the points $x_1, \ldots, x_n \in [0,1)^d$ for $n = b^m$ are a $(t, m, d)$-net in base $b$ if

$$\sum_{i=1}^{n} 1\{x_i \in E(k, c)\} = b^{m-|k|}$$

holds for every elementary interval $E(k, c)$ from (6) with $|k| \leq m - t$.

An elementary interval of volume $b^{-|k|}$ should ideally contain $nb^{-|k|} = b^{m-|k|}$ points from $x_1, \ldots, x_n$. In a digital net, every elementary interval that should ideally contain $b^t$ of the points does so. For any given $b$, $m$ and $d$, smaller $t$ imply finer equidistribution. It is not always possible to attain $t = 0$.

**Definition 2.** For integers $t \geq 0$, $b \geq 2$ and $d \geq 1$, the points $x_i \in [0,1)^d$ for $i \geq 1$ are a $(t, m, d)$-sequence in base $b$ if every subsequence of the form $x_{(r-1)b^m+1}, \ldots, x_{rb^m}$ for integers $m \geq t$ and $r \geq 1$ is a $(t, m, d)$-net in base $b$.

The best known values of $t$ for nets and sequences are recorded in the online resource MinT described in [40], which also includes lower bounds. The Sobol’ sequences of [42] are $(t, d)$-sequences in base $b = 2$. There are newer versions of Sobol’s sequence with improved ‘direction numbers’ in [20, 45]. The Faure sequences [10] have $t = 0$ but require that the base be a prime number $b \geq d$. Faure’s construction was generalized to prime powers $b \geq d$ [28]. The best presently attainable values of $t$ for base $b = 2$ are in the Niederreiter-Xing sequences of [30, 31].

Randomizations of digital nets and sequences operate by applying certain random permutations to their base $b$ expansions. For details see the survey in [35]. We will consider the ‘nested uniform’ scramble from [32].

If $a_1, \ldots, a_n$ is a $(t, m, d)$-net in base $b$ then after applying a nested uniform scramble, the resulting points $x_1, \ldots, x_n$ are a $(t, m, d)$-net in base $b$ with probability one [32]. If $a_i$ for $i \geq 1$ are a $(t, d)$-sequence in base $b$ then after applying a nested uniform scramble, the resulting points $x_i$ for $i \geq 1$ are a $(t, d)$-sequence in base $b$ with probability one [32].

If $f \in L^2 : [0,1]^d$ and $\hat{\mu}^{\text{RQMC}}_n$ is based on a nested uniform scramble of a $(t, d)$-sequence in base $b$ with sample sizes $n = b^k$ for integers $k \geq m$, then $E((\hat{\mu}^{\text{RQMC}}_n - \mu)^2) = o(1/n)$ as $n \to \infty$. For smooth enough $f$, $E((\hat{\mu}^{\text{RQMC}}_n - \mu)^2) = O(n^{-3+\epsilon})$ for any $\epsilon > 0$ [33, 38].
The main result that we will use is as follows. Let \( f \in L^2[0,1]^d \) and write \( \sigma^2 \) for the variance of \( f(x) \) when \( x \sim U[0,1]^d \). Then for a \((t,m,d)\)-net in base \( b \) scrambled as in [32] we have

\[
E((\hat{\mu}_{RQMC}^n - \mu)^2) \leq \frac{\Gamma \sigma^2}{n}
\]

for some \( \Gamma < \infty \) [34, Theorem 1]. That is, the RQMC estimate for these scrambled nets cannot have more than \( \Gamma \) times the mean squared error that MC has. The value of \( \Gamma \) is found using some conservative upper bounds. We can use \( \Gamma = b^t[(b+1)/(b-1)]^s \). If \( t = 0 \) then we can take \( \Gamma = [b/(b-1)]^s \), and for \( d = 1 \) we can take \( \Gamma = b^t \). The quantity \( \Gamma \) arises as an upper bound on an infinite set of ‘gain coefficients’ relating the RQMC variance to the MC variance for parts of a basis expansion of \( f \).

4 RQMC laws of large numbers

If \( f \in BVHK[0,1]^d \), then there is an SLLN for RQMC from the Koksma-Hlawka inequality (4) because the RQMC points have \( \Pr(\lim_{n \to \infty} D_n^*(x_1, \ldots, x_n) = 0) = 1 \). The same argument handles the more general case of Riemann integrable \( f \).

