Quasi-Monte Carlo on Product Spaces

Kinjal Basu

Department of Statistics
Stanford University

Joint work with Prof. Art Owen

Będlewo, Poland
29th April, 2015
Overview

1. Introduction
2. Scrambled Geometric Net
   - Function $\phi$
   - Nested Uniform Scrambling
3. Results
   - Results in $L^2$ not requiring smoothness
   - Scrambled Net Variance for smooth functions
4. Proof Idea
   - Using ANOVA and Multiresolution
   - Variance and gain coefficients
   - Technical Challenges
5. Summary
Problem: Quasi-Monte Carlo integration over product spaces of the form $\mathcal{X}^s$ where $\mathcal{X} \subseteq \mathbb{R}^d$. 

$$
\mu = \frac{1}{\text{vol}(\mathcal{X})^s} \int_{\mathcal{X}^s} f(x) \, dx.
$$
Usual Approach for $s = 1$

- Measure preserving transformation $\tau : [0, 1]^d \rightarrow \mathcal{X}$. 

Usual Approach for $s = 1$

- Measure preserving transformation $\tau : [0, 1]^d \rightarrow \mathcal{X}$.
- $u \sim U[0, 1]^d$ then $x = \tau(u) \sim U(\mathcal{X})$. 
Usual Approach for $s = 1$

- Measure preserving transformation $\tau : [0, 1]^d \to X$.
- $u \sim U[0, 1]^d$ then $x = \tau(u) \sim U(X)$.
- Now use
  $$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \frac{1}{n} \sum_{i=1}^{n} f(\tau(u_i))$$
Usual Approach for $s = 1$

- Measure preserving transformation $\tau : [0, 1]^d \to \mathcal{X}$.
- $u \sim U[0, 1]^d$ then $x = \tau(u) \sim U(\mathcal{X})$.
- Now use
  $$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \frac{1}{n} \sum_{i=1}^{n} f(\tau(u_i))$$

- This is fine for MC. For QMC we often find that $f \circ \tau \notin BVHK$

[Pillards and Cools (2005), Fang and Wang (1994)]
Our estimates are equal weight rules

\[ \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f(x_i), \quad \text{where} \quad x_i = \phi(u_i) \]

for random points \( u_i \in [0, 1]^s \).
Our estimates are equal weight rules

\[ \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f(x_i), \quad \text{where} \quad x_i = \phi(u_i) \]

for random points \( u_i \in [0, 1]^s \).

Most interesting case: \( d = 2 \), such as, triangles, spherical triangles and discs.
For any $f \in L^2(\mathcal{X}^s)$, $\text{Var}(\hat{\mu}) = o(1/n)$. 
Overview of Results

- For any $f \in L^2(\mathcal{X}^s)$, $\text{Var}(\hat{\mu}) = o(1/n)$.
- For each finite $n$, $\text{Var}(\hat{\mu})$ is bounded by a finite multiple of the Monte Carlo variance, uniformly over all $f \in L^2(\mathcal{X}^s)$. 
Overview of Results

- For any $f \in L^2(\mathcal{X}^s)$, $\text{Var}(\hat{\mu}) = o(1/n)$.
- For each finite $n$, $\text{Var}(\hat{\mu})$ is bounded by a finite multiple of the Monte Carlo variance, uniformly over all $f \in L^2(\mathcal{X}^s)$.
- Under smoothness conditions on $f$ and a sphericity constraint on the partitioning of $\mathcal{X}$ we show
  $$\text{Var}(\hat{\mu}) = O\left(\frac{(\log n)^{s-1}}{n^{1+2/d}}\right).$$
Overview

1. Introduction

2. Scrambled Geometric Net
   - Function $\phi$
   - Nested Uniform Scrambling

3. Results
   - Results in $L^2$ not requiring smoothness
   - Scrambled Net Variance for smooth functions

4. Proof Idea
   - Using ANOVA and Multiresolution
   - Variance and gain coefficients
   - Technical Challenges

5. Summary
Special Case: \( s = 1, \ d = 2, \mathcal{X} \) is Triangle.

Figure: Kronecker Triangular Lattice using \( 3\pi/8 \)

- Kronecker Points. Uses badly approximable numbers.
Special Case: \( s = 1, d = 2, \mathcal{X} \) is Triangle.

