

*Numerical Methods in Economics*

MIT Press, 1998

**Notes for Chapter 9: quasi-Monte Carlo Methods**

April 13, 2011

# Quasi-Monte Carlo Methods

- Observation:
  - MC uses “random” sequences to satisfy i.i.d. premise of LLN
  - Integration only needs sequences which are good for integration
  - Integration does not care about i.i.d. property
- Idea of quasi-Monte Carlo methods
  - Explicitly construct a sequence designed to be good for integration.
  - Do not leave integration up to mindless random choices
- Pseudorandom sequence are not random.
  - von Neumann: “Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin.”
  - “pseudo” means “false, feigned, fake, counterfeit, spurious, illusory”
  - Neither LLN nor CLT apply
  - Visual similarities are not mathematically relevant; for example, “optical regression” is not a valid technique.

- Monte Carlo Propaganda
  - Best deterministic methods converge at rate  $N^{-1/d}$
  - MC converges at rate  $N^{-1/2}$  for any dimension  $d$
  - So, MC is far better than any deterministic scheme
- Observations about Monte Carlo Propaganda
  - Implementations of MC use pseudorandom (hence, deterministic) sequences instead of random numbers
  - Implementations of MC converge at rate  $N^{-1/2}$  for any dimension  $d$
  - Therefore, there exist deterministic methods which converge at rate  $N^{-1/2}$  for any dimension  $d$ .
  - Therefore, under MC propaganda logic,  $1/2 = 1/d$  for all  $d > 1$

- Questions
  - What is rate of convergence when using pseudorandom numbers?
  - Why do deterministic pseudorandom methods converge at rate  $N^{-1/2}$  in practice?
- Answer: MC propagandists pull a bait-and-switch
  - They use worst-case analysis in “Best deterministic methods for integrating  $C^0$  functions converge at rate  $N^{-1/d}$ ”
  - They use probability-one criterion when they say “MC methods converge at rate  $N^{-1/2}$ ”
- Mathematical Facts:
  - MC worst-case convergence rate is  $N^{-0}$  - *no convergence* - there is some sequence where MC does not converge
  - Some pseudorandom methods converge at  $N^{-1/2}$  for smooth functions *in worst case*; proofs are number-theoretic.
  - If  $f$  is  $C^k$  and periodic, then there are deterministic rules converging at rate  $N^{-k}$  *independent of dimension*
- Practical facts
  - qMC has been used for many high-dimension (e.g., 360) problems.
  - pMC asymptotics kick in early; qMC asymptotics take longer
  - Therefore, pMC methods have *finite sample advantages*, not asymptotic advantages.
  - “quasi-MC” is bad name since qMC methods have no connection to probability theory

## Equidistributed Sequences

**Definition 1** A sequence  $\{x_j\}_{j=1}^{\infty} \subset R$  is equidistributed over  $[a, b]$  if

$$\lim_{n \rightarrow \infty} \frac{b-a}{n} \sum_{j=1}^n f(x_j) = \int_a^b f(x) dx \quad (9.1.1)$$

for all Riemann-integrable  $f(x)$ . More generally, a sequence  $\{x^j\}_{j=1}^{\infty} \subset D \subset R^d$  is equidistributed over  $D$  iff

$$\lim_{n \rightarrow \infty} \frac{\mu(D)}{n} \sum_{j=1}^n f(x^j) = \int_D f(x) dx \quad (9.1.2)$$

for all Riemann-integrable  $f(x) : R^d \rightarrow R$ , where  $\mu(D)$  is the Lebesgue measure of  $D$ .

- Examples:

- $0, 1/2, 1, 1/4, 3/4, 1/8, 3/8, 5/8, 7/8, \dots$ , is *not* equidistributed over  $[0, 1]$  since  $\frac{b-a}{n} \sum_{j=1}^n x_j$ , the approximation to  $\int_0^1 x dx$ , oscillates.
- Weyl sequence: for  $\theta$  irrational

