Bounds for Star-Discrepancy with Dependence on the Dimension

Michelle Mastrianni

University of Minnesota

October 21, 2020

Discrepancy Notation

Definition (Local Discrepancy)

If $\mathcal{S}$ is a class of sets in $[0,1]^d$, then for $S \in \mathcal{S}$, the local discrepancy of a point set $P \subseteq [0,1]^d$ with respect to S, with $|P| = n$, is

$$D(P, S) = |S| - n^{-1}|P \cap S|$$

(Note that this is the "normalized" version of discrepancy).

This talk will focus on discrepancy with respect to the set $\mathcal{C}$ of corners (boxes anchored at the origin) in the unit cube $[0,1]^d$. We define a corner $C_x \in \mathcal{C}$ in terms of its largest point $\mathbf{x} = (x_1, \cdots, x_d)$:

$$C_x = \{y \in [0,1]^d : 0 \leq y_i < x_i \text{ for } i = 1, \cdots, d\}$$

Discrepancy Notation

Definition (Star-Discrepancy with respect to $\mathcal{S}$)

$$D^\star_\infty(P, \mathcal{S}) = \sup_{S \in \mathcal{S}} |D(P, S)|$$

Definition (Minimal Star-Discrepancy with respect to $\mathcal{S}$ of an n-point subset in $[0,1)^d$)

$$D^\star_\infty(n, d) = \inf\{D^\star_\infty(P, \mathcal{S}) : P \subset I^d, |P| = n\}$$

Discrepancy results can also be stated in terms of the 'inverse discrepancy,' which gives the minimum number of points with discrepancy at most ϵ.

Definition (Inverse Discrepancy)

$$N^\star_\infty(d, \epsilon) = \min\{n : D^\star_\infty(n, d) \leq \epsilon\} = \min\{|P| : P \subset I^d, D^\star_\infty(P) \leq \epsilon\}$$

Search for Bounds Based on Dimension

- The asymptotic behavior of $D_\infty^\star(P, \mathcal{C}_d)$ (where $\mathcal{C}$ is the class of corners in $[0,1]^d$) is of order at most $n^{-1}\log(n)^{d-1}$: points which achieve this are called low-discrepancy points

- For some applications, the dimension d may be huge

- Then the usual discrepancy bounds are of no help as $n^{-1}\log(n)^{d-1}$ is increasing for $n \leq \exp(d-1)$: in practice, $\exp(d-1)$ is prohibitively large

Vapnik-Cervonenkis Classes

Definition (VC Class)

A countable family $\mathcal{F}$ of measurable subsets of X is a VC-class if there is a nonnegative integer v such that

$$|\{A \cap F : F \in \mathcal{F}\}| < 2^{v+1}$$

for any point subset $A \subset X$ with $|A| = v + 1$. The smallest such v is called the VC-dimension of $\mathcal{F}$.

In other words, $\mathcal{F}$ is a VC-class if there is some integer v such that any subset of cardinality $v + 1$ cannot be "shattered" by sets in $\mathcal{F}$.

Definition (Shattering)

$\mathcal{F}$ is said to shatter a point set A, $|A| = n$, if

$$|\{A \cap F : F \in \mathcal{F}\}| = 2^n$$

Examples of the VC Dimension

Class of Sets	**VC Dimension**	
Semi-Infinite Intervals in $\mathbb{R}$	1	
Closed Intervals in $\mathbb{R}$	2	
Corners in $\mathbb{R}^d$	d	

Examples of the VC Dimension

Class of Sets	**VC Dimension**	
Closed Halfspaces in $\mathbb{R}^d$	$d+1$	
Axis-Parallel Boxes in $\mathbb{R}^d$	$2d$	
Disks in $\mathbb{R}^2$	3	
Balls in $\mathbb{R}^d$	$d+1$	
Convex Sets in $\mathbb{R}^d$	∞	

Important Inequality for VC Classes

Lemma (Sauer-Shelah Lemma)

$$|\{A \cap F : F \in \mathcal{F}\}| \leq \sum_{i=0}^{v} \binom{|A|}{i} = O(|A|^v)$$

(In other words: if the VC-dimension of $\mathcal{F}$ is v, then $\mathcal{F}$ can consist of at most $O(|A|^v)$ sets).