**Figure:** Kronecker Triangular Lattice using \( 3\pi/8 \)

- Kronecker Points. Uses badly approximable numbers.
- Convergence rate \( O(\log n/n) \)
Special Case: $s = 1, d = 2, \mathcal{X}$ is Triangle.

Figure: Kronecker Triangular Lattice using $3\pi/8$

- Kronecker Points. Uses badly approximable numbers.
- Convergence rate $O(\log n/n)$
- Fails for $d > 2$ and also for $d = 2, s > 1$ due to Littlewood conjecture.

[B. and Owen (2015a)]
Special Case: $s = 1, d = 2, \mathcal{X}$ is Triangle.
Special Case: $s = 1, d = 2, \mathcal{X}$ is Triangle.

- Triangular van der Corput
Special Case: $s = 1, d = 2, \mathcal{X}$ is Triangle.

- Triangular van der Corput
- $n \geq 0$ in a base 4 representation $n = \sum_{k \geq 1} d_k 4^{k-1}$ where $d_k \in \{0, 1, 2, 3\}$
- $x_n = 0.d_1d_2\ldots$
Special Case: $s = 1, d = 2, \mathcal{X}$ is Triangle.

- Triangular van der Corput
- $n \geq 0$ in a base 4 representation $n = \sum_{k \geq 1} d_k 4^{k-1}$ where $d_k \in \{0, 1, 2, 3\}$
- $x_n = 0.d_1d_2\ldots$
- Discrepancy: $O(1/\sqrt{n})$. RMSE under randomization: $O(1/n)$.

[Brandolini et. al. (2013)]
Splits on the Triangle

**Figure:** Splits of a triangle $\mathcal{X}$ for bases $b = 2$, 3 and 4. The subtriangles $\mathcal{X}_j$ are labeled by the digit $j \in \mathbb{Z}_b$. 
Recursive Splits on the Triangle

2^6 Decomposition  

3^3 Decomposition  

4^3 Decomposition

Figure: The base $b$ splits from previous figure carried out to $k = 6$ or 3 or 4 levels.
Definition 1

Let $X$ be a recursive split of $X \in \mathbb{R}^d$ in base $b$. Then $X$ satisfies the sphericity condition if there exists $C < \infty$ such that $\text{diam}(X_{a_1,\ldots,a_k}) \leq C b^{-k/d}$ holds for all cells $X_{a_1,\ldots,a_k}$ in $X$. 
Definition 1

Let $\mathcal{X}$ be a recursive split of $\mathcal{X} \in \mathbb{R}^d$ in base $b$. Then $\mathcal{X}$ satisfies the sphericity condition if there exists $C < \infty$ such that $\text{diam}(\mathcal{X}_{a_1,\ldots,a_k}) \leq C b^{-k/d}$ holds for all cells $\mathcal{X}_{a_1,\ldots,a_k}$ in $\mathcal{X}$.

Definition 2

A recursive split $\mathcal{X}$ in base $b$ is convergent if for every infinite sequence $a_1, a_2, a_3, \ldots \in \mathbb{Z}_b$, the cells $\mathcal{X}_{a_1,a_2,\ldots,a_K}$ converges a point as $K \to \infty$. That point is denoted $\lim_{K \to \infty} \mathcal{X}_{a_1,a_2,\ldots,a_K}$.
**Definition 1**

Let $X$ be a recursive split of $X \in \mathbb{R}^d$ in base $b$. Then $X$ satisfies the *sphericity condition* if there exists $C < \infty$ such that $\text{diam}(X_{a_1, \ldots, a_k}) \leq C b^{-k/d}$ holds for all cells $X_{a_1, \ldots, a_k}$ in $X$.

