*Numerical Methods in Economics*
MIT Press, 1998

# Chapter 9 Notes
# Quasi-Monte Carlo Methods

Kenneth L. Judd
Hoover Institution

October 26, 2002

# 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
- 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
- Retort to 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 do exist deterministic methods which converge at rate $N^{-1/2}$ for any dimension $d$.
  - Therefore, under MC propaganda logic, $0 = 1$.

- Questions
  - What is rate of convergence when using pseudorandom numbers?
  - Why do pseudorandom methods converge at rate $N^{-1/2}$ in practice even though they are deterministic?
- Answer: MC propagandists pull a bait-and-switch
  - They use worst-case analysis when they say "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* - since there always is some sequence where MC does not converge
  - Some pseudorandom methods converge at $N^{-1/2}$ *in worst case*; proofs are number-theoretic.
  - If $f$ is $C^k$ there are deterministic rules which converge 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 logical connection to probability theory

## Equidistributed Sequences

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

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

*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*

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

*for all Riemann-integrable* $f(x) : R^d \to R$*, 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, etc., 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

    $$x_n = \{n\theta\},\ \ n = 1, 2, \cdots, \tag{9.1.3}$$

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

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

    is equidistributed

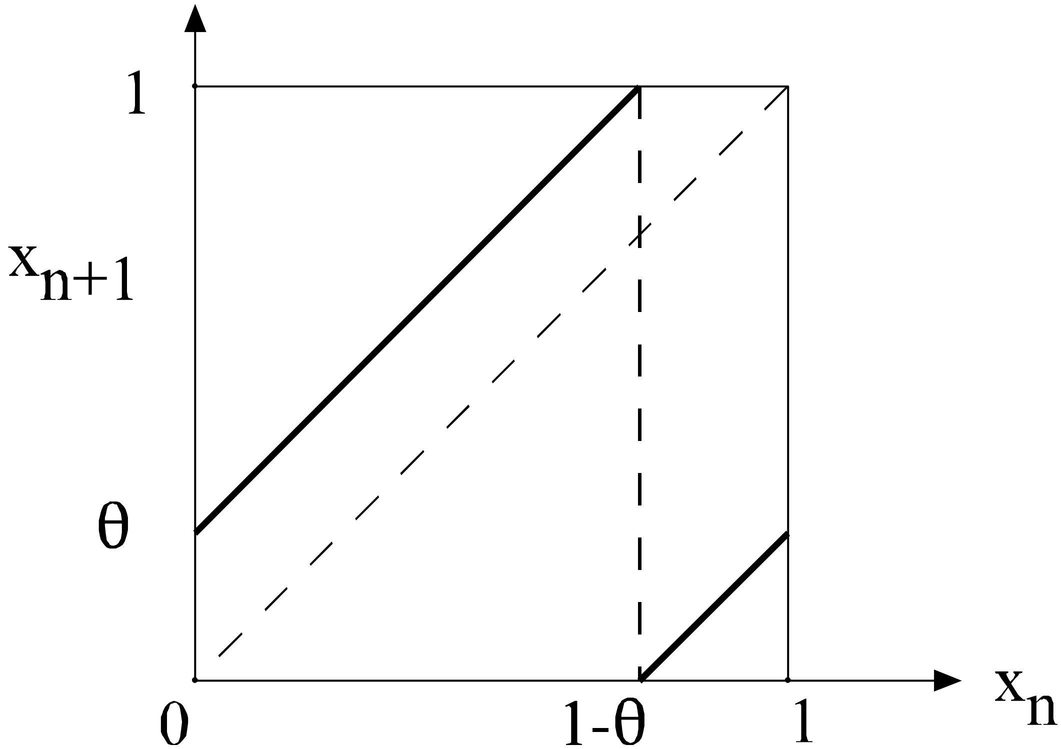
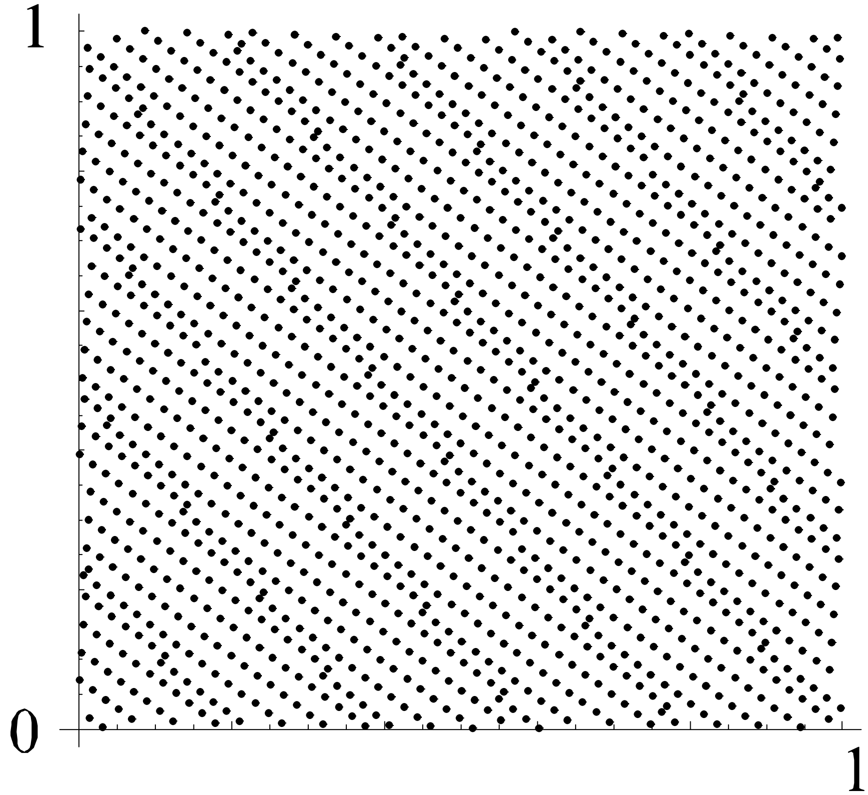