$$x_n = \{n\theta\}, \quad n = 1, 2, \dots, \quad (9.1.3)$$

where  $\{x\}$  is *fractional part of  $x$*  and defined by

$$\{x\} \equiv x - \max\{k \in Z \mid k \leq x\}$$

is equidistributed

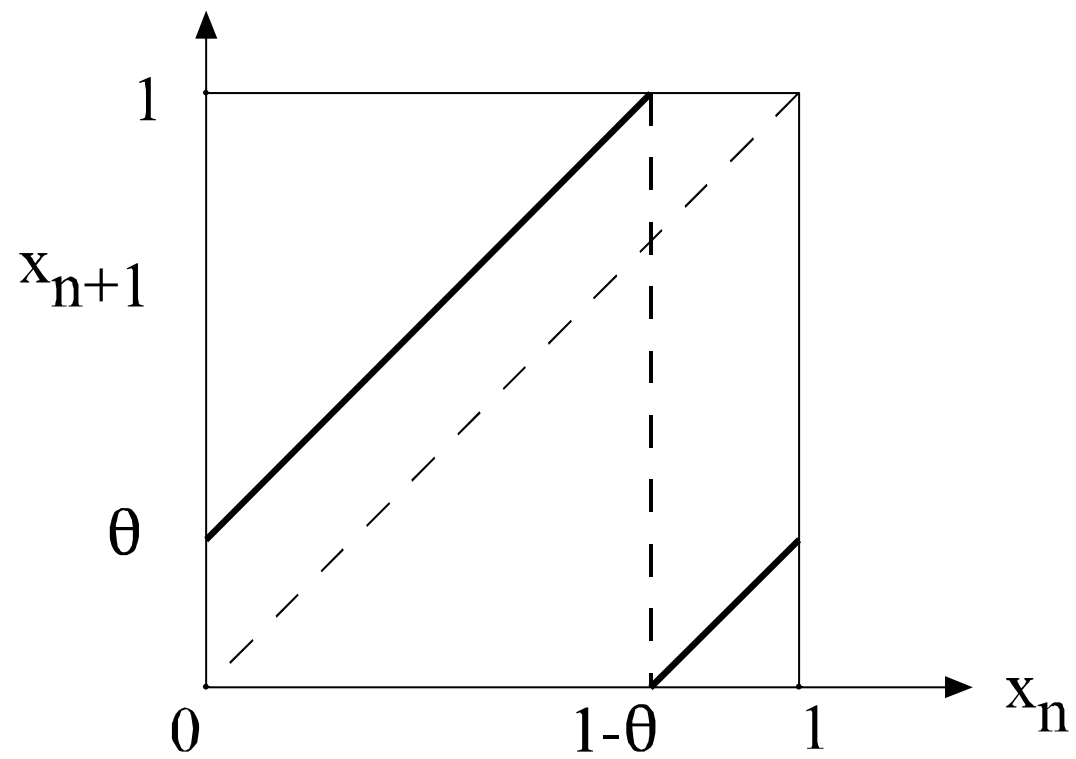
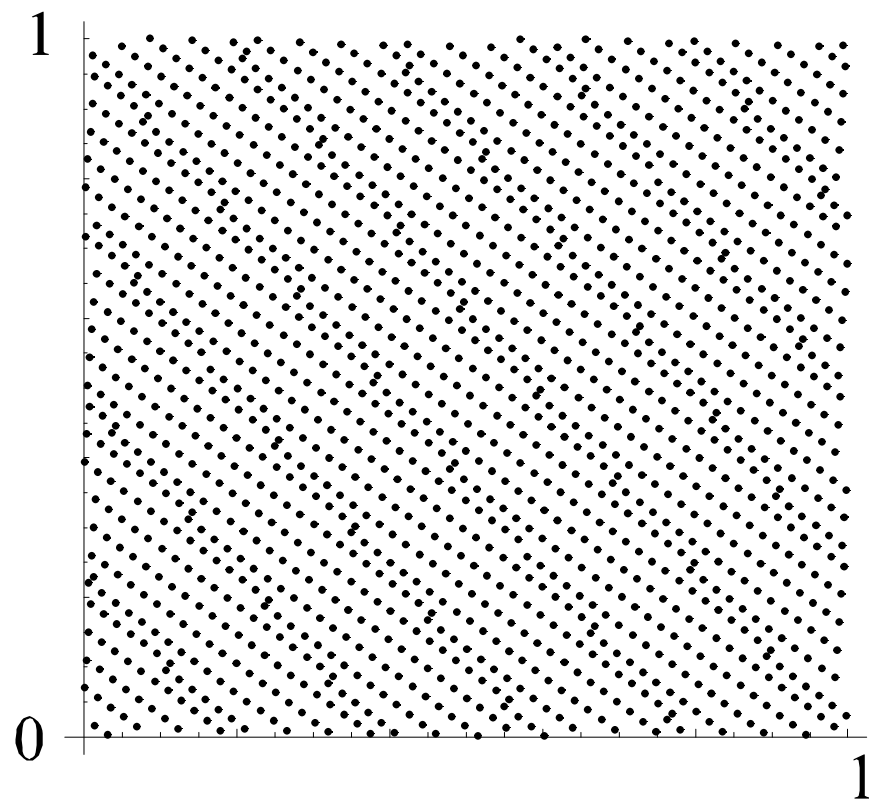
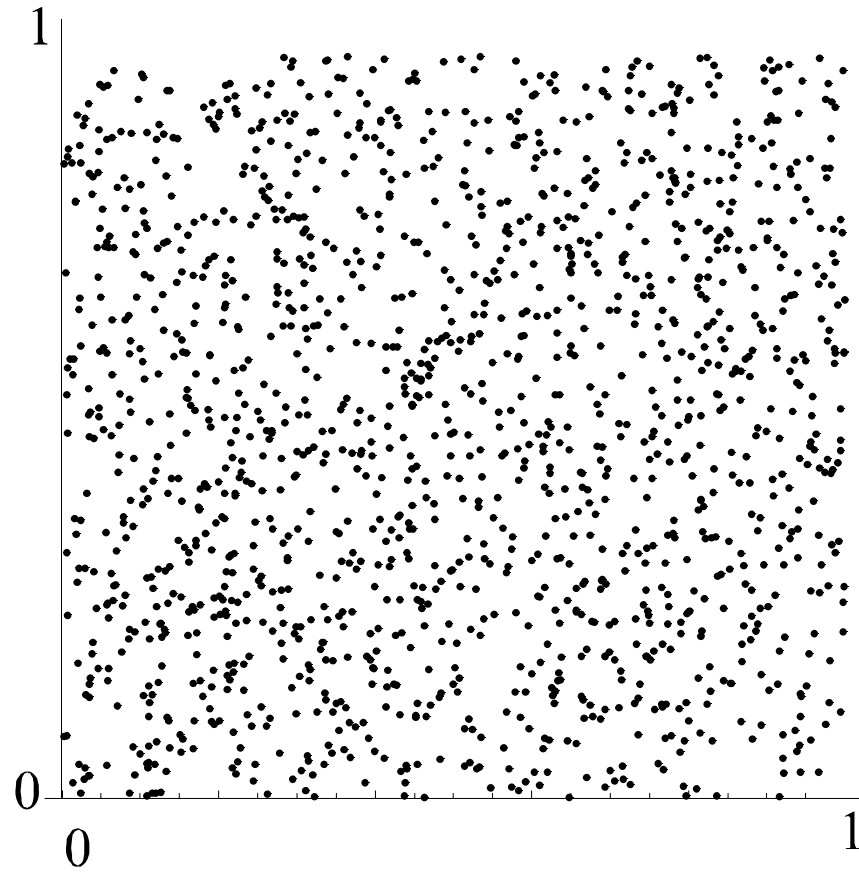


