MATH-414 – Stochastic simulation

Lecture 8: Quasi Monte Carlo

Prof. Fabio Nobile

EPFL

Outline

Quasi Monte Carlo

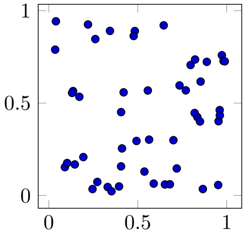
Quasi Monte Carlo

Setting:

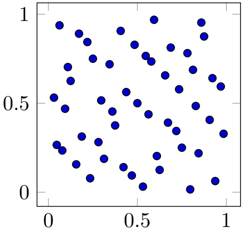
- $Z = \psi(\boldsymbol{X})$ with $\boldsymbol{X} = (X_1, \ldots, X_d) \sim \mathcal{U}([0,1]^2)$
- **Goal**: compute $\mathbb{E}[Z] = \int_{[0,1]^d} \psi(x_1, \ldots, x_d) dx_1 \cdots dx_d$

  using a "Monte Carlo like" estimator $\hat{\mu}_{QMC} = \frac{1}{N} \sum_{i=1}^{N} \psi(\boldsymbol{X}^{(i)})$

Question: Can we improve the Monte Carlo convergence rate using a "better" desing than a purely random one? $\rightsquigarrow$ similar direction as stratification



(a) random sampling



(b) QMC Sampling (Sobol sequence)

Discrepancy function

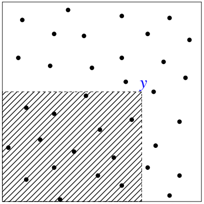
Notation

- $\mathcal{P} = \{\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}\}$: point set (design) in $[0,1]^d$
- for $\boldsymbol{y} \in [0,1]^d$, denote $[\mathbf{0}, \boldsymbol{y}] = \prod_{i=1}^{d} [0, y_i]$ (hyper-rectangle of vertices $\mathbf{0}$ and $\boldsymbol{y}$)
- $\mathrm{Vol}([\mathbf{0}, \boldsymbol{y}]) = \prod_{i=1}^{d} y_i$, volume of hyper-rectangle $[\mathbf{0}, \boldsymbol{y}]$
- empirical volume based on point-set $\mathcal{P}$

$$\widehat{\mathrm{Vol}}_{\mathcal{P}}([\mathbf{0}, \boldsymbol{y}]) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[\mathbf{0}, \boldsymbol{y}]}(\boldsymbol{X}^{(i)}) = \frac{\#\{\boldsymbol{X}^{(i)} \in [\mathbf{0}, \boldsymbol{y}]\}}{N}.$$

Discrepancy function: $\Delta_{\mathcal{P}} : [0,1]^d \to [-1, 1]$

$$\begin{aligned} \Delta_{\mathcal{P}}(\boldsymbol{y}) &= \widehat{\mathrm{Vol}}_{\mathcal{P}}([\mathbf{0}, \boldsymbol{y}]) - \mathrm{Vol}([\mathbf{0}, \boldsymbol{y}]) \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[\mathbf{0}, \boldsymbol{y}]}(\boldsymbol{X}^{(i)}) - \prod_{j=1}^{d} y_j. \end{aligned}$$



Discrepancy function

Measures of discrepancy of a point set $\mathcal{P}$:

L_q-discrepancy: $$D_{N,q}(\mathcal{P}) = \|\Delta_{\mathcal{P}}\|_{L^q} = \left(\int_{[0,1]^d} |\Delta_{\mathcal{P}}(\boldsymbol{y})|^q \, d\boldsymbol{y} \right)^{1/q}, \ 1 \leq q < \infty$$

$*$-discrepancy: $$D_N^*(\mathcal{P}) = \|\Delta_{\mathcal{P}}\|_{L^\infty} = \sup_{\boldsymbol{y} \in [0,1]^d} |\Delta_{\mathcal{P}}(\boldsymbol{y})|.$$

Zaremba's identity (in 1D)

Lemma

- $\psi : [0,1] \to \mathbb{R}$ *absolutely continuous with integrable derivatives*
- $\mathcal{P} = \{X^{(1)}, \dots, X^{(N)}\}$ *point set in* $[0,1]$

$$\int_0^1 \psi(x)\,dx - \frac{1}{N}\sum_{i=1}^{N}\psi(X^{(i)}) = \int_0^1 \psi'(y)\Delta_{\mathcal{P}}(y)\,dy - \Delta_{\mathcal{P}}(1)\psi(1)$$

