METHODS FOR ANALYZING RANDOM DESIGNS - IIAS SPECIAL DAY ON COMBINATORIAL DESIGN THEORY

MICHAEL SIMKIN

Abstract. Keevash's breakthrough proof of the existence and enumeration of designs has opened the door to the analysis of properties of typical designs. We briefly review three methods for so doing. As an example, we apply a method of Kwan to prove that, asymptotically almost surely, a uniformly random Steiner triple system contains $(1 - o(1))\, n^2/24$ Pasch configurations.

1. Introduction

Definition 1.1. An (n, q, r, t)**-design** is a q-uniform hypergraph on $[n]$, s.t. every r-set in $[n]$ is contained in exactly t edges.

An order-n **Steiner triple system** (STS) is an $(n, 3, 2, 1)$-design, i.e., a triangle-decomposition of the complete graph K_n.

A d**-dimensional permutation** is a $[n]^{d+1}$ $(0, 1)$ array with exactly one 1 in each axis-parallel line. These can be viewed as partite $(n, q, q-1, 1)$-designs. A **Latin square** is a 2-dimensional permutation, which is equivalent to a triangle decomposition of the complete tripartite graph $K_{n,n,n}$.

Peter Keevash's breakthrough works [4] [5] [6] provided existence and enumeration results for (n, q, r, t)-designs for all fixed parameters (q, r, t) (and $n \to \infty$, subject to necessary divisibility conditions - see the introduction to [4]). This opens the door to the study of *typical* designs.

The plan for today's talk is first to list three methods for studying typical designs, and then apply the third to prove a nice theorem about typical STSs. The methods are applicable to designs with arbitrary parameters, but we will always have in mind the concrete case of STSs. In particular, we will give the nitty gritty details of the third method only for STSs. Although the existence of STSs was known before Keevash (this is a result of Kirkman from 1847), and some weaker enumeration results were known previously [11], Keevash's method gives both a stronger enumeration and a way to understand the structure of typical STSs.

2. Proving a Property Holds for Typical Designs

Let P be a property of (n, q, r, t)-designs. How might we go about proving that P holds asymptotically almost surely (a.a.s.) for a uniformly random (n, q, r, t)-design? We remark that throughout the talk asymptotics will be with (q, r, t) fixed and $n \to \infty$ ranging over "permissible" numbers, i.e., those satisfying the divisibility conditions.

We've identified three methods used in the literature to analyze typical designs.

(a) Show that given any design A, P holds a.a.s. for a uniformly random element of A's (hypergraph) isomorphism class.
   - **Advantage**: Given A, a uniformly random element of A's isomorphism class can be generated by drawing $\pi \in S_n$ uniformly at random (u.a.r.) and considering $\pi \cdot A$.
   - **Disadvantage**: Most interesting properties are of isomorphism classes, not individual designs.

Date: March 19, 2018.

Example. In almost every (a.e.) d-dimensional permutation the length of the longest coordinatewise-monotone sequence of tuples is $\Theta\left(n^{d/(d+1)}\right)$ [9].

(b) **Counting**: Show that the number of designs not satisfying P is significantly less than the total number of designs.

Example. • A.e. STS is asymmetric (an order-n STS A is **asymmetric** if $|S_n \cdot A| = n!$) [1].

- A.e. Latin square is asymmetric [10]. Given Keevash's enumeration of high-dimensional permutations, one can similarly show that almost every d-dimensional permutation is asymmetric.

- **Disadvantage**: Typically one needs to obtain estimates significantly less than the number of designs. For example, the estimate Babai obtained on the number of non-asymmetric STSs was $n^{5n^2/48+o(n^2)}$, whereas the number of STSs is $n^{n^2/6\pm o(n^2)}$.

(c) **Approximation by the random greedy algorithm (RGA)**: Recall that the (n, q, r, t) RA is the iterative process wherein q-sets of $[n]$ are selected in one at a time u.a.r. subjecto to the restriction that no set is chosen that will cause an r-set to be covered $t + 1$ times. In particular, the $(n, 3, 2, 1)$ RGA is simply the Triangle removal process (TRP) [2].

Kwan [7] showed If a property holds with exceedingly high probability in the first εn^2 steps of the TRP then it holds a.a.s. for a uniformly random set of εn^2 edges chosen from a uniformly random STS. Keevash's theorem are a major ingredient of this technique. This is a little surprising, since Keevash's methods produce designs that are far from random. Nevertheless, using them in a clever way allows one to study typical designs.

Example. This method was used by Kwan [7] to show that for $n = 0 \bmod 3$). a.e. order-n STS contains a perfect matching.

Kwan's proof that a.e. STS contains a perfect matching combines method (c) with the absorbing method. Presenting this in a single hour would require glossing over too many details. Instead, we'll state a weak form of the "Approximation Theorem" that provides the engine for method (c). We'll then apply it to count Pasch configurations in typical STSs. Finally, we'll sketch a proof of the Approximation Theorem.

3. Pasch Configurations in Typical Steiner Triple Systems

Definition 3.1. A **Pasch configuration** (PC) is a triangle-decomposition of the one-skeleton of an octahedron.

We seek to answer the question, "How many copies of the Pasch configuration does a typical STS contain?" To the best of our knowledge, the following result is new (although almost all the ideas are present in Kwan's paper).

Theorem 3.2. *Almost every order-n Steiner triple system contains at least* $(1 - o(1))\frac{n^2}{24}$ *Pasch configurations.*

Remark 3.3. Theorem 3.2 provides only a lower bound, and an upper bound is conspicuously absent. In fact, Kwan's method, applied naively, cannot provide a matching upper bound due to the "infamous upper tail" issue (see [3]). We'll come back to this presently.

