Configurations of points with respect to discrepancy and uniform distribution

Wöden Kusner

Institute for Analysis and Number Theory
Graz University of Technology

March 2016

Abstract

In the theory of uniform distribution, one important measure of the quality of a point set is its discrepancy. This quantifies how well the counting measure of a point set can approximate volume with respect to some collection of regions. For the purposes of Quasi-Monte Carlo integration, we would like to find point sets with low discrepancy.

We'll look at the implementation of some algorithms for explicitly computing discrepancy, analyze some problems related to high quality point sets in the compact setting and discuss some of the open questions that go with them.

Based on various work with Peter Grabner, Johann Brauchart, Alden Walker

Discrepancy

- You may be familiar with discrepancy as a function

$$D : \mathbb{I}^{\mathbb{N}} \times \mathbb{N} \to \mathbb{I}$$

which measures the irregularity of the distribution of a sequence $\{x_i\}_{i=1}^{\infty}$ when truncated to length N

$$D_N(\{x_i\}_{i=1}^{\infty}) := \sup_{0 \leq a < b \leq 1} \left| \frac{\#(\{x_i\}_{i \leq N} \cap [a,b])}{N} - (b-a) \right|.$$

- This notion can be generalized in many ways to other spaces. Identify the endpoints of the interval and including all connected subsets containing the point $\{0 = 1\}$ gives a discrepancy for points in $\mathbb{S}^1$, the *spherical cap discrepancy.*

Discrepancy

We would like low discrepancy sequences in order to replace uniformly distributed random points.

For reasonable $f : \mathbb{I} \to \mathbb{R}$, the error in the approximate integral

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) + err = \int_0^1 f(x)dx$$

is bounded by the discrepancy up to a factor of the total variation of f, a version of the Koksma-Hlawka inequality.

Spherical Cap Discrepancy

On a unit sphere $\mathbb{S}^d$ in $\mathbb{R}^d + 1$ with uniform measure σ and an spherical cap C in $\mathbb{S}^d$, the *local spherical cap discrepancy* of a set X_N of N distinct points is given by

$$D_C[X_N] := |\sigma(C) - \frac{1}{N}\#(X_N \cap C)]|.$$

This is the normalized difference between the expected and the actual number of points found in cap C

$$D_C[X_N] := \frac{1}{N}|\mathbb{E}[\#(X_N \cap C)] - \#(X_N \cap C)]|$$

Remark

Integrating over the space of caps of a fixed size, we can define

$$\mathbb{V}_{\sigma(C)}[X_N] = \int_{\mathbb{S}^d} D_C^2 \, d\sigma.$$

Spherical Cap Discrepancy

These give several natural measures of the quality of X_N.

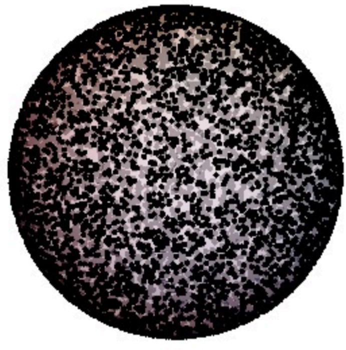
- The classical *spherical cap discrepancy* given as

$$D(X_N) := \sup_{C \in \mathbb{S}^d \times [-1,1]} D_C[X_N].$$

- The L^2*-discrepancy*, given as

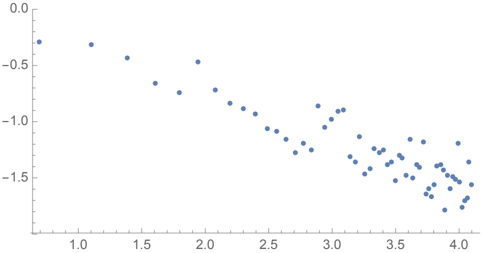
$$D_{L^2}[X_N] := \sqrt{\int_{-1}^{1} \mathbb{V}_C[X_N] \, \mathrm{d}t}.$$

- It is know there exist good point sets X_N on the sphere satisfying $D(X_N)$ bounded by $O(N^{-3/4}\sqrt{\log N})$ and a lower bound for all points sets of $\Omega(N^{-3/4})$.
- Random are bounded by $O(N^{-1/2}\sqrt{\log^2 N})$ *a.s.*.



