Part 2: Discrepancy of stratified samples from partitions of the unit cube

arXiv:2008.12026 (joint work with M. Kiderlen)

Randomized quasi-Monte Carlo I

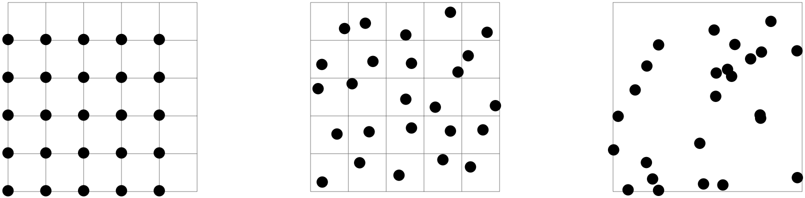
Randomized quasi-Monte Carlo (RQMC) sampling is a popular method to randomize deterministic point sets.

RQMC basically takes a deterministic QMC point set as an input and uses a randomisation technique (e.g. a random shift modulo 1 or a so-called digital shift etc) to generate a new point set, which can be shown to have improved uniform distribution properties (on average) compared to Monte Carlo samples, while still enjoying the advantages of being 'random' in theoretical analysis.

Randomized quasi-Monte Carlo II

The starting point for our work, is a classical and basic RQMC technique (already discussed by **Haber** in 1966).

Jittered sampling for $N = m^d$ combines the simplicity of grids with uniform random sampling by partitioning $[0, 1]^d$ into m^d axis-aligned congruent cubes and placing a random point inside each of them.



Stratified sampling

We take a systematic look at jittered sampling and its extension based on more general partitions $\mathbf{\Omega} = (\Omega_1, \ldots, \Omega_N)$ of $[0,1]^d$.

We consider stratified sampling, where $[0,1]^d$ is partitioned into N subsets $\Omega_1, \ldots, \Omega_N$ and the ith point in $\mathcal{P}$ is chosen uniformly in the ith set of the partition (and stochastically independent of the other points), $i = 1, \ldots, N$.

As ground model for comparison we consider the set of Monte Carlo samples $\mathcal{P}_N$ consisting of N i.i.d. (independent and identically distributed) uniform random points in the unit cube.

Stratified sampling II

Let $d \geq 2$ be given. We consider partitions $\mathbf{\Omega} = \{\Omega_1, \ldots, \Omega_N\}$ of the unit cube in $\mathbb{R}^d$ into N Lebesgue-measurable sets, i.e.

$$[0,1]^d = \bigcup_{i=1}^{N} \Omega_i,$$

and the sets do not overlap in the L^1-sense, so $\Omega_i \cap \Omega_j$ is a Lebesgue-null set for all $i \neq j$ in $\{1, \ldots, N\}$.

Any such partition gives rise to an N-element **stratified sample** $\mathcal{P}_{\mathbf{\Omega}}$ of N random points derived from the partition by picking a random uniform point from each Ω_i in a stochastically independent manner.

Relation to literature

Note that sequences of partitions that can be used in stratified sampling are more general than those in Kakutani's splitting procedure and its variants.

Apart from the obvious difference that these procedures restrict considerations to $d = 1$, the partitions in the present paper need not be nested.

This means that the partition in step $N + 1$ is not necessarily obtained as a refinement of the partition in step N.

We discuss the technical differences at length in our preprint on arXiv!

Questions I

It was shown that the asymptotic order of the star-discrepancy of a point set obtained from jittered sampling is $\mathcal{O}(N^{-\frac{1}{2}-\frac{1}{2d}})$.

Thus, taking partitions can significantly improve the expected discrepancy of (random) point sets in small dimensions $d \geq 2$. We are interested in the following main questions:

1. In which sense are sequences of stratified sample points uniformly distributed as their number N increases? What is the connection to similar notions for partitions in the literature?

Questions II

2. Does stratified sampling yield smaller or larger mean discrepancy than Monte Carlo sampling with N i.i.d. uniform random points? Are there discrepancy notions and assumptions assuring that stratified sampling is strictly better?
3. Is there a best partition for a given N in terms of a chosen mean discrepancy?
4. Is there a simple family of partitions $\{\mathbf{\Omega}^{(N)}\}_{N\geq 1}$ that gives reasonable results for all N and not just for square numbers of points as in the case of classical jittered sampling?
5. Can we improve classical jittered sampling with stratified sampling?