**Definition 2**

A recursive split $X$ in base $b$ is convergent if for every infinite sequence $a_1, a_2, a_3, \ldots \in \mathbb{Z}_b$, the cells $X_{a_1, a_2, \ldots, a_K}$ converges a point as $K \to \infty$. That point is denoted $\lim_{K \to \infty} X_{a_1, a_2, \ldots, a_K}$.

\[ \phi(x) = \lim_{K \to \infty} X_{x_1, x_2, \ldots, x_K} \text{ where } x \text{ has the base } b \text{ representation } 0.x_1x_2\ldots. \]
Figure: A recursive binary equal area splitting of the unit disk, keeping the aspect ratio close to unity.
For $a \in [0, 1)$ let $a = (0.a_1a_2\ldots)_b$. 
For $a \in [0, 1)$ let $a = (0.a_1 a_2 \ldots)_b$.

Random permutations to the digits $a_k$ yielding $x_k \in \mathbb{Z}_b$ and deliver $x = (0.x_1 x_2 \ldots)_b$. 
For $a \in [0, 1)$ let $a = (0.a_1a_2 \ldots)_b$.

Random permutations to the digits $a_k$ yielding $x_k \in \mathbb{Z}_b$ and deliver $x = (0.x_1x_2 \ldots)_b$.

$x_{k+1} = \pi \cdot a_1a_2 \ldots a_k(a_{k+1})$ where all of these permutations are independent and uniform.
Nested Uniform Scrambling

- For $a \in [0, 1)$ let $a = (0.a_1a_2\ldots)_b$.
- Random permutations to the digits $a_k$ yielding $x_k \in \mathbb{Z}_b$ and deliver $x = (0.x_1x_2\ldots)_b$.
- $x_{k+1} = \pi \cdot a_1a_2\ldots a_k (a_k+1)$ where all of these permutations are independent and uniform.
- For $a = (a_1, \ldots, a_s) \in [0, 1)^s$, we apply nested uniform scramble component-wise.
For \( a \in [0, 1) \) let \( a = (0.a_1 a_2 \ldots)_b \).

Random permutations to the digits \( a_k \) yielding \( x_k \in \mathbb{Z}_b \) and deliver \( x = (0.x_1 x_2 \ldots)_b \).

\[ x_{k+1} = \pi \cdot a_1 a_2 \ldots a_k (a_k + 1) \]

where all of these permutations are independent and uniform.

For \( a = (a_1, \ldots, a_s) \in [0, 1)^s \), we apply nested uniform scramble component-wise.

A nested uniform scramble of \( a_1, \ldots, a_n \in [0, 1)^s \) applies the same set of permutations to the digits of all \( n \) of those points.
For $s \geq 1$, let $a_i$ be $(t, m, s)$-net in base $b$
Scrambled Geometric nets in $\mathcal{X}^s$ via splitting

- For $s \geq 1$, let $a_i$ be $(t, m, s)$-net in base $b$
- Let $u_i \in [0,1]^s$ - nested uniform scrambling.
Scrambled Geometric nets in $\mathcal{X}^s$ via splitting

- For $s \geq 1$, let $a_i$ be $(t, m, s)$-net in base $b$
- Let $u_i \in [0, 1]^s$ - nested uniform scrambling.
- $x_{ij} = \phi(u_{ij})$ for $j = 1, \ldots, s$. Then $x_i \in \mathcal{X}^s$ and we use

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

Overview

1. Introduction
2. Scrambled Geometric Net
   - Function \( \phi \)
   - Nested Uniform Scrambling
3. Results
   - Results in \( L^2 \) not requiring smoothness
   - Scrambled Net Variance for smooth functions
4. Proof Idea
   - Using ANOVA and Multiresolution
   - Variance and gain coefficients
   - Technical Challenges
5. Summary
Theorem 1

