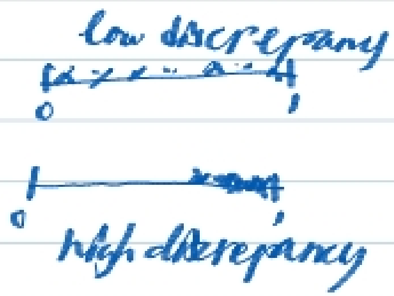
8 Low discrepancy numbers



Low discrepancy numbers: numbers that are "evenly spread"

How is it with $R(0,1)$ random numbers?

CLT: $d=1$, $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ approximately $N(0,1)$

d = dimension

i.e. $\bar{x}-\mu$ proportional to $1/\sqrt{n}$.

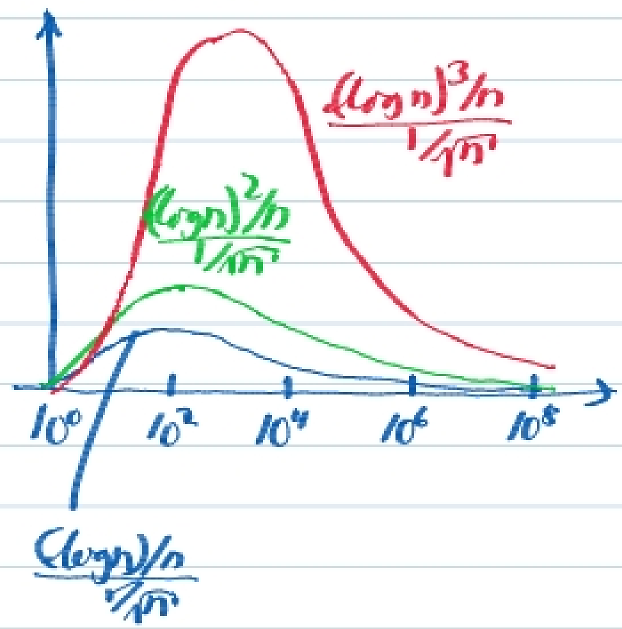
$\bar{x}-\mu$ proportional to $1/\sqrt{n}$ also for $d>1$

For low discrepancy numbers $\bar{x}-\mu$ proportional to $c(d)(\log n)^d/n$.

For fixed d,

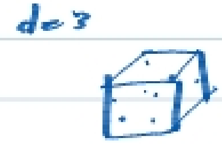
$$\frac{(\log n)^d/n}{1/\sqrt{n}} \to 0 \quad \text{as } n\to\infty$$

but the convergence will be slower as d increases.



8.1 Discrepancy

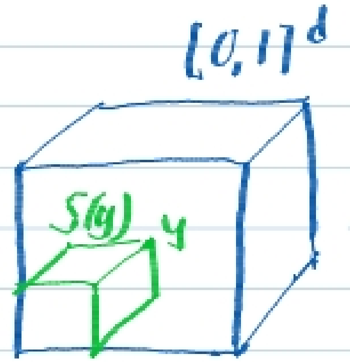
We want to quantify discrepancy/inhomogenity/spread.



Given N d-dimensional numbers $r_1, \dots, r_N$ in $[0,1]^d$.
They may be for example pseudo random numbers.

L^2-discrepancy norm

$$T_N^{(d)} = \left(\int_{[0,1]^d} \left(\frac{n_{S(y)}}{N} - \prod_1^d y_i \right)^2 dy \right)^{1/2},$$



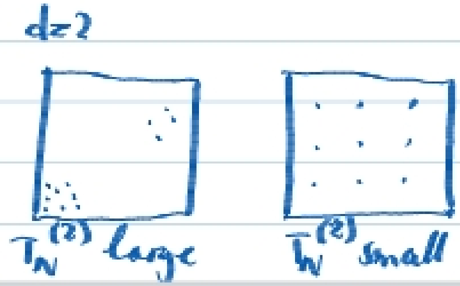
$$n_{S(y)} = \sum_1^N \mathbb{1}\{r_i \in S(y)\},$$

the number of points of $r_1, \dots, r_N$ that are in $S(y)$,

$$S(y) = [0, y_1) \times \dots \times [0, y_d) \subseteq [0,1]^d$$

subhypercube in $[0,1]^d$,

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} \in [0,1]^d$$

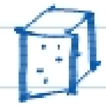


The more evenly spread the data are, the smaller $T_N^{(d)}$.

L^∞ discrepancy norm

$$D_N^{(d)} = \sup_{y\in[0,1]^d} \left| \frac{n_{S(y)}}{N} - \prod_1^d y_i \right|.$$

Clearly $D_N^{(d)} \geq T_N^{(d)}$. $D_N^{(d)}$ hard to evaluate. $T_N^{(d)}$ can be computed explicitly.

For $r_1, \dots, r_N$ $R([0,1]^d)$ (uniformly distributed in $[0,1]^d$), 

$$\mathbb{E}\left[(T_N^{(d)})^2\right] = \frac{1}{N}\left(2^{-d} - 3^{-d}\right).$$

Koksma-Hlawka inequality (Glasserman '03: MC methods and Financial Engineering):

$$\left| \int_{[0,1]^d} g(x)\,dx - \frac{1}{N}\sum_1^N g(r_i) \right| \leq V(g)\, D_N^{(d)},$$

$V(g)$ the variation of g. $d=1$: $V(g) = \int_0^1 |g'(y)|\,dy$ (does not depend on $r_1, \dots, r_N$)

$d=1$: For any sequence $r_1, \dots, r_N$:

$$D_N^{(1)} \geq c_1 \frac{\log N}{N}, \quad \text{equality if } r_1 = \frac{1}{N+1}, \dots, r_N = \frac{N}{N+1}.$$

This can motivate why better to use quadrature rules (ex Simpson's approximation) than MC-estimate for approximation of $\int_{[0,1]^d} g(x)\,dx$ for $d=1$.

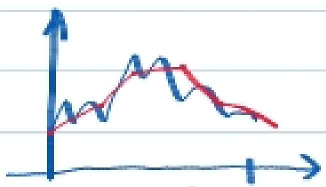
For $d \geq 2$ it is (still) generally believed that for any sequence $r_1, \dots, r_N$ in $[0,1]^d$,

$$D_N^{(d)} \le C_d (\log N)^d / N.$$

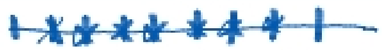
A sequence $r_1, \dots, r_N$ is called a low discrepancy sequence if $D_N^{(d)} \le C_d \frac{(\log N)^d}{N}$.

<u>Discussion point</u> (Asmussen '07, Stochastic simulation)

To compute $\int_{[0,1]^d} g(x)\,dx$ we have at least three methods,



quadrature



quasi MC

$\frac{1}{N}\sum_1^N g(x_i)$

using a low discrepancy sequence

MC

$\frac{1}{N}\sum_1^N g(x_i)$

using pseudo random numbers

Rule of thumb:

	d small	quadrature best
$d \approx 12, 15$?	d intermediate	quasi MC best
	d large	MC best

For approximation of $\int_A g(x)\,dx$ for $A \subseteq \mathbb{R}^d$ we may transform it into $\int_{[0,1]^d} g(x(y))\, h(y)\, dy$