Figure 1: Weyl function



First 1500 Weyl points



1500 Points generated by LCM



Table 9.1: Equidistributed Sequences in  $R^d$

Name:	Formula for $x^n$ :
Weyl	$\left( \left\{ n p_1^{1/2} \right\}, \dots, \left\{ n p_d^{1/2} \right\} \right)$
Haber	$\left( \left\{ \frac{n(n+1)}{2} p_1^{1/2} \right\}, \dots, \left\{ \frac{n(n+1)}{2} p_d^{1/2} \right\} \right)$
Niederreiter	$\left( \left\{ n 2^{1/(d+1)} \right\}, \dots, \left\{ n 2^{d/(d+1)} \right\} \right)$
Baker	$(\{n e^{r_1}\}, \dots, \{n e^{r_d}\})$ , $r_j$ rational and distinct

- MC vs qMC

- qMC are not serially uncorrelated

- Similar iterations for Weyl since  $x_{n+1} = (x_n + \theta) \bmod 1$ , but slope term is 1, not some big number.

# Discrepancy

We want measures of deviation from uniformity for sets of points

**Definition 2** *The discrepancy  $D_N$  of the set  $X \equiv \{x_1, \dots, x_N\} \subset [0, 1]$  is*

$$D_N(X) = \sup_{0 \leq a < b \leq 1} \left| \frac{\text{card}([a, b] \cap X)}{N} - (b - a) \right|.$$

**Definition 3** *If  $X$  is a sequence  $x_1, x_2, \dots \subset [0, 1]$ , then  $D_N(X)$  is  $D_N(X^N)$  where  $X^N = \{x_j \in X \mid j = 1, \dots, N\}$ .*

- **Small discrepancy sets**

- On  $[0, 1]$ , the set with minimal  $D_N$  is  $\left\{ \frac{1}{N+1}, \frac{2}{N+1}, \dots, \frac{N}{N+1} \right\}$

- Discrepancy of lattice point set

$$U_{d,m} = \left\{ \left( \frac{2m_1 - 1}{2m}, \dots, \frac{2m_d - 1}{2m} \right) \mid 1 \leq m_j \leq m, j = 1, \dots, d \right\}$$

is  $\mathcal{O}(m^{-1}) = \mathcal{O}(N^{-1/d})$

- Star discrepancy of  $N$  random points is  $\mathcal{O}(N^{-\frac{1}{2}}(\log \log N)^{1/2})$ , a.s.