Proof. Using identity $\psi(x) = \psi(1) - \int_x^1 \psi'(y)\,dy$ in left hand side

$$\int_0^1 \psi(x)\,dx - \frac{1}{N}\sum_{i=1}^{N}\psi(X^{(i)})$$

$$= \psi(1) - \int_0^1\int_x^1 \psi'(y)\,dy\,dx - \frac{1}{N}\sum_{i=1}^{N}\psi(1) + \frac{1}{N}\sum_{i=1}^{N}\int_{X^{(i)}}^1 \psi'(y)\,dy$$

$$= \int_0^1 \psi'(y)\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{[0,y]}(X^{(i)}) - y\right]dy - \psi(1)\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{[0,1]}(X^{(i)}) - 1\right]$$

EPFL

Koksma-Hlawka inequality

$$\left| \int_0^1 \psi(x)\, dx - \frac{1}{N} \sum_{i=1}^{N} \psi(X^{(i)}) \right| \leq \|\psi'\|_{L_p} \|\Delta_{\mathcal{P}}\|_{L_q}, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

In particular, if ψ' is integrable (or ψ has bounded total variation) then

$$\left| \int_0^1 \psi(x)\, dx - \frac{1}{N} \sum_{i=1}^{N} \psi(X^{(i)}) \right| \leq \|\psi\|_{\mathrm{TV}} D_N^*(\mathcal{P}).$$

The QMC quadrature error is proportional to the $*$-discrepancy of the point set, provided that the function ψ has bounded total variation.

Generalization to higher dimension – Hlawka's identity

Notation

- $\mathbf{u} = \{u_1, \ldots, u_k\} \subset \{1, \ldots, d\}$: subset of dimensions (without repetition)
- $|\mathbf{u}| = k$: number of dimensions in $\mathbf{u}$
- For $\boldsymbol{x} = (x_1, \ldots, x_d) \in [0,1]^d$, denote
  - $\boldsymbol{x}_{\mathbf{u}} = (x_{u_1}, \ldots, x_{u_k}) \in [0,1]^k$
  - $\boldsymbol{z} = (\boldsymbol{x}_{\mathbf{u}}, 1)$: vector s.t. $z_j = x_j$ if $j \in \mathbf{u}$ and $z_j = 1$ if $j \notin \mathbf{u}$

Lemma (**Hlawka's identity**)

Let $\psi : [0,1]^d \to \mathbb{R}$ be an integrable function with integrable mixed first order derivatives of any order, and let $\mathcal{P} = \{\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}\}$ be an arbitrary point set in $[0,1]^d$. Then

$$\frac{1}{N}\sum_{i=1}^{N}\psi(\boldsymbol{X}^{(i)}) - \int_{[0,1]^d}\psi(\boldsymbol{x})\,d\boldsymbol{x} = \sum_{\mathbf{u}\subset\{1,\ldots,d\}}(-1)^{|\mathbf{u}|}\int_{[0,1]^{|\mathbf{u}|}}\frac{\partial^{|\mathbf{u}|}\psi}{\partial \boldsymbol{x}_{\mathbf{u}}}(\boldsymbol{x}_{\mathbf{u}},1)\Delta_{\mathcal{P}}(\boldsymbol{x}_{\mathbf{u}},1)\,d\boldsymbol{x}_{\mathbf{u}}$$

where $\frac{\partial^{|\mathbf{u}|}\psi}{\partial \boldsymbol{x}_{\mathbf{u}}} = \frac{\partial^k \psi}{\partial x_{u_1}\ldots\partial x_{u_k}}$ is a mixed first order derivative.

Multidimensional Koksma-Hlawka's inequality

Define the following norm

$$\|\psi\|_{p,p'} = \left(\sum_{\mathbf{u} \subset \{1,\dots,d\}} \left(\int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial^{|\mathbf{u}|}}{\partial \boldsymbol{x}_{\mathbf{u}}} \psi(\boldsymbol{y}_{\mathbf{u}}, 1) \right|^p d\boldsymbol{y}_{\mathbf{u}} \right)^{p'/p} \right)^{1/p'} .$$

Multidimensional **Koksma-Hlawka inequality**, provided $\|\psi\|_{p,p'} < +\infty$

$$\left| \int_{[0,1]^d} \psi(\boldsymbol{x})\, d\boldsymbol{x} - \frac{1}{N} \sum_{i=1}^{N} \psi(\boldsymbol{X}^{(i)}) \right| \leq \|\psi\|_{p,p'} \|\Delta_{\mathcal{P}}\|_{q,q'}, \qquad \frac{1}{p} + \frac{1}{q} = \frac{1}{p'} + \frac{1}{q'} = 1$$

In particular, if $\|\psi\|_{1,1} < +\infty$, then

$$\left| \int_{[0,1]^d} \psi(\boldsymbol{x})\, d\boldsymbol{x} - \frac{1}{N} \sum_{i=1}^{N} \psi(\boldsymbol{X}^{(i)}) \right| \leq \|\psi\|_{1,1} D_N^*(\mathcal{P}).$$

Again, the QMC quadrature error is proportional to the $*$-discrepancy $D_N^*(\mathcal{P})$ provided ψ has *integrable mixed first order derivatives*.