Proof: Induction on n+v (where $|A| = $ n)

Covering Numbers

Definition (Covering Number)

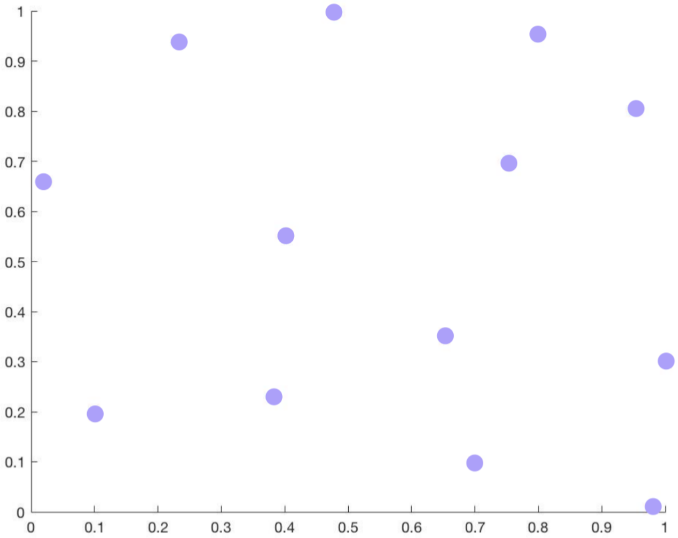
The covering number $\mathcal{N}(M, d, \epsilon)$ of a metric space (M, d) is the smallest number of closed ϵ-balls

$$B(x, \epsilon) := \{y \in M : d(x, y) \leq \epsilon\}$$

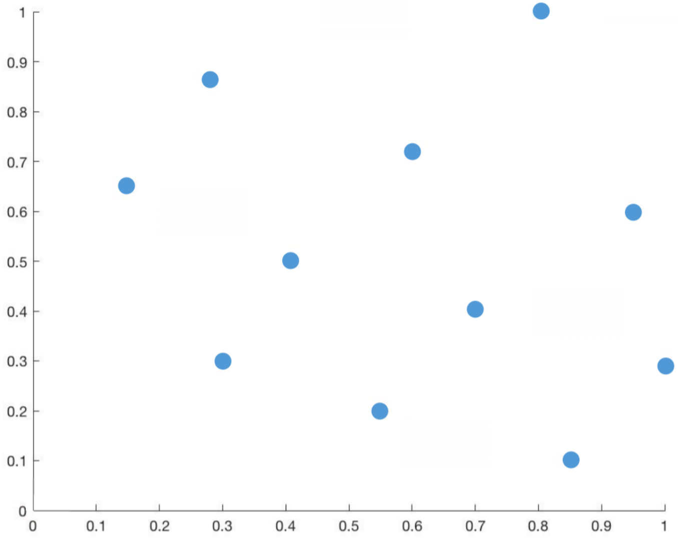
that cover M. If there is no finite cover we set $\mathcal{N}(M, d, \epsilon) = \infty$.

There are nice connections between covering/packing numbers and VC-classes.

