Multi–Color Discrepancies

Benjamin Doerr*

1 Introduction

We extend the notion of combinatorial discrepancy of hypergraphs to arbitrary numbers of colors. Unless otherwise stated, the following results are joint work with Anand Srivastav and appeared in [DS03]. Let $\mathcal{H} = (X, \mathcal{E})$ denote a finite *hypergraph*, i. e., X is a finite set and $\mathcal{E}$ is a family of subsets of X. Put $n = |X|$ and $m = |\mathcal{E}|$. A c–coloring of $\mathcal{H}$ is a mapping $\chi : X \to M$, where M is any set of cardinality c. Usually, we take $M = [c] := \{1, \ldots, c\}$. The basic idea of measuring the deviation from perfect balance motivates these definitions of the *discrepancy of* $\mathcal{H}$ *with respect to* χ and the *discrepancy of* $\mathcal{H}$ *in* c *colors*:

$$\begin{aligned} \operatorname{disc}(\mathcal{H}, \chi, c) &:= \max_{i \in M, E \in \mathcal{E}} \left| |\chi^{-1}(i) \cap E| - \frac{|E|}{c} \right|, \\ \operatorname{disc}(\mathcal{H}, c) &:= \min_{\chi : X \to [c]} \operatorname{disc}(\mathcal{H}, \chi, c). \end{aligned}$$

Let us start with an example which shows that a hypergraph may have very different discrepancies in different numbers of colors.

Example: Let $k \in \mathbb{N}$ and $n = 4k$. Set $\mathcal{H}_n = ([n], \{X \subseteq [n] \,|\, |X \cap [\frac{n}{2}]| = |X \setminus [\frac{n}{2}]|\})$. Obviously, $\mathcal{H}_n$ has 2–color discrepancy zero, but $\operatorname{disc}(\mathcal{H}_n, 4) = \frac{1}{8}n$.

In fact, such examples exist for nearly any two numbers of colors. Unless c_1 divides c_2, there are hypergraphs $\mathcal{H}_n$ on n vertices having discrepancy $\Theta(n)$ in c_1 colors and zero discrepancy in c_2 colors. This has been investigated in [Doe02].

2 Recursive Coloring

For some 2–color discrepancy results the proofs seem to rely heavily on the fact that only two colors are used. This applies in particular to those where the partial coloring method

*Christian–Albrechts–Universität zu Kiel, 24098 Kiel, Germany.

introduced by Beck [Bec81] is used. A key step there is to construct a low discrepancy partial coloring $\chi := \frac{1}{2}(\chi_1 - \chi_2)$ from two colorings χ_1, χ_2 with $\chi_1(E) \approx \chi_2(E)$ for all $E \in \mathcal{E}$. It is not clear to us how this idea can be extended to c colors.

The idea of recursive coloring is to successively enlarge the number of partition classes. We start with a suitable 2–coloring of X with color classes X_1, X_2 and then iterate this process on the subhypergraphs induced by X_1 and X_2. If the weighted 2–color discrepancies of the induced subhypergraphs are bounded, such a recursive approach can be analyzed, even if c is not a power of 2. For $p \in [0,1]$, denote the discrepancy of $\mathcal{H}$ with respect to the weight $(p, 1-p)$ by $\operatorname{disc}(\mathcal{H}, (p, 1-p)) = \min_{\chi: X \to [2]} \max_{E \in \mathcal{E}} \big| |E \cap \chi^{-1}(1)| - p|E| \big|$.

Theorem 1. *Let* $\operatorname{disc}(\mathcal{H}_0, (p, 1-p)) \leq K$ *for all induced subgraphs* $\mathcal{H}_0$ *of* $\mathcal{H}$ *and all* $p \in [0,1]$. *Then* $\operatorname{disc}(\mathcal{H}, c) \leq 2.0005K$ *holds for all numbers* c *of colors.*

For many classical results, a refinement of the above ideas yields even stronger bounds that decrease for larger numbers of colors. For reasons of space we are not able to state the general result precisely. Roughly speaking, we have that if induced subhypergraphs on n_0 vertices have 2–color discrepancy at most $O(n_0^\alpha)$ for some $\alpha \in]0,1[$, then $\operatorname{disc}(\mathcal{H}, c) = O((\frac{n}{c})^\alpha)$. This gives, among many others, the following bounds. In all cases, the implicit constants do not depend on c.