Low discrepancy point sets and sequences

Definition.

- *A family $\mathcal{P} = \{\mathcal{P}_N\}_{N\in\mathbb{N}}$ of non-nested point sets $\mathcal{P}_N = \{\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\} \subset [0,1]^d$ is called a low discrepancy family of point sets if*

$$D_N^*(\mathcal{P}) = O\left(\frac{(\log N)^{d-1}}{N}\right)$$

- *A point sequence $\mathcal{S} = \{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots\} \subset [0,1]^d$ is called a low discrepancy sequence if the corresponding family $\mathcal{P} = \{\mathcal{P}_N\}_{N\in\mathbb{N}}$ of point sets given by $\mathcal{P}_N = \{\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\}$ (first N terms of the sequence) satisties*

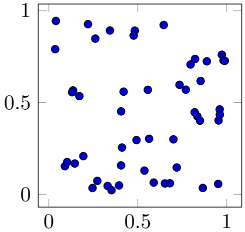
$$D_N^*(\mathcal{S}) = O\left(\frac{(\log N)^{d}}{N}\right)$$

These bounds are believed to be the best possible for point sets and sequences.

The term $(\log N)^d$ growns exponentially with the dimension and the bound becomes useless for $d > \log N$. However, this degeneracy with the dimension is not observed in most applications.

Example 1 – random points

Consider the sequence $\mathcal{S} = \{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots, \boldsymbol{X}^{(N)}, \dots\} \subset [0,1]^d$, with $\boldsymbol{X}^{(i)} \overset{\text{iid}}{\sim} \mathcal{U}([0,1]^d)$, i.e. a random design.

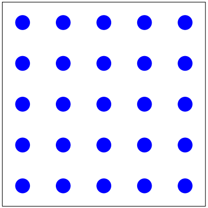


How does $D_N^*(\mathcal{S})$ behaves?

Example 2 – regular lattice

Consider the family $\mathcal{P} = \{\mathcal{P}_N\}_{N\in\mathbb{N}}$ of point sets given by regular lattices

$$\mathcal{P}_N = \left\{ \left(\frac{k_1 + 1/2}{m}, \ldots, \frac{k_d + 1/2}{m} \right), 0 \le k_j \le m-1, j = 1, \ldots, d \right\}, \quad N = m^d$$



How does $D_N^*(\mathcal{P})$ behaves?

Hint: observe that

$$D_N^*(\mathcal{P}) = \sup_{\boldsymbol{y}\in[0,1]^d} |\Delta_{\mathcal{P}}(\boldsymbol{y})| \ge \sup_{t\in[0,1]} |\Delta_{\mathcal{P}}(t, 1, \ldots, 1)|$$

Van der Corput sequence

- Take $b \in \mathbb{N}$, $b \geq 2$.
- Any $n \in \mathbb{N}_0$ can be written as $n = n_0 + n_1 b + n_2 b^2 + \dots$ (b-adic expansion).

Definition. *We define radical inverse of n, denoted $\phi_b(n)$ as*

$$\varphi_b(n) = \frac{n_0}{b} + \frac{n_1}{b^2} + \dots .$$

Obviously $\varphi_b : \mathbb{N}_0 \to [0,1)$.

b-adic Van der Corput sequence (in 1D)

$$\mathcal{S} = \{\varphi_b(0),\ \varphi_b(1),\ \varphi_b(2),\ \dots\}$$

Example for $b = 2$: $0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \dots$

It is easy to check that $D_N^*(\mathcal{S}) = \frac{1}{b^{\lfloor \log_b N \rfloor}} = O(\frac{1}{N})$

Halton sequence; Hammersley point set

Halton sequence:

Let $b_1, \ldots, b_d \geq 2$ be integers pairwise relatively prime.

$$\mathcal{S} = \{\boldsymbol{X}^{(n)}, n \in \mathbb{N}_0\}, \qquad \boldsymbol{X}^{(n)} = (\varphi_{b_1}(n), \varphi_{b_2}(n), \ldots, \varphi_{b_d}(n))$$

- It achieves the optimal bound $D_N^*(\mathcal{S}) \leq c(d)\frac{(\log N)^d}{N}$