**Theorem 1.** Let \( f : [0,1]^d \to \mathbb{R} \) be Riemann integrable. Let \( x_i \) for \( i \geq 1 \) be RQMC points with \( \Pr(\lim_{n \to \infty} D_n^*(x_1, \ldots, x_n) = 0) = 1 \). Then

\[
\Pr\left(\lim_{n \to \infty} \hat{\mu}_{RQMC}^n = \mu\right) = 1.
\]

**Proof.** On the event \( D_n^*(x_1, \ldots, x_n) \to 0 \), which has probability one, we have \( \hat{\mu}_{RQMC}^n \to \mu \) by (3). \( \square \)

The usual results for RQMC show that \( E((\hat{\mu}_{RQMC}^n - \mu)^2) \to 0 \) as \( n \to \infty \) for \( f \in L^2[0,1]^d \). From that a WLLN follows by Chebychev’s inequality.

An SLLN is more complicated. We will provide one for points taken from a scrambled digital sequence using a subset of sample sizes. For integers \( b \geq 2 \) and \( M \geq 1 \), let

\[ N = N(b,M) = \{mb^k \mid m \in \{1, \ldots, M\}, k \in \mathbb{N} \}. \]

Denote the unique elements of \( N \) by \( n_\ell \) for \( \ell \in \mathbb{N} \) with \( n_1 < n_2 < \ldots \). The \( m \) used in sample sizes should not be confused with the \( m \) used in the definition of nets.

Restricting the sample sizes is not a serious impediment to using RQMC. It is actually a best practice when the point sets are designed for sample sizes equal to a power of a prime number. Sobol’ [44] recommends considering sample sizes in a geometric progression such as \( n_\ell = 2^\ell \), not an arithmetic one. One of his points is that when errors smaller than \( o(1/n) \) are possible they cannot be expected to occur at consecutive sample sizes \( n \) and \( n+1 \) for then the \( n+1 \)st
point on its own would be unrealistically good. See [39] for more details on how rate optimal sample sizes for equal weight rules must be geometrically spaced. The empirical results for scrambled Faure sequences in [32] were noticeably better for sample sizes \( n = b^m \).

**Theorem 2.** Let \( x_1, x_2, \ldots \) be a \((t,d)\)-sequence in base \( b \), with gain coefficients no larger than \( \Gamma < \infty \) and randomized as in [32]. Let \( f \in L^2[0,1]^d \) with 
\[
\int_{[0,1]^d} f(x) \, dx = \mu \quad \text{and} \quad \int_{[0,1]^d} (f(x) - \mu)^2 \, dx = \sigma^2.
\]
Then
\[
\Pr\left( \lim_{\ell \to \infty} \hat{\mu}_{RQMC}^{n_\ell} = \mu \right) = 1,
\]
for any set of sample sizes \( n_\ell \) of the form \( N(b,M) \).

**Proof.** For \( 1 \leq r \leq M \) and \( k \geq 0 \) let \( \hat{\mu}_{r,k} = b^{-k} \sum_{i=(r-1)b^k+1}^{rb^k} f(x_i) \). For \( n_\ell = mb^k \) we have
\[
\hat{\mu}_{n_\ell} = \frac{1}{m} \sum_{r=1}^{m} \hat{\mu}_{r,k}.
\]
Because RQMC is unbiased, we have \( \mathbb{E}(\hat{\mu}_{n_\ell}) = \mu \). Next, define
\[
\tau_{r,k}^2 = \text{var}(\hat{\mu}_{r,k}) \leq \frac{\Gamma \sigma^2}{b^k}.
\]
Then because the correlations among \( \hat{\mu}_{r,k} \) cannot be larger than one,
\[
\text{var}(\hat{\mu}_{n_\ell}) \leq \frac{1}{m^2} \sum_{r=1}^{m} \sum_{s=1}^{m} \frac{\Gamma \sigma^2}{b^k} = \frac{m \Gamma \sigma^2}{b^k} = \frac{m \Gamma \sigma^2}{n_\ell}.
\]
For any \( \epsilon > 0 \), Chebychev’s inequality yields
\[
\Pr(|\hat{\mu}_{n_\ell} - \mu| > \epsilon) \leq \frac{m \Gamma \sigma^2}{\epsilon^2 n_\ell}.
\]
Then
\[
\sum_{\ell=1}^{\infty} \Pr(|\hat{\mu}_{n_\ell} - \mu| > \epsilon) \leq \sum_{k=0}^{\infty} \frac{M \Gamma \sigma^2}{\epsilon^2 n_\ell} \leq \sum_{k=0}^{M} \sum_{m=1}^{b^k} \frac{m \Gamma \sigma^2}{\epsilon^2 (mb^k)} = \frac{b}{b-1} M \Gamma \sigma^2.
\]
The second inequality above is not necessarily an equality because summing over \( k \) and \( m \) may count some \( n_\ell \) more than once. Because that sum is finite, \( \Pr(|\hat{\mu}_\ell - \mu| > \epsilon \text{ infinitely many } \ell) = 0 \) by the Borel-Cantelli lemma. Therefore \( \Pr(\hat{\mu}_{n_\ell} \to \mu) = 1 \).