- Roth (1954) and Kuipers and Niederreiter (1974):

$$D_N^* > 2^{-4d} ((d-1) \log 2)^{(1-d)/2} N^{-1} (\log N)^{(d-1)/2}. \quad (9.2.1)$$

which is much lower than the Chung-Kiefer result on randomly generated point sets.

- The Halton sequence in  $I^d$  has discrepancy

$$\begin{aligned} D_N &< \frac{d}{N^2} + \frac{1}{N} \prod_{j=1}^d \left( \frac{p_j - 1}{2 \log p_j} \log N + \frac{p_j + 1}{2} \right) \\ &\sim \frac{(\log N)^d}{N} \leq \mathcal{O}(N^{-1+\varepsilon}) \end{aligned} \quad (9.2.4)$$

- Bound not good for moderate  $N$  and large  $d$ .

## Variation and Integration

**Theorem 4** *The total variation of  $f$ ,  $V(f)$ , on  $[0, 1]$  is*

$$V(f) = \sup_n \sup_{0 \leq x_0 < x_1 < \dots < x_n \leq 1} \sum_{j=1}^n |f(x_j) - f(x_{j-1})|$$

**Theorem 5 (Koksma)** *If  $f$  has bounded total variation, i.e.,  $V(f) < \infty$ , on  $I$ , and the sequence  $x_j \in I$ ,  $j = 1, \dots, N$ , has discrepancy  $D_N^*$ , then*

$$\left| N^{-1} \sum_{j=1}^N f(x_j) - \int_0^1 f(x) dx \right| \leq D_N^* V(f) \quad (9.2.5)$$

Can generalize variation to multivariate functions,  $V^{HK}(f)$ .

**Theorem 6 (Hlawka)** *If  $V^{HK}(f)$  is finite and  $\{x^j\}_{j=1}^N \subset I^d$  has discrepancy  $D_N^*$ , then*

$$\left| \frac{1}{N} \sum_{j=1}^N f(x^j) - \int_{I^d} f(x) dx \right| \leq V^{HK}(f) D_N^*.$$

Product rules use lattice sets, which have discrepancy  $O(N^{-1/d})$ , not as good as some other sets with discrepancy  $\mathcal{O}(N^{-1+\varepsilon})$

# Monte Carlo versus Quasi-Monte Carlo

Table 9.2: Integration Errors for  $\int_{I^d} d^{-1} \sum_{j=1}^d |4x_j - 2| dx$

N(1000s)   MC   Weyl   Haber   Niederreiter

d = 10:

1	1(-3)	3(-4)	4(-4)	4(-4)
10	2(-4)	6(-5)	1(-3)	3(-5)
100	1(-3)	7(-6)	2(-4)	2(-6)
1000	4(-5)	6(-7)	2(-4)	2(-7)

d = 40:

1	3(-3)	4(-4)	3(-3)	2(-4)
10	3(-4)	6(-5)	1(-3)	2(-6)
100	4(-6)	5(-6)	3(-4)	9(-6)
1000	1(-4)	6(-7)	1(-5)	4(-7)

Table 9.3: Integration Errors for  $\int_{I^d} \prod_{j=1}^d \left(\frac{\pi}{2} \sin \pi x_j\right) dx$

N(1000s)   MC   Weyl   Haber   Niederreiter

d = 10:

1	1(-2)	6(-2)	8(-2)	9(-3)
10	3(-2)	8(-3)	5(-3)	5(-4)
100	9(-3)	2(-3)	1(-3)	6(-4)
1000	2(-3)	3(-5)	6(-3)	2(-4)

d = 40:

1	4(-1)	5(-1)	5(-2)	7(-1)
10	2(-1)	4(-1)	4(-1)	8(-2)
100	1(-2)	2(-1)	3(-3)	5(-2)
1000	3(-2)	2(-1)	3(-2)	4(-3)

## Fourier Analytic Methods

- Consider  $\int_0^1 \cos 2\pi x \, dx = 0$  and its approximation  $N^{-1} \sum_{n=1}^N \cos 2\pi x_n$