Covering Numbers: ℓ^2 Distance



Covering Numbers: Symmetric Difference of Corners



VC-Classes and Discrepancy

The proofs we will discuss are heavily dependent on the fact that the set of corners in $[0,1]^d$ is a VC-class of dimension d. Some important features that make VC-classes interesting/natural to study in discrepancy theory:

- Low complexity
- As discussed above, known geometric properties (i.e. connections with covering numbers)
- Direct implications for combinatorial discrepancy

Upper Bounds for Discrepancy Based on Dimension

Theorem (Heinrich, Novak, Wasilkowski, Wozniakowski, '01, Thm 1)

For all $n, d \in \mathbb{N}$,

$$D_\infty^\star(n,d) \leq 2\sqrt{2}n^{-1/2}\left(d\log\left(\lceil\frac{dn^{1/2}}{(2\log 2)^{1/2}}\rceil + 1\right) + \log 2\right)^{1/2}$$

Theorem (Heinrich, Novak, Wasilkowski, Wozniakowski, '01, Thm 3)

There is a positive number c such that for all $n, d \in \mathbb{N}$,

$$D_\infty^\star(n,d) \leq cd^{1/2}n^{-1/2} \ (\text{and } n_\infty^\star(d,\epsilon) \leq \lceil c^2 d\epsilon^{-2}\rceil).$$

Theorem 1 above can be proven using basic features of empirical process theory, while the proof of Theorem 3 requires deeper results that make use of the fact that the class of corners $\mathcal{C}$ is a VC-class of dimension d.

Hoeffding's Inequality vs. Bernstein's Inequality

For $X_1, \cdots, X_n$ i.i.d random, $\mathbb{E}[X_i] = 0$, $|X_i| \leq 1$ for each i, we have

Hoeffding's Inequality

$$\mathbb{P}(|\sum_{i=1}^{n} X_i| \geq t) \leq \exp\big(-2t^2/n\big)$$

Bernstein's Inequality

$$\mathbb{P}(|\sum_{i=1}^{n} X_i| \geq t) \leq 2\exp\Big(-\frac{t^2/2}{\sum_{i=1}^{n}\mathbb{E}[X_i^2] + t/3}\Big)$$

Thus, Bernstein's inequality gives good bounds for random variables with low variance.

Proof of Theorem 1 (weak upper bound)

- Take $\delta > 0$ and replace set of corners with a finite set Γ_m, the equidistant grid on $[0,1]^d$ with mesh-size $1/m$, where $m = \lceil d/\delta \rceil$
- Then, we can show that the supremum in the definition of $\star$-discrepancy can be replaced with the maximum over the finite set Γ_m with a possible decrease of the $\star$-discrepancy by δ:

$$D_\infty^\star(\mathbf{p}_1, \cdots, \mathbf{p}_n) \leq \max_{\mathbf{x} \in \Gamma_m} \left| x_1 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{C_\mathbf{x}}(\mathbf{p}_i) \right| + \delta$$

Proof of Theorem 1 (weak upper bound)

- Let $\tau_1, \cdots, \tau_n$ be iid uniform random variables and take
$$\xi_{\mathbf{x}}^{(i)} = x_1 \cdots x_d - 1_{[0,\mathbf{x})}(\tau_i), \qquad i = 1, \cdots, n$$
- Obtain, via Hoeffding: $\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n \xi_x^{(i)}\right| \geq \delta\right) \leq 2\exp(-\delta^2 n/2)$
- Then $\mathbb{P}(D_\infty^\star(\tau_1, \cdots, \tau_n) \leq 2\delta) \geq \mathbb{P}(\max_{x\in\Gamma_m}\left|n^{-1}\sum_{i=1}^n \xi_x^i\right| \leq \delta)$
$= 1 - \mathbb{P}(\max_{x\in\Gamma_m}\left|n^{-1}\sum_{i=1}^n \xi_x^i\right| > \delta) \geq 1 - 2(m+1)^d \exp(-\delta^2 n/2)$
- This is strictly positive for $\delta > \delta_0$, where
$$\delta_0^2 = 2n^{-1}\left(d\log\lceil\frac{dn^{1/2}}{2(\log 2)^{1/2}} + 1\rceil + \log 2\right)$$
- Hence, for any $\delta > \delta_0$ there are points $\tau_1, \cdots, \tau_n$ such that $D_\infty^\star(\tau_1, \cdots, \tau_n) \leq 2\delta$. Thus $D_\infty^\star(n, d) \leq 2\delta_0$ and the proof is complete

Sketch of Proof of Theorem 3 (strong upper bound)

To obtain the stronger upper bound $D_\infty^\star(n,d) \leq cd^{1/2}n^{-1/2}$, we use that the set of corners in $[0,1]^d$ is a VC-class of dimension d. Central to the proof is the following VC-type inequality from Talagrand (combined with a result of Haussler on covering numbers), from which the discrepancy bound follows fairly easily.