10000 psudo-random points

Spherical Cap Discrepancy

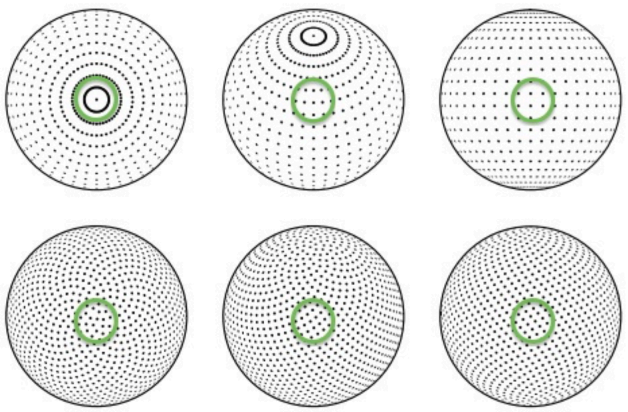


$\log(N)$-$\log(D)$ plot for Random Point Discrepancy

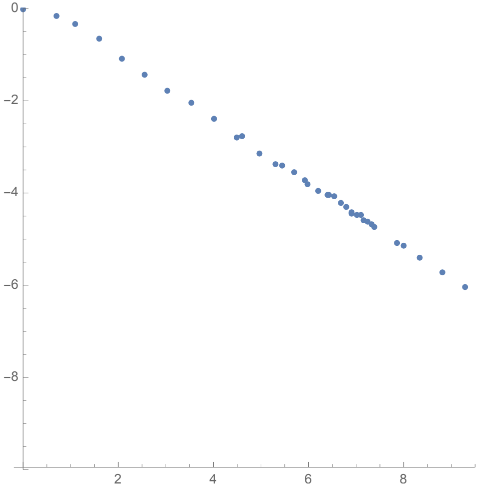
Spherical Cap Discrepancy

- It is know there exist good point sets X_N on the sphere satisfying $D(X_N)$ bounded by $O(N^{-3/4}\sqrt{\log N})$ and a lower bound for all points sets of $\Omega(N^{-3/4})$.
- Random points are bounded by $O(N^{-1/2}\sqrt{\log^2 N})$ *a.s.*.
- It is conjectured that the Fibonacci points on the sphere have low discrepancy, perhaps optimal, but they are only known to be as good as random points.
  (Aisleitner, Brauchart, Dick)

Spherical Cap Discrepancy



Spiral Points vs. Latitude/Longitude (González, 09)



$\log(N)$-$\log(D)$ plot for Spiral Point Discrepancy

Computation

- Computing the L^2-discrepancy of a set of points is fairly easy. It follows from the *Stolarsky Invariance Principle*, which roughly states

$$\frac{1}{N^2}\sum_{i\neq j}|x_i - x_j| + D_{L^2}[X_N] = C_d$$

- Based on results for discrepancy in boxes, it is likely that computing the spherical cap discrepancy is NP-hard.

Remark

There is still a nice algorithm for computing the discrepancy.

Computation

In the case of the star discrepancy, an algorithm was described by Niederreiter that exactly computes the star discrepancy.

- Note that the discrepancy function achieves local extrema when the measuring sets are "captured."
- This is a finite set, so enumerate and take the maximum.

A similar approach works for spherical caps. It is a brute force approach that is exponential in dimension, but it is polynomial in the number of points. To "capture" a cap on $\mathbb{S}^d$ requires $d+1$ points or fewer, and $\binom{N}{d+1} = O(N^{d+1})$.

