On further aspects of dispersion

Jan Vybíral
Czech Technical University
Prague, Czech Republic

joint work with **A. Hinrichs** (U Linz, Austria), **J. Prochno** (U Hull, UK/Uni Graz, Austria), **M. Ullrich** (U Linz, Austria)

Monte Carlo & Quasi-Monte Carlo Methods
in Scientific Computing
July 2018, Rennes, France

Outline

Dispersion - once again
- Definition
- Overview of results
- Further directions

Random vs. deterministic
- $\log(d)$ by random constructions
- $\log(d)$ deterministically

k-dispersion
- Definition
- Results

Dispersion: definition

- Let $d \geq 2$ and let $[0,1]^d$ be the unit cube in $\mathbb{R}^d$
- Let $\mathcal{P}_n = \{x_1, \dots, x_n\} \subset [0,1]^d$ be a set of n points
- Let $\mathcal{B}^d_{ax}$ be all boxes in $[0,1]^d$ with sides parallel to coordinate axes
- Then

  $$\mathrm{disp}(\mathcal{P}_n, d) := \sup \Big\{ \mathrm{vol}(\mathbb{B}) : \mathbb{B} \in \mathcal{B}^d_{ax} \text{ with } \#(\mathcal{P}_n \cap \mathbb{B}) = \emptyset \}$$

  is the volume of the largest box not intersecting $\mathcal{P}_n$
- A well-spread point set should have a small dispersion - i.e. no big holes

  $$\mathrm{disp}(n,d) := \inf_{\substack{\mathcal{P}_n \subset [0,1]^d \\ \#\mathcal{P}_n = n}} \mathrm{disp}(\mathcal{P}_n, d).$$

Dispersion: recent results

- Trivial pigeonhole principle:

$$1 \geq \mathrm{disp}(n, d) \geq \frac{1}{n+1}$$

- Aistleitner, Hinrichs, and Rudolf (2017):

$$\mathrm{disp}(n, d) \geq \frac{\log_2(d)}{4(n + \log_2(d))}$$

- Larcher (unpublished):

$$\mathrm{disp}(n, d) \leq \frac{2^{7d+1}}{n}$$

Dispersion: even more recent results

- Rudolf ($n > 2d$, 2018):

$$\mathrm{disp}(n,d) \le \frac{4d}{n}\log_2\left(\frac{9n}{d}\right)$$

- Sosnovec (2018): Randomized point set with at most $c_\varepsilon \log_2(d)$ points in $[0,1]^d$ and dispersion at most $\varepsilon \in (0, \frac{1}{4}]$
- Ullrich, V. (2018): Refining the argument of Sosnovec

$$\#\mathcal{P} \le 2^7 \frac{(1+\log_2(\varepsilon^{-1}))^2}{\varepsilon^2}\log_2(d)$$

i.e.

$$\mathrm{disp}(n,d) \le c\log_2(n)\sqrt{\frac{\log_2(d)}{n}}$$

Dispersion: even more recent results

- Calculation of the dispersion for well-known sets and lattices (Krieg 2018; Temlyakov 2018)
- Essentially - estimates good in n are not optimal in d and vice versa
- Conjecture - at least for a large range of parameters:

$$\mathrm{disp}(n,d) \approx \min\left\{1, \frac{\log_2(d)}{n}\right\}$$

Further directions

- Closing the gaps
- Relations to discrepancy
- Further test sets
- Generalizations
- Deterministic constructions

$\log(d)$: Random constructions

Sosnovec (2018): The points of $\mathcal{P}$ are chosen uniformly from

$$\left\{\frac{1}{2^m}, \frac{2}{2^m}, \dots, \frac{2^m-1}{2^m}\right\}^d = M_m^d \subset [0,1]^d$$

If $\mathrm{vol}(\mathbb{B}) = |I_1| \cdot \dots \cdot |I_d| > \frac{1}{2^m}$, then most (for d large) of the intervals I_j cover all the points $M_m = \{1/2^m, \dots, (2^m-1)/2^m\}$.

This fails, if we choose the points uniformly from the whole $[0,1]^d$!

$\implies$ For every $\mathbb{B}$ with $|\mathbb{B}| > 2^{-m}$, there are “active coordinates” j (whose number does not depend on d!), such that I_j does *not* cover the whole M_m

$\implies$ The infinite number of $\mathbb{B}$'s is divided into finitely many groups according to their “active coordinates” and the position of the corresponding I_j's

$\implies$ Union bound!

$\log(d)$: Deterministic constructions I

We "derandomize" the construction of Sosnovec!
Let $\varepsilon = 2^{-m}$, $M_m := \{1/2^m, \dots, (2^m-1)/2^m\}$ and

$$\Omega_m := \left\{ \mathbb{B} = I_1 \times \dots \times I_d \subset [0,1]^d \,:\, \mathrm{vol}(\mathbb{B}) > \frac{1}{2^m} \right\}.$$

- For $\mathbb{B} \in \Omega_m$ at most $m2^m$ of the intervals $I_j, j \in \mathcal{A}$ do not include all the points from M_m;
- take $z \in I_{j_1} \times \dots \times I_{j_n}$, $\mathcal{A} = \{j_1, \dots, j_n\}$;
- every $x \in M_m^d$ with $x|_{\mathcal{A}} = z$ lies also in $\mathbb{B}$;
- Sosnovec actually constructs randomly a set of points $x^1, \dots, x^N \in \{1, 2, \dots, 2^m - 1\}^d$ such that for every $\mathcal{A} \subset \{1, \dots, d\}$ with $|\mathcal{A}| = m2^m$ the set of restrictions $x^1|_{\mathcal{A}}, \dots, x^N|_{\mathcal{A}}$ contains all $(2^m-1)^{m2^m}$ possible values.