Triangular arrays

Now let $\{\mathbf{\Omega}^{(N)}\}_{N\geq 1}$ be a sequence of finite partitions of the unit cube and let $\mathcal{P}_{\mathbf{\Omega}^{(N)}} = \{\boldsymbol{X}_1^{(N)}, \ldots, \boldsymbol{X}_N^{(N)}\}$ be the stratified sample associated to the Nth partition.

Partitions for different N need not be related to one another; hence the set of all sampling points forms a **triangular array**.

Definition

A triangular array $\widehat{\boldsymbol{x}} = \left(\boldsymbol{x}_1^{(N)}, \ldots, \boldsymbol{x}_N^{(N)}\right)_{N\in\mathbb{N}}$ with points in $[0,1]^d$ is said to be **uniformly distributed**, if for every cube $[\boldsymbol{x}, \boldsymbol{y}[\subset [0,1[^d$ we have

$$\lim_{N\to\infty} \frac{\#\left(\{\boldsymbol{x}_1^{(N)}, \ldots, \boldsymbol{x}_N^{(N)}\} \cap [\boldsymbol{x}, \boldsymbol{y}[\right)}{N} = \lambda\big([\boldsymbol{x}, \boldsymbol{y}[\big).$$

It follows from the strong law of large numbers that a sequence of Monte Carlo samples $(\mathcal{P}_N)$ is almost surely uniformly distributed.

A corresponding statement for the sequences from triangular arrays is based on the **strong law of large numbers for triangular arrays**.

Theorem

Consider a sequence of partitions $\{\boldsymbol{\Omega}^{(N)}\}_{N\geq 1}$ *of* $[0,1]^d$*, with* $\boldsymbol{\Omega}^{(N)} = (\Omega_1^{(N)}, \ldots, \Omega_N^{(N)})$ *consisting of Lebesgue-sets with positive content and let* $\boldsymbol{X}^{(N)} = (\boldsymbol{X}_1^{(N)}, \ldots, \boldsymbol{X}_N^{(N)})$ *be the vector of stratified sampling points based on* $\boldsymbol{\Omega}^{(N)}$*. Then the* **triangular array** $\widehat{\boldsymbol{X}} = \left(\boldsymbol{X}^{(N)}\right)_{N\in\mathbb{N}}$ *is* **almost surely uniformly distributed** *if and only if*

$$\lim_{N\to\infty} \sum_{i=1}^{N} \frac{1}{N|\Omega_i^{(N)}|} |\Omega_i^{(N)} \cap [\boldsymbol{x}, \boldsymbol{y}[| = \lambda([\boldsymbol{x}, \boldsymbol{y}[)$$

for all cubes $[\boldsymbol{x}, \boldsymbol{y}[\subset [0,1]^d$.

In particular, if $\{\mathbf{\Omega}^{(N)}\}_{N\geq 1}$ is a sequence of finite partitions of the unit cube such that all partitions are **equivolume**, then $\lambda\left(\Omega_i^{(N)}\right) = 1/N$, so the condition is satisfied even without taking the limit.

As a consequence, sequences of equivolume partitions are **uniformly distributed**.

$\mathcal{L}_p$-discrepancy

Given a finite set $\mathcal{P} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ of N points in $[0,1]^d$ one way to quantify how well-spread these points are, is to calculate the **$\mathcal{L}_p$-discrepancy**

$$\mathcal{L}_p(\mathcal{P}) := \left(\int_{[0,1]^d} \left| \frac{\#\left(\mathcal{P} \cap [0, \boldsymbol{x}[\right)}{N} - \lambda\big([0, \boldsymbol{x}[\big) \right|^p \delta \boldsymbol{x} \right)^{1/p},$$

of $\mathcal{P}$, in which $1 \leq p < \infty$, i.e. the $\mathcal{L}_p$ norm of the discrepancy function.

The strong partition principle I

Certainly, a stratified sample need not be better than a Monte Carlo sample.