The set \( N \) of sample sizes could be made even larger by allowing \( M \) to depend on \( k \). It would suffice to have \( \sum_{k=0}^{\infty} M_k b^{-k} < \infty \). For instance \( M_k \sim b^k/(k+1)^{1+\delta} \) for some \( \delta > 0 \) or even \( M_k = b^k/(k+1) \log(k+2) \) would still yield an SLLN.
A \((t, d)\)-sequence in base \(b\) is extensible in that after using \(n\) points of the sequence we can obtain an \(n' > n\) point rule by simply adjoining the next \(n' - n\) points in the sequence. We can get an SLLN for a non-extensible “triangular array” construction where the points in the rule with \(n = n_\ell\) are not necessarily also used for \(n' = n_{\ell'}\) for \(\ell' > \ell\). For \(\ell \in \mathbb{N}\) let \(x_{\ell, i}, \ldots, x_{\ell, n_\ell}\) be the union of \(M_\ell\) digital nets, each with gain coefficient \(\Gamma_\ell < \infty\), and given a nested uniform scramble. Then the proof in Theorem 2 will go through so long as
\[
\sum_{\ell=1}^{\infty} \frac{\Gamma_\ell M_\ell}{n_\ell} < \infty.
\]

5 Discussion

The strong law for averages over nested uniform scrambling of digital nets makes two restrictions not needed in the classical strong law for averages of independent and identically distributed random variables. It requires \(f \in L^2[0,1]^d\) instead of just \(f \in L^1[0,1]^d\), and it uses a strict subset of sample sizes \(n\) instead of all \(n \geq 1\). The SLLN in Theorem 2 applies to a sufficiently large collection of sample sizes for applications. It would be useful to weaken the assumption that \(f \in L^2[0,1]^d\). Even a WLLN for \(f \in L^p[0,1]^d\) for some \(p \in [1, 2)\), especially \(p = 1\), would be valuable.

There are other ways to scramble digital nets and sequences. The linear scrambles of [26] require less space than the nested uniform scramble. They have the same mean squared discrepancy as the nested uniform scramble [19] and so they might also satisfy an SLLN. A digital shift [24, 35] does not produce the same variance as the nested uniform scramble and it does not satisfy the critically important bound (7) on gain coefficients, so the methods used here would not provide an SLLN for it. The nested uniform scramble is the only one for which central limit theorems have been proved [25, 2].

A second major family of RQMC methods has been constructed from lattice rules [41]. Points \(a_1, \ldots, a_n\) on a lattice in \([0, 1]^d\) are randomized into \(x_i = a_i + u\) mod 1, for \(u \sim U[0, 1]^d\). That is, they are shifted with wraparound in what is known as a Cranley-Patterson rotation [5]. For an extensible version of shifted lattice rules, see [18]. The Cranley-Patterson rotation does not provide a \(\Gamma\) bound like (7) because there are functions \(f \in L^2[0,1]^d\) with \(\text{var}(\hat{\mu}_n^{\text{RQMC}}) = \sigma^2(f)\) [24], and so a proof of an SLLN for this form of RQMC would require a different approach. The fact that \(\text{var}(\hat{\mu}_n^{\text{RQMC}}) = \sigma^2(f)\) is possible does not provide a counter-example to an SLLN because this equality might only hold for a finite number of \(n_\ell\) in the infinite sequence. Given a class of functions \(F\) with \(\text{var}(\hat{\mu}_n^{\text{RQMC}}) \leq B\sigma^2(f)/n_\ell\) for all \(f \in F\), all \(\ell \geq 1\), and some \(B < \infty\), we get an SLLN for \(f \in F\) if \(\sum_{\ell=1}^{\infty} 1/n_\ell < \infty\). Some such bounds \(B\) for randomly shifted lattices appear in [24] though they hold for specific \(n_\ell\) not necessarily an infinite sequence of them.
Acknowledgments

I thank Max Balandat for posing the problem of finding and SLLN for randomized QMC. This work was supported by grant IIS-1837931 from the U.S. National Science Foundation.

References