Let $u_1, \ldots, u_n$ be a nested uniform scramble of a $(t, m, s)$-net in base $b \geq 2$ and let $x_i = \phi(u_i)$ componentwise. Then for any $f \in L^2(\chi^{1:s})$,

$$
\mathbb{E}(\hat{\mu}) = \mu \\
\text{Var}(\hat{\mu}) = o\left(\frac{1}{n}\right)
$$

as $n \rightarrow \infty$. 
Theorem 2

Under the conditions of Theorem 1

\[ \text{Var}(\hat{\mu}) \leq b^t \left( \frac{b + 1}{b - 1} \right)^{s-1} \frac{\sigma^2}{n}, \]

where \( \sigma^2 = \text{Var}(f(x)) \) for \( x \sim U(\mathcal{X}^{1:s}) \). If \( t = 0 \), then \( \text{Var}(\hat{\mu}) \leq e\sigma^2/n \approx 2.718\sigma^2/n \).
Main Theorem

Theorem : B. and Owen (2015b)

Let \( u_1, \ldots, u_n \) be the points of a randomized \((t, m, s)\)-net in base \( b \). Let \( x_i = \phi(u_i) \in \mathcal{X}^{1:s} \) for \( i = 1, \ldots, n \) where \( \phi \) is the componentwise application of the transformation from convergent recursive splits in base \( b \). Then for a smooth \( f \) on \( \mathcal{X}^{1:s} \),

\[
\text{Var}(\hat{\mu}) = O \left( \frac{(\log n)^{s-1}}{n^{1+2/d}} \right).
\]
A closed set $\mathcal{X} \subseteq \mathbb{R}^m$ with non-empty interior is said to be *Sobol’ extensible* if there exists a point $c \in \mathcal{X}$ such that $z \in \mathcal{X}$ implies $\text{rect}[c, z] \subseteq \mathcal{X}$. The point $c$ is called the anchor.
Sobol Extensible $\mathcal{X}$

Figure: Sobol’ extensible regions. At left, $\mathcal{X}$ is the triangle with vertices $(0, 0), (0, \sqrt{2}), (\sqrt{2}, 0)$ and the anchor is $c = (0, 0)$. At right, $\mathcal{X}$ is a circular disk centered its anchor $c$. The dashed lines depict some rectangular hulls joining selected points to the anchor.
Non Sobol Extensible $\mathcal{X}$

**Figure:** Non-Sobol’ extensible regions. At left, $\mathcal{X}$ is an annular region centered at the origin. At right, $\mathcal{X}$ is the unit square exclusive of an $\epsilon$-wide strip centered on the diagonal.
Smoothness Condition

- Let $\mathcal{X} \subseteq \mathbb{R}^m$ for $m \in \mathbb{N}$ be Sobol extensible. The function $f : \mathcal{X} \rightarrow \mathbb{R}$ is said to be smooth if $\partial^{1:m} f$ is continuous on $\mathcal{X}$. 
Smoothness Condition

- Let $\mathcal{X} \subseteq \mathbb{R}^m$ for $m \in \mathbb{N}$ be Sobol extensible. The function $f : \mathcal{X} \rightarrow \mathbb{R}$ is said to be smooth if $\partial^{1:m}f$ is continuous on $\mathcal{X}$.
- For general $\mathcal{X}$, $f$ is said to be smooth if $f \in C^m(\mathcal{X})$. 
Overview

1. Introduction
2. Scrambled Geometric Net
   - Function \( \phi \)
   - Nested Uniform Scrambling
3. Results
   - Results in \( L^2 \) not requiring smoothness
   - Scrambled Net Variance for smooth functions
4. Proof Idea
   - Using ANOVA and Multiresolution
   - Variance and gain coefficients
   - Technical Challenges
5. Summary
Proof Idea of Main Theorem

- If $X$ is convergent and $f \in L^2(X)$ then,

$$f(x) = \langle f, \varphi \rangle \varphi(x) + \sum_{k=1}^{\infty} \sum_{t=0}^{b^k-1} \sum_{c=0}^{b-1} \langle f, \psi_{ktc} \rangle \psi_{ktc}(x).$$
If $\mathbb{X}$ is convergent and $f \in L^2(\mathcal{X}')$ then,

$$f(x) = \langle f, \varphi \rangle \varphi(x) + \sum_{k=1}^{\infty} \sum_{b^k-1}^{b-1} \sum_{t=0}^{b-1} \sum_{c=0}^{b^k-1} \langle f, \psi_{ktc} \rangle \psi_{ktc}(x).$$