Theorem (Heinrich, Novak, Wasilkowski, Wozniakowski, '01, Thm 2)

There is a positive number K such that for each countable VC-class $\mathcal{F}$ (subsets of X) and for each probability measure P on X, the following holds: For all $s \geq Kv(\mathcal{F})^{1/2}$ and all natural n,

$$\mathbb{P}\left(\omega : \sup_{F\in\mathcal{F}} \left| P(F) - \frac{1}{n}\sum_{i=1}^{n} 1_F(X_i(\omega))\right| \geq sn^{-1/2}\right) \leq \frac{1}{s}\left(\frac{Ks^2}{v(\mathcal{F})}\right)^{v(\mathcal{F})} e^{-2s^2}$$

where $v(\mathcal{F})$ is the VC-dimension of $\mathcal{F}$.

Sketch of Proof of Theorem 3 (strong upper bound)

$$\mathbb{P}\Big(\omega : \sup_{F\in\mathcal{F}} \Big|P(F) - \frac{1}{n}\sum_{i=1}^{n} 1_F(X_i(\omega))\Big| \geq sn^{-1/2}\Big) \leq \frac{1}{s}\Big(\frac{Ks^2}{v(\mathcal{F})}\Big)^{v(\mathcal{F})} e^{-2s^2}$$

First we show how the discrepancy bound follows from this VC-type inequality:

- Approximate the set of corners $\mathcal{C}$ by

$$\mathcal{C}_{\mathbb{Q}} = \{[0, \mathbf{x}) : \mathbf{x} = (x_1, \cdots, x_d) \in ([0,1] \cap \mathbb{Q})^d$$

- Take $s = \lambda v(\mathcal{C}_{\mathbb{Q}})^{1/2}$, choose λ_0 so that $K\lambda^2 \leq e^{2\lambda^2}$ for all $\lambda \geq \lambda_0$.
- Then for $\lambda > \max(K, \lambda_0, 1)$, the RHS of the VC-inequality above is strictly smaller than 1
- This ensures the existence of a point set P with star-discrepancy less than $cd^{1/2}n^{-1/2}$, where c is some positive number.

Sketch of Proof of Talagrand's VC Inequality

Desired Result: For a class of sets $\mathcal{F}$ and i.i.d random variables such that there exists a $V > 0$ with $X_1, \cdots, X_n$ satisfying $\mathcal{N}(\mathcal{F}, d_P, \epsilon) \leq \left(\frac{V}{\epsilon}\right)^v$ where $d_P(C_1, C_2) = P(C_1 \Delta C_2)$,

$$\mathbb{P}\Big(\omega : \sup_{F \in \mathcal{F}} \Big| P(F) - \frac{1}{n} \sum_{i=1}^{n} 1_F(X_i(\omega)) \Big| \geq s n^{-1/2}\Big) \leq \frac{1}{s} \Big(\frac{Ks^2}{v(\mathcal{F})}\Big)^{v(\mathcal{F})} e^{-2s^2}$$

Note: The covering number condition for VC classes, as used in the proof of the upper bound for discrepancy with respect to corners, is a result of Haussler (1995).