Figure 1: Weyl function




First 1500 Weyl points

Table 9.1: Equidistributed Sequences in $R^d$

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

- MC vs qMC
  - qMC are not serially uncorrelated
  - Similar iterations for Weyl since $x_{n+1} = (x_n + \theta) \text{mod } 1$, but slope term is 1, not big.

## Discrepancy

We want measures of deviation from uniformity for sets of points

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

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

**Definition 3** *The star discrepancy* $D_N^*$ *of the set* $X \equiv \{x_1, \cdots, x_N\} \subset [0,1]$ *is*

$$D_N^*(X) = \sup_{0 \le t \le 1} |\frac{card([0,t) \cap X)}{N} - t|$$

**Definition 4** *If* $X$ *is a sequence* $x_1, x_2, \cdots \subset [0,1]$*, then* $D_N(X)(D_N^*(X))$ *is* $D_N(X^N)(D_N^*(X^N))$ *where* $X^N = \{x_j \in X \mid j = 1, \cdots, N\}$*.*

**Definition 5** *The star discrepancy* $D_N^*$ *of the set* $X = \{x_1, \cdots, x_N\} \subset I^d$ *is*

$$D_N^*(X) = \sup_{0 \le t_1, \cdots t_d \le 1} |\frac{card([0,\, t_1) \times \cdots \times [0, t_d) \cap X)}{N} - \Pi_{j=1}^d t_j|$$

*and* $D_N^*(X)$ *for sequences in* $I^d$ *is defined as in the case* $d = 1$*.* $D_N(X)$ *is generalized similarly from* $d = 1$*.*

- **Small discrepancy sets**
  - On $[0,1]$, the set with minimal $D_N$ and $D_N^*$ is $\left\{\frac{1}{N+1}, \frac{2}{N+1}, \cdots, \frac{N}{N+1}\right\}$
  - Discrepancy of lattice point set

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

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

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

    which is much lower than the Chung-Kiefer result on randomly generated point sets.
  - The Halton sequence in $I^d$ has discrepancy

    $$\begin{aligned} D_N &< \frac{d}{N^2} + \frac{1}{N}\prod_{j=1}^{d}\left(\frac{p_j - 1}{2\log p_j}\log N + \frac{p_j + 1}{2}\right) \qquad (9.2.4) \\ &\sim \frac{(\log N)^d}{N} \le \mathcal{O}\left(N^{-1+\varepsilon}\right) \end{aligned}$$
  - Bound not good for moderate $N$ and large $d$.

Variation and Integration

**Theorem 6** *The total variation of $f$, $V(f)$, on $[0,1]$ is*

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

**Theorem 7** *(Koksma) If $f$ has bounded total variation, i.e., $V(f) < \infty$, on $I$, and the sequence $x_j \in I$, $j = 1, \cdots, N$, has discrepancy $D_N^*$, then*

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

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

**Theorem 8** *(Hlawka) If $V^{HK}(f)$ is finite and $\{x^j\}_{j=1}^N \subset I^d$ has discrepancy $D_N^*$, then*

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

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

# Monte Carlo versus Quasi-Monte Carlo

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

| 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 $\int_{I^d} \Pi_{j=1}^{d} \left( \frac{\pi}{2} \sin \pi x_j \right) \, dx$

| 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

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

    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 $e^{2\pi i k x}$ terms; can be generalized to arbitrary Fourier series.

- The following theorem summarizes results reported in book.

**Theorem 9** *Suppose, for some integer $k$, that $f : [0,1]^d \rightarrow R$ satisfies the following two conditions:*

*1. All partial derivatives*

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

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

*2. All partial derivatives*

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

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

*Then, the error in integrating $f \in \mathcal{C}^k$ 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 $\beta \sim U\ [0,1]$
$$I(f) \equiv \int_D f(x)\ dx = E\{Q(f;\beta)\} \tag{9.5.1}$$
  - Then
$$\hat{I} \equiv \frac{1}{m}\sum_{j=1}^{m} Q(f;\beta_j) \tag{9.5.2}$$
    is an unbiased estimator of $I(f)$ with standard error $\sigma_{\hat{I}}$ approximated by
$$\hat{\sigma}_{\hat{I}}^2 \equiv \frac{\sum_{j=1}^{m}(Q(f;\beta_j) - \hat{I})^2}{m-1} \tag{9.5.3}$$
- 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 $O\left(N^{-1/2}\right)$ convergence; this is proven for LCM
- There are $O\left(N^{-1}\right)$ 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