$\psi_{ktc}$ are Haar wavelets in base $b$. 
Proof Idea of Main Theorem

- Multiresolution of $L^2(\mathcal{X}^{1:s})$ is

$$f(x) = \sum_{u \subseteq 1:s} \sum_{\kappa \mid u} \sum_{\tau \mid u, \kappa} \sum_{\gamma \mid u} \langle \psi_{\kappa \tau \gamma}, f \rangle \psi_{\kappa \tau \gamma}(x)$$

$$= \mu + \sum_{|u| > 0} \sum_{\kappa \mid u} \sum_{\tau \mid u, \kappa} \sum_{\gamma \mid u} \sum_{\kappa \mid u} \langle \psi_{\kappa \tau \gamma}, f \rangle \psi_{\kappa \tau \gamma}(x).$$

The sum over $\kappa$ is over all possible values of $\kappa$ given the subset $u$. 

$\psi_{\kappa \tau \gamma}$ are tensor products of Haar wavelets in base $b$. 

$\langle \psi_{\kappa \tau \gamma}, f \rangle$
Proof Idea of Main Theorem

- Multiresolution of $L^2(\mathcal{X}^{1:s})$ is

$$f(x) = \sum_{u \subseteq 1:s} \sum_{|u| > 0} \sum_{\kappa|u} \sum_{\tau|u, \kappa} \sum_{\gamma|u} \langle \psi_{u\kappa\tau\gamma}, f \rangle \psi_{u\kappa\tau\gamma}(x)$$

$$= \mu + \sum_{u \subseteq 1:s} \sum_{|u| > 0} \sum_{\kappa|u} \sum_{\tau|u, \kappa} \sum_{\gamma|u} \langle \psi_{u\kappa\tau\gamma}, f \rangle \psi_{u\kappa\tau\gamma}(x).$$

The sum over $\kappa$ is over all possible values of $\kappa$ given the subset $u$.

- $\psi_{u\kappa\tau\gamma}$ are tensor products of Haar wavelets in base $b$. 
Variance expression

\[ \text{Var}(\hat{\mu}) = \mathbb{E}\left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{|u| > 0} \sum_{\kappa | u \tau | u, \kappa \gamma | u} \sum_{|u'| > 0} \sum_{\kappa' | u' \tau' | u', \kappa' \gamma' | u'} \left( \langle f, \psi_{u\kappa\tau\gamma} \rangle \langle f, \psi_{u'\kappa'\tau'\gamma'} \rangle \psi_{u\kappa\tau\gamma}(x_i) \psi_{u'\kappa'\tau'\gamma'}(x_{i'}) \right) \right) \]
Variance expression

\[
\text{Var}(\hat{\mu}) = \mathbb{E} \left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{|u|>0} \sum_{\kappa |u} \sum_{\tau |u, \kappa} \sum_{\gamma |u} \sum_{|u'|>0} \sum_{\kappa' |u'} \sum_{\tau'|u', \kappa'} \sum_{\gamma'|u'} \langle f, \psi_{u\kappa \tau \gamma} \rangle \langle f, \psi_{u'\kappa' \tau' \gamma'} \rangle \psi_{u \kappa \tau \gamma} (x_i) \psi_{u' \kappa' \tau' \gamma'} (x_{i'}) \right).
\]

\[
\text{Var}(\hat{\mu}) = \sum_{|u|>0} \sum_{\kappa |u} \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} \nu_{u \kappa}(x_i) \right),
\]

where \( \nu_{u \kappa}(x) = \sum_{\tau |u, \kappa} \sum_{\gamma |u} \langle f, \psi_{u \kappa \tau \gamma} \rangle \psi_{u \kappa \tau \gamma} (x) \) with \( \nu_{\emptyset}() = \mu \).
Simplification of Variance expression

- \( \sigma_{uk}^2 = \int_{\mathcal{X}^{1:s}} \nu_{uk}^2(x) \, dx. \)
- The multiresolution-based ANOVA decomposition is

\[
\sigma^2 = \int_{\mathcal{X}^{1:s}} (f(x) - \mu)^2 \, dx = \sum_{|u| > 0} \sum_{\kappa |u} \sigma_{uk}^2.
\]
Simplification of Variance expression

- \( \sigma_{u\kappa}^2 = \int_{\chi^{1:s}} \nu_{u\kappa}^2(x) \, dx \).