Sketch of Proof of Talagrand's VC Inequality

Idea: first study Gaussian processes $(X_t)_{t\in T}$ for which $N(T, d, \epsilon) \leq (A/\epsilon)^v$ for some constants A, v and $0 < \epsilon < \sigma = (\sup_{t\in T} \mathbb{E}[X_t^2])^{1/2}$

- Obtain bounds on tails of supremum by breaking index set into suitable pieces
- Partitioning the index set T into N pieces each of diameter $\leq a \leq \sigma$, we can obtain a bound on $\mathbb{P}(\sup_{t\in T} X_t \geq u)$ that depends on $\mathbb{E}[\sup_{t\in T} X_t]$.
- We need to control both N, the number of pieces, and the expectation.
- Instead of the crude partition, take an (essentially) dyadic partition: the union of sets $\mathcal{P}_l$, for $p \leq l \leq q$, such that each set in $\mathcal{P}_l$ has diameter $\leq 4^{-l+1}$.

Sketch of Proof of Talagrand's VC Inequality

For the general (non-Gaussian) case:

- Use same partitioning scheme
- Bound on $\mathbb{P}\left(\sup_{t \in T} X_t \geq u\right)$ looks slightly different (a bit more complicated) in non-Gaussian case, but still requires controlling $\mathbb{E}[\sup_{t \in T} X_t]$
- For indices whose corresponding random variables have small variance, it is easy (via Bernstein's inequality) to control the tails and get a good bound on $\mathbb{E}[\sup_{t \in T} X_t]$.
- Then, bound the cardinality on the set of remaining random variables (with large variance): its contribution is controlled.

Aistleitner's Improvement with Dyadic Partitioning

Theorem (Aistleitner, 2011)

$$D^{\star}_{\infty}(n, d) \leq 9.65 d^{1/2} n^{-1/2}$$

- Approximate with a finite set – roughly $N^{d/2}$ sampling points are needed, as in the previously known proofs
- Use direct dyadic partitioning argument rather than "black-box" of VC-inequality
- Express the indicator function 1_{C_x} as a sum of indicator functions for a few sets with large variance and many with small variance. Use Bernstein's inequality (rather than Hoeffding) on the many variables with small variance leads to an improvement in the bound.

Sketch of Aistleitner's Proof

Use Gnewuch's bounds on covering and bracketing numbers.

(Bracketing is a more restricted notion of covering: a finite set Δ of pairs of points of $[0,1]^d$ is an ϵ-bracketing cover if for every pair $(x,z) \in \Delta$, the estimate $\lambda(C_z) - \lambda(C_x) \leq \epsilon$ holds, and for every $y \in [0,1]^d$, there is a pair (x,z) such that $x \leq y \leq z$.)

Gnewuch bounds:

$$\mathcal{N}([0,1]^d, \lambda, \epsilon) \leq (2e)^d (\epsilon^{-1} + 1)^d$$

$$\mathcal{N}_{[]}([0,1]^d, \lambda, \epsilon) \leq 2^{d-1} e^d (\epsilon^{-1} + 1)^d$$

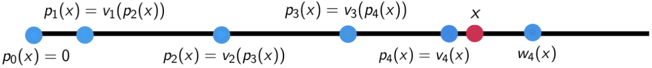
Sketch of Aistleitner's Proof

- d=1 and d=2: theorem is clear (known examples), so assume $d \geq 3$
- Let $K = \lceil (\log_2 n - \log_2 d)/2 \rceil$.
- For $1 \leq k \leq K - 1$, let Γ_k be a 2^{-k} cover for which $|\Gamma_k|$ satisfies Gnewuch covering bound
- Let Γ_K be a 2^{-K} bracketing cover for which $|\Gamma_K|$ satisfies Gnewuch bracketing bound

Fix $x \in [0,1]^d$, pick the pair $(v_K, w_K) = (v_K(x), w_K(x))$ in the bracketing cover for which $v_K \leq x \leq w_K$. Then define