Consider for instance a partition $\mathbf{\Omega}$ with N sets in $[0,1]^2$ where the $N-1$ **partitioning sets $\Omega_1, \ldots, \Omega_{N-1}$ are all subsets of $[\delta, 1]^2$** with some $\delta \in]0,1[$. Then the mean $\mathcal{L}_2$-discrepancy satisfies

$$\mathbb{E}\mathcal{L}_2(\mathcal{P}_{\mathbf{\Omega}})^2 \geq \mathbb{E}\int_{[0,\delta]^2} \left(\frac{1_{[0,\boldsymbol{x}[}(X_N^{(N)})}{N} - \lambda([0,\boldsymbol{x}[) \right)^2 \delta\boldsymbol{x}$$
$$= \frac{\delta^4}{4N} + \frac{\delta^6}{9}\left(1 - \frac{2}{N}\right) \geq \frac{\delta^4}{4N} > \mathbb{E}\mathcal{L}_2(\mathcal{P}_N)^2,$$

for all $\delta > (5/9)^{1/4} \approx 0.86$ and $N \geq 2$.

The strong partition principle II

In contrast, stratified samples from **equivolume** partitions are never worse than Monte Carlo samples in terms of the mean $\mathcal{L}_2$-discrepancy according to the **Partition Principle** shown by **Steinerberger & P.**

We **strengthen this result** in two directions showing firstly that stratified samples from equivolume partitions are **strictly** better, and secondly that $\mathcal{L}_2$-discrepancy can be replaced by **$\mathcal{L}_p$-discrepancy** with arbitrary $p > 1$.

Theorem (Strong Partition Principle)

For any equivolume partition $\mathbf{\Omega}$ with $N \geq 2$ sets we have

$$\mathbb{E}\mathcal{L}_p(\mathcal{P}_{\mathbf{\Omega}})^p < \mathbb{E}\mathcal{L}_p(\mathcal{P}_N)^p$$

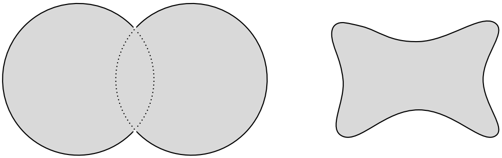
for all $p > 1$.

Partitions with best average discrepancy I

We assume that there is $r > 0$ such that the sets $\Omega_1, \ldots, \Omega_N$ of the partition have **reach** at least r.

This means that for any point $\boldsymbol{x}$ with distance less than r from Ω_i there is a **unique** closest point to $\boldsymbol{x}$ in Ω_i, $i = 1, \ldots, N$.

This class is very general and contains for instance all closed convex sets (which actually have infinite reach). Also, any compact closed set in $\mathbb{R}^2$ whose boundary is a piecewise C^2-curve of $\mathbb{R}^d$ such that its finitely many vertices are 'convex', has positive reach.



Partitions with best average discrepancy II

Theorem

Let $r > 0$, $N \in \mathbb{N}$, and $p \geq 1$ be given. Let $\mathfrak{P}_N(r)$ denote the set of all equivolume partitions *of $[0,1]^d$ into N* sets of reach at least r*. Then there exists (at least) one partition $\mathbf{\Omega}^* \in \mathfrak{P}_N(r)$ that minimizes the mean $\mathcal{L}_p$-discrepancy on $\mathfrak{P}_N(r)$; i.e.*

$$\min_{\mathbf{\Omega} \in \mathfrak{P}_N(r)} \mathbb{E}\mathcal{L}_p(\mathcal{P}_{\mathbf{\Omega}})^p = \mathbb{E}\mathcal{L}_p(\mathcal{P}_{\mathbf{\Omega}^*})^p.$$

A corresponding statement holds true for the mean star-discrepancy.

Partitions with best average discrepancy III

Also smaller classes of partitions can be treated, such as **convex partitions**.

This is relevant for applications, as all sets constituting a convex partition of $[0,1]^d$ are actually polytopes, with their numbers of vertices being uniformly bounded when N is given.

Theorem

Let $N \in \mathbb{N}$, and $p \geq 1$ be given. Let $\mathfrak{P}_N^{\mathrm{conv}}$ denote the set of all **equivolume partitions** *of $[0,1]^d$ into N* **convex sets**. *Then there exists (at least) one partition $\mathbf{\Omega}^* \in \mathfrak{P}_N^{\mathrm{conv}}$ that minimizes the mean $\mathcal{L}_p$-discrepancy on $\mathfrak{P}_N^{\mathrm{conv}}$; i.e.*

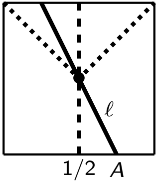
$$\min_{\mathbf{\Omega} \in \mathfrak{P}_N^{\mathrm{conv}}} \mathbb{E}\mathcal{L}_p(\mathcal{P}_{\mathbf{\Omega}})^p = \mathbb{E}\mathcal{L}_p(\mathcal{P}_{\mathbf{\Omega}^*})^p.$$

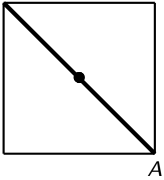
A corresponding statement holds true for the mean star-discrepancy.