- The multiresolution-based ANOVA decomposition is
  \[
  \sigma^2 = \int \chi^{1:s} (f(x) - \mu)^2 \, dx = \sum_{|u| > 0} \sum_{\kappa | u} \sigma_{u\kappa}^2.
  \]

- From the equidistribution properties of \( a_i \):
  \[
  \text{Var}(\hat{\mu}) = \frac{1}{n} \sum_{|u| > 0} \sum_{\kappa | u} \Gamma_{u,\kappa} \sigma_{u\kappa}^2 \leq \frac{b^t}{n} \left( \frac{b + 1}{b - 1} \right)^s \sum_{|u| > 0} \sum_{\kappa + |u| > m-t} \sigma_{u\kappa}^2.
  \]
Proof Idea Continued

- **Bound** $\sigma_{u\kappa}^2$.

\[
\sigma_{u\kappa}^2 = \int_{\mathcal{X}^{1:s}} \nu_{u\kappa}^2(x) \, dx
\]

\[
= \sum_{\tau \mid u, \kappa} \sum_{\gamma, \gamma'} \langle f, \psi_{u\kappa\tau\gamma} \rangle \langle f, \psi_{u\kappa\tau\gamma'} \rangle \int_{\mathcal{X}^{1:s}} \psi_{u\kappa\tau\gamma}(x) \psi_{u\kappa\tau\gamma'}(x) \, dx
\]

\[
= \sum_{\tau \mid u, \kappa} \sum_{\gamma, \gamma'} \langle f, \psi_{u\kappa\tau\gamma} \rangle \langle f, \psi_{u\kappa\tau\gamma'} \rangle \prod_{j \in u} (1 \gamma_j = \gamma'_j - b^{-1})
\]

- **Bound** $|\langle f, \psi_{u\kappa\tau\gamma} \rangle|$
The coefficients $|\langle f, \psi_{uK}\gamma \rangle|$ decay quickly.
Proof Idea Continued

- The coefficients $|\langle f, \psi_{\kappa T \gamma} \rangle|$ decay quickly.
- Extend the region from $\mathcal{X}^s$ to a bounding box.
- Extend the integrand to that bounding box.
Proof Idea Continued

- The coefficients $|\langle f, \psi_{uk\tau} \rangle|$ decay quickly.
- Extend the region from $\mathcal{X}^s$ to a bounding box.
- Extend the integrand to that bounding box.
- Sobol’s low variance extension given by

$$\tilde{f}(x) = \sum_{u \subseteq 1:m} \int_{[c_u, x_u]} \partial^u f(c_{-u} \cdot y_u) 1_{c_{-u} \cdot y_u \in \mathcal{X}} \, dy_u$$

- Whitney extension requiring extra smoothness.
Summary

- RQMC available for $\mathcal{X}^s$ and $\mathcal{X} \subset \mathbb{R}^d$
- Root mean squared error of $O(n^{-1/2 - 1/d}(\log(n))^{(s-1)/2})$.
- We can extend it to unequal $d_j$ by taking $d = \max_{j \in 1:s} d_j$.
- QMC and composition mapping, $O(n^{-1} \log(n)^{sd-1})$.
- Advantage for $d = 1$ and $2$
- Convergence guarantee as well as error estimates.
- Given additional smoothness perhaps higher order nets would work.

[Dick and Baldeaux (2009)]
Thank you!

A very big thank you to

- Prof. Leszek Plaskota and all other organizers of the conference.
- Prof. Art Owen for his discussions and support.
- National Science Foundation (NSF)
- Banach Center for organizing this conference.