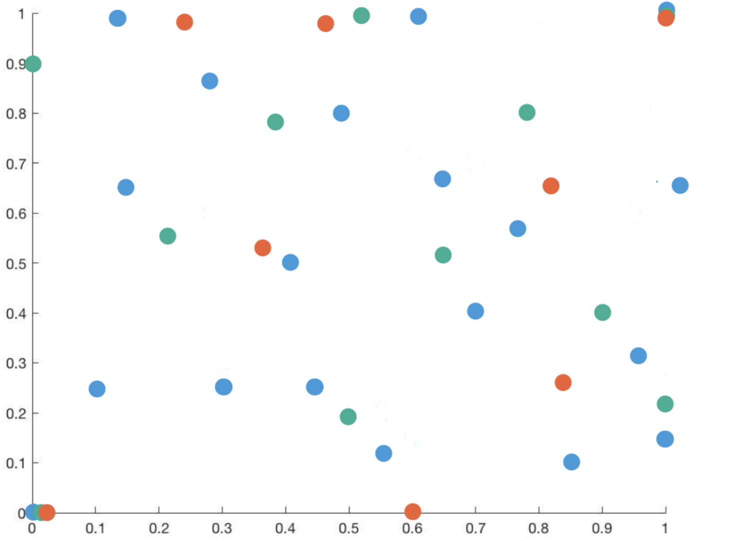
$$p_K(x) = v_K(x)$$
$$p_{K-1}(x) = v_{K-1}(p_K(x))$$
$$p_{K-2}(x) = v_{K-2}(p_{K-1}(x))$$
$$\vdots$$
$$p_1(x) = v_1(p_2(x))$$

Suppose $K = 4$.



where, in the above, $v_k(x) \in \Gamma_k$ for $1 \leq k \leq K - 1$

Sketch of Aistleitner's Proof



Sketch of Aistleitner's Proof

- For $0 \le k \le K-1$, if A_k is the set of all sets of the form $C_{p_{k+1}(x)} \setminus C_{p_k(x)}$, A_k has at most $|\Gamma_{k+1}| \le (2e)^d(2^{k+1}+1)^d$ elements.
- If A_K is the set of all sets of the form $C_{w_K(x)} \setminus C_{p_K(x)}$, A_K has at most $|\Gamma_K| \le 2^{d-1}e^d(2^K+1)^d$ elements.

Let $X_1, \cdots, X_n$ be i.i.d. uniformly distributed random variables on $[0,1]^d$.

Use Bernstein's inequality to bound, for I in A_k with $k \ge 2$

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_I(X_j) - \lambda(I)\Big| > t\Big)$$

(for $k = 0, 1$, the intervals are larger, and Hoeffding's inequality gives a better bound)

Sketch of Aistleitner's Proof

Then...

- Writing $B_k = \bigcup_{I \in A_k} \left(\left| \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_I(X_j) - \lambda(I) \right| > c_k \sqrt{d/n} \right)$
- Choose constants c_k appropriately so that $\mathbb{P}(B_0), \mathbb{P}(B_1), \mathbb{P}(B_2) \leq \frac{1}{4}$, and $\mathbb{P}(B_k) \leq 2^{-k}$ for higher k. Then

$$\sum_{k=0}^{K} \mathbb{P}(B_k) \leq \frac{3}{4} + \sum_{k=3}^{K} 2^{-k} < 1$$

- This ensures the existence of a realization $X_1(\omega), \cdots, X_n(\omega)$, such that $\omega \notin \bigcup_{k=0}^{K} B_k$
- Then for any $x \in [0,1]^d$, use the estimates (with constants) from the finite partition to find c such that

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{C_x}(X_j(\omega)) - x \right| > c\sqrt{d/n} \right)$$

- In the end, $c \approx 9.65$.