Example: Convex equivolume partitions into two sets

Left: Model for all convex partitions into two sets with equal volume.

Right: The partition $\boldsymbol{\Omega}_*^{(2)}$ of this family with the smallest expected discrepancy.





Lemma

For $d = 2$ we have

$$\min_{\boldsymbol{\Omega} \in \mathfrak{P}_2^{\mathrm{conv}}} \mathbb{E}\mathcal{L}_2^2(\mathcal{P}_{\boldsymbol{\Omega}}) = \mathbb{E}\mathcal{L}_2^2(\mathcal{P}_{\boldsymbol{\Omega}_*^{(2)}}) = 0.05.$$

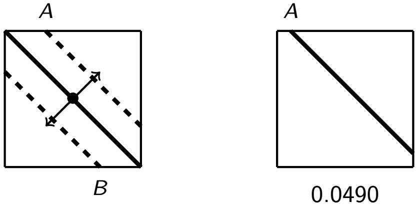
Example: Relaxing the volume constraint

Lemma

We have

$$\min_{v\in[0,\sqrt{2}]} \mathbb{E}\mathcal{L}_2^2(\mathcal{P}_{\mathbf{\Omega}_v^{(2)}}) = \mathbb{E}\mathcal{L}_2^2(\mathcal{P}_{\mathbf{\Omega}_{v^*}^{(2)}}) \approx 0.049,$$

for $v^* = 0.793398\ldots$.

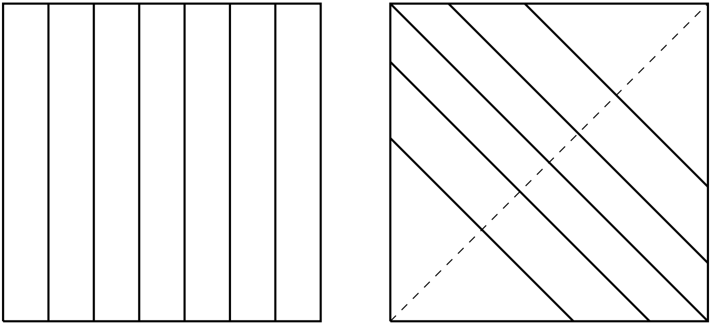


Left: One-parameter model of partitions used in Lemma.

Right: The partition of this family with the smallest expected discrepancy.

Example: Infinite families

Left: Illustration of the vertical strip partition in $[0,1]^2$ for $N = 7$.
Right: Illustration of partition $\mathbf{\Omega}_*^{(6)}$ into equivolume slices that are orthogonal to the diagonal for $N = 6$.



Numerical results

Expected $\mathcal{L}_2$-discrepancy of different point sets, in which N stands for the number of points. The empirical values are calculated as the mean of the discrepancy of 500 samples. We calculated the discrepancy of individual samples with **Warnock's formula**.

$\mathbb{E}\mathcal{L}_2(\cdot)$	$\mathcal{P}_N$	$\mathcal{P}_{\text{vert}}$	$\mathcal{P}_{\mathbf{\Omega}_*^{(N)}}$	$\mathcal{P}_{\text{jit}}$
N			empirical	empirical
50	0.00277778	0.00168889	0.00137637	
100	0.00138889	0.000838889	0.000699558	0.000163637
150	0.000925926	0.000558025	0.000471159	
200	0.000694444	0.000418056	0.000356743	
256	0.000542535	0.000326369	0.000269319	0.0000403301
300	0.000462963	0.000278395	0.000228231	
350	0.000396825	0.000238549	0.000201676	
400	0.000347222	0.000208681	0.000172704	0.0000206345
450	0.000308642	0.00018546	0.000159365	

References

M. Kiderlen, F. Pausinger, **Discrepancy of stratified samples from partitions of the unit cube.** Submitted, arXiv:2008.12026

F. Pausinger, M. Rachh S. Steinerberger, **Optimal Jittered Sampling for two points in the unit square.** Stat. & Prob. Letters (132) 2018, 55-61.

F. Pausinger, S. Steinerberger, **On the discrepancy of jittered sampling.** Journal of Complexity (33) 2016, 199-216.