Hammersley point sets:

$$\mathcal{P}_N = \{\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(N-1)}\}, \quad \text{with} \quad \boldsymbol{X}^{(n)} = \left(\frac{n}{N}, \varphi_{b_1}(n), \ldots, \varphi_{b_{d-1}}(n)\right)$$

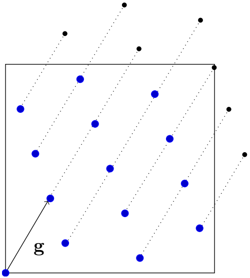
- The family $\mathcal{P} = \{\mathcal{P}_N\}$ is non-nested
- It achieves the better bound $D_N^*(\mathcal{P}) = c(d)\frac{(\log N)^{d-1}}{N}$

Rank-1 lattice rules

- take $N \in \mathbb{N}$ (usually prime)
- $\boldsymbol{g} \in \mathbb{N}^d$, $\boldsymbol{g} = (g_1, \ldots, g_d)$ such that g_j has no factor in common with N.
- Notation: for $X \in [0,1]$ denote $\{X\}$ the fractional part of X

Rank-1 lattice rule with generating vector $\boldsymbol{g}$:

$$\mathcal{P}_N = \left\{ \frac{n\boldsymbol{g}}{N} \right\}_{n=0}^{N-1}$$



Good choices of $\boldsymbol{g}$ lead to low discrepancy non-nested point sets.

$(t-m-d)$ nets

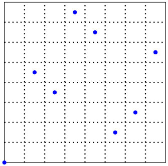
Definition. *Let $0 \leq t \leq m \in \mathbb{N}$ and $b \geq 2$. A $(t\text{-}m\text{-}d)$-net in base b is a point set $\mathcal{P}_N$ consisting of $N = b^m$ points such that each elementary rectangle of volume b^{t-m},*

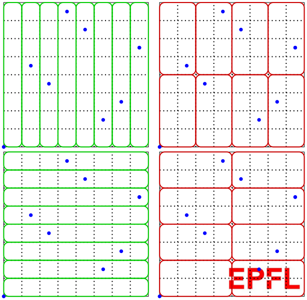
$$R_a = \prod_{j=1}^{d} \left[\frac{a_j - 1}{b^{p_j}}, \frac{a_j}{b^{p_j}} \right), \qquad a_j = 1, \ldots, b^{p_j}$$

with $p_1 + p_2 + \ldots + p_d = m - t$ contains exactly b^t points.

Example: (0-3-2)-net in base $b = 2$:

- point set with $N = 2^3 = 9$ points
- each elementary rectangle with volume $2^{-(m-t)} = 1/8$ contains exactly $2^t = 1$ point.





$(t-m-d)$ nets and $(t-d)$ sequences

Definition. *A (t-d) sequence in base b is a sequence* $\mathcal{S} = \{\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)}, \dots\}$ *such that for any* $m > t$*, every block of* b^m *points* $\{\boldsymbol{X}^{(\ell b^m)}, \dots, \boldsymbol{X}^{((\ell+1)b^m-1)}\}$, $\ell \in \mathbb{N}$ *is a (t-m-d)-net in base b.*

- The star-discrepancy of a $(t\text{-}m\text{-}d)$-net satisfies

$$D_N^*(\mathcal{P}) = O\left(b^t \frac{(\log N)^{d-1}}{N}\right)$$

- The star-discrepancy of a $(t\text{-}d)$-sequence satisfies

$$D_N^*(\mathcal{S}) = O\left(b^t \frac{(\log N)^{d}}{N}\right)$$

Famous $(t\text{-}d)$-sequences are those of Sobol, Niederreiter and Faure.

Controlling the error in QMC

Let

- $\mu = \mathbb{E}\left[\psi(X)\right] = \int_{[0,1]^d} \psi(\boldsymbol{x}) d\boldsymbol{x}$
- $\hat{\mu}_{QMC} = \frac{1}{N}\sum_{i=1}^{N} \psi(\boldsymbol{X}^{(i)})$ for a given point set $\mathcal{P} = \{\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\}$

We have seen an "a-priori" error stimate for QMC:

$$|\mu - \hat{\mu}_{QMC}| = \left| \int_{[0,1]^d} \psi(\boldsymbol{x})\, d\boldsymbol{x} - \frac{1}{N}\sum_{i=1}^{N} \psi(\boldsymbol{X}^{(i)}) \right| \leq \|\psi\|_{1,1} D_N^*(\mathcal{P}).$$

Not of practical use as $\|\psi\|_{1,1}$ and $D_N^*(\mathcal{P})$ ar not known.