Comparing points gives an extra factor of N, so the runtime is $O(N^{d+1})$. There are possible improvements with clever sorting, at some cost in memory (cf. Dobkin et. al.).

Computation

- This algorithm illustrates some of the issues with large constants in polynomial time algorithms, and with the implementation. Just optimizing the code for the brute force algorithm reduced the time for several thousand points from several hundred years to a few hours.
- Ways to reduce the extra factor of N at the cost of some memory: Sort with respect to a known set using the triangle inequality. Use a sweep based on pivoting on d points.
- It is also a massively parallel problem. In principle, we can compute discrepancy for millions of points relatively quickly.

Optimization

This algorithm hints at some optimization methods, but it really seems to depends on the labeled partitioning of N. For a simplex, it works out nicely.

Four points on $\mathbb{S}^2$ can be optimized by hand (almost). For a lower bound, consider that discrepancy is symmetric across a cap C, which allows one to pass its supporting points across the boundary.

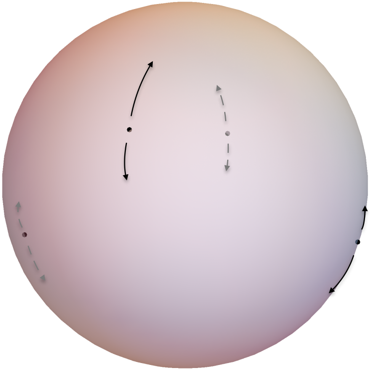
This gives an easy system to solve:

$$|\mathrm{Vol}(C) - 1/4| = |1 - \mathrm{Vol}(C)| \implies \mathrm{Vol}(C) = 5/8,\ D(X_N) > 3/8.$$

Remark

This does not hold for the regular simplex.

Optimization



A very simple family of four points.

Optimization

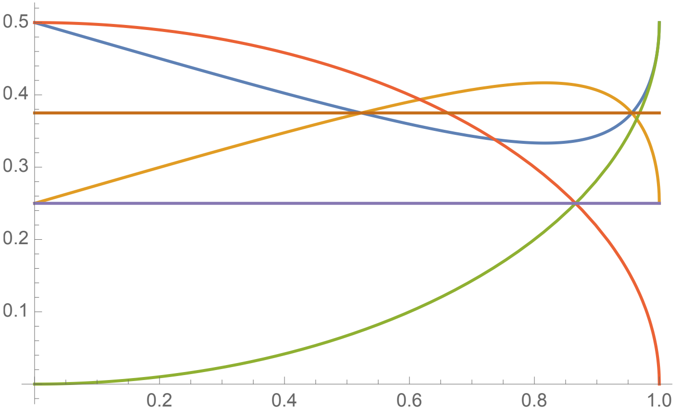
We can symbolically solve to give 3-point discrepancies of

$$\frac{512 + 19\sqrt{466 - 38\sqrt{105}} + \sqrt{210\left(233 - 19\sqrt{105}\right)}}{2048}$$

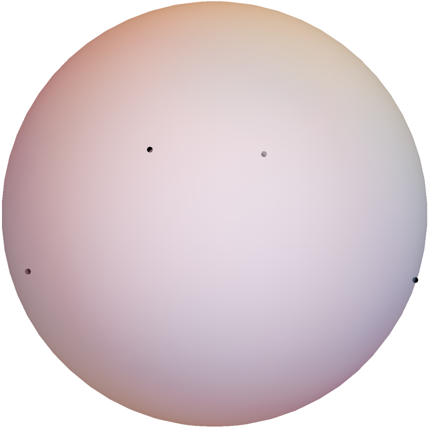
$$\frac{1024 - 19\sqrt{466 - 38\sqrt{105}} - \sqrt{210\left(233 - 19\sqrt{105}\right)}}{2048}$$

(which also happen to be 3/8!)

Optimization



Tetrahedra family vs. 3-point and 2-point discrepancies



A minimal discrepancy set


Thank
You!