Low-discrepancy sampling

Jan-Willem Buurlage

Laboratory Class Scientific Computing (WISM454)

Recap of last week

- Quadrature methods
  - Trapezoidal rule
  - Error depends on $\frac{1}{k^2} \max_x |f''(x)|$
  - $n = k^d$ points in d dimensions
  - To achieve error ϵ we need

$$n \propto \left(\frac{1}{\epsilon}\right)^{d/2}$$

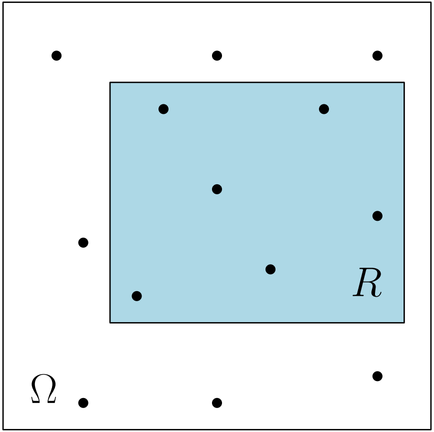
- Monte Carlo methods
  - Hit-or-miss
  - Simple sampling
  - To achieve error ϵ we need

$$n \propto \left(\frac{1}{\epsilon}\right)^{2}$$

Discrepancy

- With Monte Carlo methods, sample points are selected randomly, is this optimal?
- Intuitively, the *discrepancy* of a sequence is a measure of the gaps that a sequence leaves
- *Sampling for low discrepancy* is the subject of today

Discrepancy



- We estimate the area of R by hit-or-miss sampling with sequence of points

Discrepancy definition

- Let $\Omega = [0,1]^d$. For some sampling sequence $\{\vec{x}_j\}$, what is the largest error in estimating rectangular volumes?
- $R = [a_1, b_1] \times \ldots \times [a_d, b_d]$, volume is

$$V(R) = \prod_{i=1}^{d} (b_i - a_i).$$

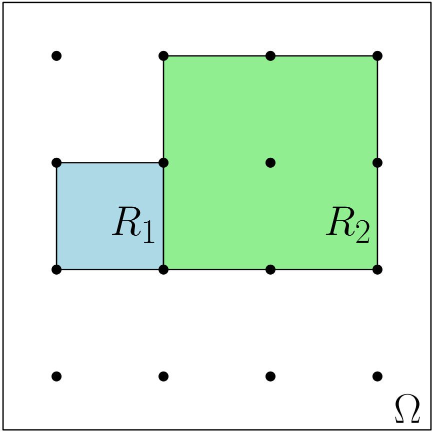
- Simple sampling with first n elements of the sequence gives:

$$\tilde{V}_n(R) = |\{j \leq n \,|\, \vec{x}_j \in R\}|.$$

- Discrepency D defined as

$$D_n = \sup_{\text{rectangles } R} |\tilde{V}_n(R) - V(R)|.$$

Discrepancy Example (uniform)



- $V(R_1) = \frac{1}{16}$, $V(R_2) = \frac{1}{4}$, $\tilde{V}(R_1) = 0$, $\tilde{V}(R_2) = \frac{1}{16}$.

Discrepency for first n points

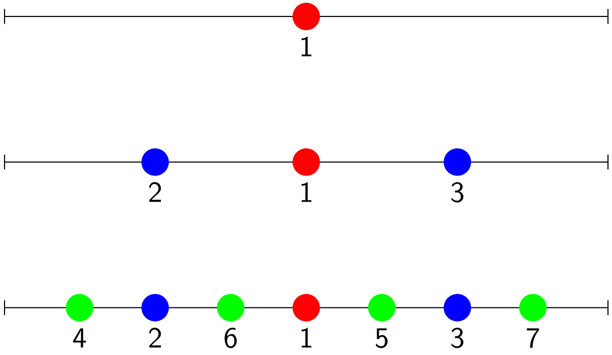
- We want a sequence that has low discrepency for all n
- Instead of a random sequence, we can start with something uniform, and then start filling in the gaps
- There are various deterministic sequences that obtain low discrepancy

Van der Corput sequence

- Exercise 2.13
- $\pi(b_{n-1} \dots b_0) = 0.b_0 b_1 \dots b_{n-1}$.
- The sequence $\{\pi(1), \pi(2), \pi(3), \dots\}$ is the van der Corput sequence.
- Example of a deterministic uniform distribution
- This coincides with the 'uniform distribution then fill up gaps' for $d = 1$!

Example van der Corput sequence

- First elements are $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \ldots$



Sampling for low discrepancy

- We want to extend this idea to $d > 1$.
- Prime number p, base-p expansions. Change of notation:

$$\pi_2((b_{n-1} \ldots b_0)_2) = (0.b_0 b_1 \ldots b_{n-1})_2.$$

- This is for binary representation, but we can do this for arbitrary base p:

$$\pi_p((a_{n-1} \ldots a_0)_p) = (0.a_0 a_1 \ldots a_{n-1})_p.$$

- More explicitely:

$$\pi_p\left(\sum_{i=0}^{n-1} a_i p^i\right) = \sum_{i=0}^{n-1} a_i p^{-i-1}.$$

Halton sequence

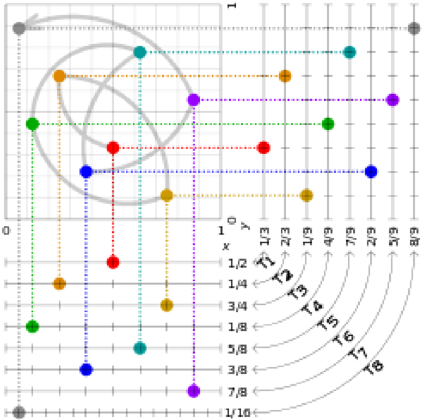
- Let $p_1, \ldots, p_d$ be the first d primes (i.e. $2, 3, 5, 7, 11, \ldots$).
- *Halton* sequence is:

$$\vec{x}_j = \left(\pi_{p_1}(j), \pi_{p_2}(j), \ldots, \pi_{p_d}(j)\right)^T.$$

- Note that this is different from the 'uniform then fill gaps' idea!

Example Halton sequence

$$\left\{ \left(\frac{1}{2}, \frac{1}{3}\right), \left(\frac{1}{4}, \frac{2}{3}\right), \left(\frac{3}{4}, \frac{1}{9}\right), \left(\frac{1}{8}, \frac{4}{9}\right), \left(\frac{5}{8}, \frac{7}{9}\right), \left(\frac{3}{8}, \frac{2}{9}\right), \left(\frac{7}{8}, \frac{5}{9}\right), \left(\frac{1}{16}, \frac{8}{9}\right), \left(\frac{9}{16}, \frac{1}{27}\right), \ldots \right\}$$



Halton discrepancy

- As we have seen, for Monte Carlo the (expected) error (and discrepency) is of
$$\mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$
- For Halton we instead have (deterministically)
$$\mathcal{O}\left(\frac{\log^d(n)}{n}\right).$$
- This is almost a quadratic improvement!

Exercise 3.8

- Last week, we already needed code to generate d-dimensional points
- Implement the Halton sequence in d-dimensions:
  - How does this tie into your RNG code?
- Find the volume of the d-dimensional sphere using
  1. Random sequence
  2. Halton sequence
- Plot the error for both methods