Improvements on Constant

A number of improvements have been made on the constant in the previous proof:

- $c \approx 9$: Pasing and Weiss 2018 – uses Aistleitner's method combined with an improvement on the bound for bracketing covers

$$\mathcal{N}_{[]}(d, \epsilon) \leq 2^{d-2} e^d (\epsilon^{-1} + 1)^d + \frac{1}{2}(\epsilon^{-1} + 1)$$

- $c \approx 2.7868$: Doerr 2016
- $c \approx 2.5287$: Gnewuch and Hebbinghaus 2019
- $c \approx 2.4968$: Pasing and Weiss 2020 – again, Aistleitner's method combined with a new improvement on the bound for bracketing covers

$$\mathcal{N}_{[]}(d, \epsilon) \leq \max(1.1^{d-101}, 1) \frac{d^d}{d!} (\epsilon^{-1} + 1)^d$$

Lower Bound

Hinrichs built upon the ideas in HNWW by using the VC-property of the class of corners to improve the lower bound on the star-discrepancy that is polynomial in d/n (like the upper bound).

Theorem (Hinrichs, 2003, Thm 1)

There exist constants $c, \epsilon_0 > 0$ *such that*

$$D^{\star}_{\infty}(n, d) \geq \min(\epsilon_0, cd/n)$$

There is also a theorem of Doerr regarding the lower bound for the expected discrepancy of a random point set:

Theorem (Doerr, 2014, Thm 1)

There is constant k *such that if* $d \leq n$ *and* P *is an* n*-point subset chosen independently and uniformly at random from* $[0, 1]^d$,

$$\mathbb{E}[D^{\star}_{\infty}(P, \mathcal{C}_d)] \geq K\sqrt{d/n}.$$

Lower Bound

Theorem (Hinrichs, 2003, Thm 4)

Let $\mathcal{F}$ be a VC-class of dimension v which is closed under intersections and let $\mathbb{P}$ be a probability measure. Assume that there exists a constant $\kappa > 0$ such that $N(\mathcal{F}, d_{\mathbb{P}}, \epsilon) \geq (\kappa\epsilon)^{-v}$ for all $\epsilon > 0$. Then there exist constants $c, \epsilon_0 > 0$ such that for all n and all $P \subset X$ with $|P| = n$,

$$D^{\star}_{\infty}(P) \geq \min(\epsilon_0, cv/n).$$

Above, $d_{\mathbb{P}} = \mathbb{P}(C_1 \Delta C_2)$.

The proof of this theorem makes use of the Sauer-Shelah Lemma.
As the set of corners is a VC class, Theorem 4 implies Theorem 2 once it is shown there is some $\kappa > 0$ such that $N(\mathcal{C}, d_{\mathbb{P}}, \epsilon) \geq (\kappa\epsilon)^{-d}$ for all $\epsilon > 0$.

Open Problems

- Upper bound: $D^{\star}_{\infty}(n, d) \lesssim (d/n)^{1/2}$

  Lower bound: $D^{\star}_{\infty}(n, d) \gtrsim d/n$

  Close the gap!

- Find points that satisfy the stronger upper bound **constructively** – this is more useful for problems involving high-dimensional integration with Quasi-Monte Carlo methods

Literature

- S. Heinrich, E. Novak, G.W. Wasilkowski, H. Wozniakowski (2001): "The Inverse of the Star-Discrepancy Depends on the Dimension." Acta Arithmetica, Vol. 96, 279-302.
- M. Talagrand (1994): "Sharper bounds for Gaussian and empirical processes." Annals of Probability, Vol. 22, 28–76.
- D. Haussler (1995): "Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Cervonenkis dimension." Journal of Combinatorial Theory, Vol. 69, 217–232.
- C. Aistleitner (2011): "Covering numbers, dyadic chaining and discrepancy." Journal of Complexity, Vol. 27, Issue 6, 531-540.

Literature

- M. Gnewuch (2008): "Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy." Journal of Complexity, Vol. 24(2), 154–172.
- H. Pasing and C. Weiss (2020): "Improving a constant in High-Dimensional Discrepancy Estimates." Publications de L'Institute Mathematique, Vol. 107(121), 67-74. *Preprint*, 2016.
- S. Heinrich (2003): "Some open problems concerning the star-discrepancy." Journal of Complexity, Vol. 19, 416-419.