- Choose  $x_n = \{n\alpha\}$ , a Weyl sequence
- Periodicity of  $\cos x$  implies  $\cos 2\pi\{n\alpha\} = \cos 2\pi n\alpha$
- Periodicity of  $\cos 2\pi x$  implies Fourier series representation

$$\cos 2\pi x = \frac{1}{2}(e^{2\pi i x} + e^{-2\pi i x})$$

- Error analysis: error is approximation, and

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \frac{1}{2}(e^{2\pi i n\alpha} + e^{-2\pi i n\alpha}) \\ &= \frac{1}{2N} \sum_{n=1}^N (e^{2\pi i \alpha})^n + \frac{1}{2N} \sum_{n=1}^N (e^{-2\pi i \alpha})^n \\ &\leq \frac{1}{2N} \left( \left| \frac{e^{2\pi i N\alpha} - 1}{e^{2\pi i \alpha} - 1} \right| + \left| \frac{e^{-2\pi i N\alpha} - 1}{e^{-2\pi i \alpha} - 1} \right| \right) \tag{9.3.1} \\ &\leq \frac{1}{2N} \left( \frac{2}{|e^{2\pi i \alpha} - 1|} + \frac{2}{|e^{-2\pi i \alpha} - 1|} \right) \leq \frac{C}{N} \end{aligned}$$

for a finite  $C$  as long as  $e^{2\pi i \alpha} \neq 1$ , which is true for any irrational  $\alpha$ .

- So, convergence rate is  $N^{-1}$ .
- (9.3.1) applies to a finite sum of  $e^{2\pi i kx}$  terms; can be generalized to arbitrary Fourier series.

- The following theorem summarizes results reported in book.

**Theorem 7** *Suppose, for some integer  $k$ , that  $f : [0, 1]^d \rightarrow R$  satisfies the following two conditions:*

1. *All partial derivatives*

$$\frac{\partial^{m_1 + \dots + m_d} f}{\partial x_1^{m_1} \dots \partial x_d^{m_d}}, 0 \leq m_j \leq k - 1, 1 \leq j \leq d$$

*exist and are of bounded variation in the sense of Hardy and Krause, and*

2. *All partial derivatives*

$$\frac{\partial^{m_1 + \dots + m_d} f}{\partial x_1^{m_1} \dots \partial x_d^{m_d}}, 0 \leq m_j \leq k - 2, 1 \leq j \leq d$$

*are periodic on  $[0, 1]^d$ .*

*Then, the error in integrating  $f \in C^k$  with Korobov or Keast good lattice point set with sample size  $N$  is  $O(N^{-k}(\ln N)^{kd})$ .*

- Key observation:

- If  $f$  is  $C^k$  we can find rules with  $O(N^{-k+\varepsilon})$  convergence.
- For smooth functions, there are deterministic rules which far outperform MC
- qMC asymptotics may not kick in until  $N$  is impractically large.



## Estimating Quasi-Monte Carlo Errors

- MC rules have standard errors
- Quasi-MC rules do not have standard errors
- Add “randomization” to construct standard errors
- Suppose

– For each  $\beta$ ,

$$I(f) \doteq Q(f; \beta)$$

– For  $\beta \sim U [0, 1]$

$$I(f) \equiv \int_D f(x) dx = E\{Q(f; \beta)\} \quad (9.5.1)$$

– Then

$$\hat{I} \equiv \frac{1}{m} \sum_{j=1}^m Q(f; \beta_j) \quad (9.5.2)$$

is an unbiased estimator of  $I(f)$  with standard error  $\sigma_{\hat{I}}$  approximated by

$$\hat{\sigma}_{\hat{I}}^2 \equiv \frac{\sum_{j=1}^m (Q(f; \beta_j) - \hat{I})^2}{m - 1} \quad (9.5.3)$$

- Example: Random shifts to Weyl rules, because if  $x_j$  is equidistributed on  $[0, 1]$ , then so is  $x_j + \beta$  for any random  $\beta$ .

## Conclusion

- *All* sampling methods use deterministic sequences
- Probability theory does not apply to *any* practical sampling scheme
- Pseudorandom schemes seem to have  $O(N^{-1/2})$  convergence; this is proven for LCM
- There are  $O(N^{-1})$  schemes for continuously differentiable functions - use equidistributional sequences
- There are  $O(N^{-k})$  schemes for  $C^k$  functions - use Fourier analytic schemes
- qMC methods have done well in some problems with hundreds of dimensions
- Pseudorandom sequences appear to have finite sample advantages for very high dimension problems