- **General bound:** $\operatorname{disc}(\mathcal{H}, c) \leq 45\sqrt{\frac{n}{c}\ln(4m)} + 1$.
- **Spencer's [Spe85] six standard deviations:** For all hypergraphs $\mathcal{H}$ having $n = m$ vertices and hyperedges, $\operatorname{disc}(\mathcal{H}, c) = O(\sqrt{\frac{n}{c}\ln(c)})$.
- **Arithmetic progressions:** The hypergraph $\mathcal{A}_n$ of arithmetic progressions in $[n]$ satisfies $\operatorname{disc}(\mathcal{A}_n, c) = O(c^{-0.16} n^{0.25})$ for $c \leq n^{0.25}$. This extends the bound of Matoušek and Spencer [MS96].

3 Linear Methods and Lower Bounds

A second general approach is to mimic the proofs of two-color results. Since the choice of the colors -1 and 1 for two colors allows several powerful arguments, the key problem is to choose a suitable set of colors for the general case. The colors we use are vectors in $\mathbb{R}^c$. We obtain a multi-color analogue of the Beck–Fiala theorem showing that $\operatorname{disc}(\mathcal{H}, c) \leq 2\Delta(\mathcal{H})$ and one of the Bárány-Grunberg theorem. The latter was improved by Bárány in his talk by reducing the multiplicative dependence on the number of colors to a constant.

An analogue of an eigenvalue bound attributed to Lovász and Sós shows $\operatorname{disc}(\mathcal{H}, c) \geq \sqrt{\frac{n(c-1)}{mc^2}\lambda_{\min}(A^\top A)}$, where A is an incidence matrix of A. This can be used to show a lower bound of $0.04\frac{1}{\sqrt{c}}\sqrt[4]{n}$ for the c–color discrepancy of the arithmetic progressions in $[n]$.

For hypergraph having $n = m$ vertices and edges, using a random construction we recently showed that our upper bound in Section 2 is sharp apart from constant factors [Doe]:

Theorem 2. *For all $c \in \mathbb{N}_{\geq 2}$ and $n \geq c \ln c$, there is a hypergraph having n vertices, n hyperedges and c–color discrepancy at least $\frac{1}{40}\sqrt{\frac{n}{c}\ln c}$.*

4 A Correlation Result

In contrast to the (ordinary) c–color discrepancy, there is a strong correlation between the hereditary discrepancies of a hypergraph in different numbers of colors.

Theorem 3. *For any two numbers of colors $c_1, c_2 \in \mathbb{N}_{\geq 2}$ and all hypergraphs $\mathcal{H}$ we have*

$$\mathrm{herdisc}(\mathcal{H}, c_2) \leq 3\, c_1^2\, \mathrm{herdisc}(\mathcal{H}, c_1).$$

Hence $\mathrm{herdisc}(\cdot, c_2) = \Theta_{c_1,c_2}(\mathrm{herdisc}(\cdot, c_1))$. The proof given in [Doe04] actually solves a more general problem, namely it reduces the color rounding problem in c_2 colors to the hereditary discrepancy problem in c_1 colors. We currently have no purely combinatorial proof.

References

[Bec81] J. Beck. Roth's estimate of the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1:319–325, 1981.

[Doe] B. Doerr. Lower bounds for multi-color discrepancies. Submitted.

[Doe02] B. Doerr. Discrepancy in different numbers of colors. *Discrete Mathematics*, 250:63–70, 2002.

[Doe04] B. Doerr. The hereditary discrepancy is nearly independent of the number of colors. *Proceedings of the AMS*, 2004. To appear.

[DS03] B. Doerr and A. Srivastav. Multicolour discrepancies. *Combinatorics, Probability and Computing*, 12:365–399, 2003.

[MS96] J. Matoušek and J. Spencer. Discrepancy in arithmetic progressions. *Journal of the AMS*, 9:195–204, 1996.

[Spe85] J. Spencer. Six standard deviations suffice. *Transactions of the AMS*, 289:679–706, 1985.