$\log(d)$: Deterministic constructions II

- M. Naor, L. J. Schulman, and A. Srinivasan, Splitters and near-optimal derandomization, 1995
- (n,k)-**universal sets:** a set $T \subset \{0,1\}^n$, such that for any index set $S \subset \{1,2,\dots,n\}$ with $|S| = k$, the projection of T on S contains all possible 2^k configurations.
- (n,k,m)-**universal sets:** a set $T \subset \{0,1,\dots,m-1\}^n$, such that for any index set $S \subset \{1,2,\dots,n\}$ with $|S| = k$, the projection of T on S contains all possible m^k configurations.
- If $m = 2$: $(n,k,2)$-universal sets are (n,k)-universal sets
- If $m = 2^\mu$ is dyadic, (n,k,m)-universal sets can be obtained from $(\mu n, \mu k)$-universal sets. Indeed, we represent each $x \in \{0,1,\dots,m-1\}^n = \{0,1,\dots,2^\mu - 1\}^d$ by an $\tilde{x} \in \{0,1\}^{\mu n}$ just by writing each coordinate x_j in the dyadic representation.

$\log(d)$: Deterministic constructions III

- There are deterministic constructions of (n,k)-universal sets of size $2^k k^{O(\log k)}\log(n)$
- Therefore, there is a deterministic construction of an $(n,k,m)=(n,k,2^\mu)$-universal set with size of

$$2^{\mu k}(\mu k)^{O(\log(\mu k))}\log(\mu n)$$
$$= m^k(k\log(m))^{O(\log(k\log(m)))}\log(n\log(m)).$$

- For the derandomization of Sosnovec's proof, we need an $(d, m2^m, 2^m-1)$-universal set
- Size:

$$N = 2^{m^2 2^m}(m^2 2^m)^{O(\log(m^2 2^m))}\log(md).$$

- A logarithmic dependence on d, rather bad dependence in $2^m \approx \varepsilon^{-1}$

$\log(d)$: Deterministic constructions - updates I

- Improved randomized construction, with better dependence in ε^{-1}, given by Ullrich, V. (2018)
- We split Ω_m into

$$\Omega_m(s,p) := \Big\{\mathbb{B} \in \Omega_m : \forall l \quad \frac{s_l}{2^m} < |I_l| \le \frac{s_l+1}{2^m}, \\ p_l - \frac{1}{2^m} \le \inf I_l < p_l\Big\}$$

- $$\mathbb{B}_m(s,p) := \bigcap_{\mathbb{B}\in\Omega_m(s,p)} \mathbb{B} = \prod_{\ell=1}^{d}\Big[p_\ell, p_\ell + \frac{s_\ell - 1}{2^m}\Big].$$

- Let z be uniformly distributed in M_m^d. Then

$$\mathbb{P}\big(z \in \mathbb{B}_m(s,p)\big) \ge \frac{1}{2^{m+4}}\,.$$

$\log(d)$: Deterministic constructions - updates II

- The aim was again to use an union bound.
- We have more subgroups, but better control of a probability that some point lies in all the cubes in the subgroup.
- Final estimate

$$\#\mathcal{P} \leq 2^4 m 2^{2m} \log_2(2^{m+1} d).$$

$\log(d)$: Deterministic constructions - updates III

- More general deterministic constructions, known as k-restriction problems are known
- Our setting fits perfectly
- For $A_m = m2^m$, we get

$$|N| \approx 2^m A_m^7 \log^2(A_m) \log d.$$

- $\implies$ polynomial in $\varepsilon^{-1} = 2^m$ and logarithmic in d

k-dispersion: Definition

- $k \in \mathbb{N}_0$

$$k\text{-}\operatorname{disp}(P_n, d) := \sup\Big\{ \operatorname{vol}(\mathbb{B}) : \mathbb{B} \in \mathcal{B}_{ax}^d \text{ with } \#(P_n \cap \mathbb{B}) \le k \Big\},$$

- $k = 0$ gives the previous notion of (usual) dispersion
- The minimal k-dispersion: infimum over all point sets of cardinality n, i.e.,

$$k\text{-}\operatorname{disp}(n, d) := \inf_{\substack{P_n \subset [0,1]^d \\ \#P_n = n}} k\text{-}\operatorname{disp}(P_n, d).$$

k-dispersion: Results

- The same random point set as in Ullrich, V. (2018)
- Again, splitting Ω_m into $\Omega_m(s,p)$'s
- The union bound at the very end
- Result (Hinrichs, Prochno, Ullrich, V.) ($1 \le k < n/2$):

$$k\text{-}\operatorname{disp}(n,d) \le C \max\left\{\log_2(n)\sqrt{\frac{\log_2(d)}{n}},\ k\,\frac{\log_2(n/k)}{n}\right\}.$$

- Compare with (U.,V.)

$$\operatorname{disp}(n,d) \le c\log_2(n)\sqrt{\frac{\log_2(d)}{n}}$$

- $\implies$ the same order until $k(n,d) \le c\sqrt{n\cdot\log_2(d)}$!

Thank You!