Question: how can we estimate the error in QMC?

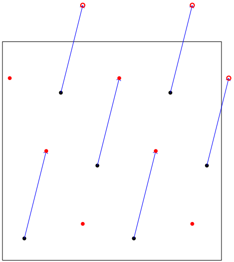
Randomly shifted QMC

Let $\mathcal{P}_N = \{\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\}$ be a low discrepancy point set. Take $\boldsymbol{U} \sim \mathcal{U}([0,1]^d)$

The randomly shifted point set

$$\mathcal{P}_N^{\boldsymbol{U}} = \{\{\boldsymbol{X}^{(1)} + \boldsymbol{U}\}, \dots, \{\boldsymbol{X}^{(N)} + \boldsymbol{U}\}\}$$

has the same discrepancy as $\mathcal{P}_N$



Idea: generate several randomly shifted point sets with independent shifts

Randomly shifted QMC

- Generate $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(k)} \overset{\text{iid}}{\sim} \mathcal{U}([0,1]^d)$
- For each shift $\boldsymbol{U}^{(j)}$, construct the randomly shifted point set

$$\mathcal{P}_N^{(j)} = \{\{\boldsymbol{X}^{(1)} + \boldsymbol{U}^{(j)}\}, \dots, \{\boldsymbol{X}^{(N)} + \boldsymbol{U}^{(j)}\}\}$$

- compute QMC estimator $\hat{\mu}_{\text{QMC}}^{(j)}$ using the point set $\mathcal{P}_N^{(j)}$

$$\hat{\mu}_{\text{QMC}}^{(j)} = \frac{1}{N} \sum_{i=1}^{N} \psi(\{\boldsymbol{X}^{(i)} + \boldsymbol{U}^{(j)}\})$$

- Estimate μ by $\quad \hat{\mu}_{\text{QMC}} = \frac{1}{k} \sum_{j=1}^{k} \hat{\mu}_{\text{QMC}}^{(j)}$
- Notice that $\{\boldsymbol{X}^{(i)} + \boldsymbol{U}^{(j)}\} \sim \mathcal{U}([0,1]^d)$ for any i and j. Hence $\hat{\mu}_{QMC}$ is unbiased and $\hat{\mu}_{QMC}^{(i)}$ are all independent
- $\mathbb{V}\text{ar}\left[\hat{\mu}_{\text{QMC}}\right] = \frac{\sigma_{\text{QMC}}^2}{k}, \qquad \sigma_{\text{QMC}}^2 = \mathbb{E}\left[(\hat{\mu}_{\text{QMC}}^{(j)} - \mu)^2\right] = O\left(\frac{(\log N)^{2(d-1)}}{N^2}\right)$
- Asymptotic $1-\alpha$ confidence interval for $\hat{\mu}_{\text{QMC}}$

$$I_\alpha = \left[\hat{\mu}_{\text{QMC}} - c_{1-\alpha/2}\frac{\hat{\sigma}_{\text{QMC}}}{\sqrt{k}},\ \hat{\mu}_{\text{QMC}} + c_{1-\alpha/2}\frac{\hat{\sigma}_{\text{QMC}}}{\sqrt{k}}\right]$$

Randomly shifted QMC

Algorithm: Randomly shifted QMC.

1 Generate point set $P_N = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)})$

2 Generate $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(k)} \overset{\text{iid}}{\sim} \mathcal{U}([0,1]^d)$;

3 For $j = 1, \ldots, k$, compute $\hat{\mu}^{(j)}_{\text{QMC}} = \frac{1}{N} \sum_{i=1}^{N} \psi(\{\boldsymbol{X}^{(i)} + \boldsymbol{U}^{(j)}\})$;

4 Compute $\hat{\mu}_{\text{QMC}} = \frac{1}{k} \sum_{j=1}^{k} \hat{\mu}^{(j)}_{\text{QMC}}$ as well as $\hat{\sigma}^2_{QMC} = \frac{1}{k-1} \sum_{j=1}^{k} (\hat{\mu}^{(j)}_{QMC} - \hat{\mu}_{QMC})^2$;

5 Output $\hat{\mu}_{QMC}$ as well as a $1 - \alpha$ confidence

$$I_\alpha = \left[\hat{\mu}_{\text{QMC}} - c_{1-\alpha/2} \frac{\hat{\sigma}_{\text{QMC}}}{\sqrt{k}},\ \hat{\mu}_{\text{QMC}} + c_{1-\alpha/2} \frac{\hat{\sigma}_{\text{QMC}}}{\sqrt{k}} \right]$$

EPFL