Remark 3.4. Kwan and Sudakov [8] showed an analogous result for Latin squares, using methods specific to Latin squares.

3.1. **An Heuristic in the Analysis of Typical Designs.** We will argue that $n^2/24$ is probably the "correct" number of PCs in a typical STS. To see why, we introduce the following heuristic to analyze typical (n, q, r, t)-designs: An (n, q, r, t)-designs is a q uniform hypergraph with $t\binom{n}{r}/\binom{q}{r}$ edges. Thus its density is approximately t/n^{q-r}. Therefore, it is not unreasonable to get a sense of its properties by studying the q-uniform binomial hypergraph

$$\mathcal{H}_q\left(n; \frac{t}{n^{q-r}}\right).$$

In the context of STSs this translates to $\mathcal{H}_3(n; 1/n)$. There are $\left(1 - O\left(\frac{1}{n}\right)\right) n^6$ *labeled* copies of the PC in the complete 3-uniform hypergraph K_n^3. The PC has automorphism group of size 24. Therefore the expected number of *unlabeled* PCs in $\mathcal{H}_3(n; 1/n)$ is $\approx n^2/24$.

3.2. **Kwan's Approximation Theorem.**

Notation. Let $N = \frac{1}{3}\binom{n}{2} \approx n^2/6$ be the number of triangles in an STS.

A **partial design** on $[n]$ is a collection of edge-disjoint triangles.

For a partial design S, we will denote by $PC(S)$ the number of PCs contained in S.

Let $\alpha \in (0, 1)$. The **Nibble** with parameters (n, α), denoted $\mathbb{N}(n, \alpha)$, is the following distribution on pairs (H, H') of hypergraphs: Let $H' \sim \mathcal{H}_3\left(n; \frac{\alpha}{n}\right)$. Let $H \subseteq H'$ be the subcollection of triangles that are edge-disjoint from all others.

The following is a weakened version of Kwan's main theorem from [7].

Theorem 3.5 (Approximation Theorem)**.** *Let $\alpha > 0$ be a sufficiently small constant. Let P be a monotone increasing property of partial designs on n vertices. Let S be a uniformly random order-n STS. Let $S_{\alpha N}$ be a uniformly random αN-subset of S. Let $(H, H') \sim \mathbb{N}(n, \alpha)$. Suppose that*

$$\mathbb{P}[H \notin P] = \exp\left(-\Omega\left(n^2\right)\right).$$

Then, a.a.s., $S_{\alpha N} \in P$.

3.3. **Proof of Theorem 3.2.** Theorem 3.2 will follow from the Approximation Theorem and the next two lemmas.

Lemma 3.6. *Let $\varepsilon, \alpha > 0$. Let S be an order-n STS containing fewer than $(1-\varepsilon)n^2/24$ PCs. Let $S_{\alpha N}$ be a uniformly random αN-subset of S. Then, a.a.s., the number of PCs in $S_{\alpha N}$ is less than $(1-\varepsilon/2)\alpha^4 n^2/24$.*

Proof. Follows from Chebychev's inequality, for example. □

Lemma 3.7. *Let $\varepsilon > 0$. There exists some $\alpha(\varepsilon) > 0$ s.t. for $(H, H') \sim \mathbb{N}(n, \alpha)$:*

$$\mathbb{P}\left[PC(H) \leq (1-\varepsilon)\alpha^4 \frac{n^2}{24}\right] = \exp\left(-\Omega\left(n^2\right)\right).$$

Remark 3.8. Perhaps surprisingly, Lemma 3.7 does not hold with the inequality sign reversed. Due to the "infamous upper tail", it is simply not true that the probability of containing more than $(1+\varepsilon)\alpha^4 \frac{n^2}{24}$ PCs is anywhere near as low as $\exp\left(-\Omega\left(n^2\right)\right)$.

Proof of Theorem 3.2. Let S be a uniformly random order-n STS. We need to show that for every $\varepsilon > 0$, $\mathbb{P}\left[PC(S) \leq (1-\varepsilon)\frac{n^2}{24}\right] = o(1)$. Well, let $\mathcal{L}$ be the collection of order-n STSs containing fewer than $(1-\varepsilon)\frac{n^2}{24}$ PCs. Let α be the constant whose existence is guaranteed by Lemma 3.7. Draw $S_{\alpha N} \in \binom{S}{\alpha N}$ u.a.r. Then, by Theorem 3.5 and Lemma 3.7:

$$\mathbb{P}\left[PC(S_{\alpha N}) \leq (1-\varepsilon)\alpha^4 \frac{n^2}{24}\right] = o(1).$$

On the other hand, by Lemma 3.6:

$$\mathbb{P}\left[PC(S_{\alpha N}) \leq (1-\varepsilon)\alpha^4 \frac{n^2}{24}\right] \geq (1-o(1))\mathbb{P}[S \in \mathcal{L}].$$

Combining the two inequalities, we have:

$$\mathbb{P}[S \in \mathcal{L}] = o(1),$$

as desired. □

Proof of Lemma 3.7. We will use the following measure concentration inequality. It is a consequence of Freedman's inequality, and is proved in [7].

Lemma 3.9. *Let $f : \{0,1\}^m \to \mathbb{R}$ be a Boolean function satisfying the Lipschitz condition $|f(x) - f(x')| \leq K$ for all x, x' that differ in exactly one coordinate. Let $x = (X_1, \dots, X_m) \sim Ber(p)$ be i.i.d. random variables. Then for all $t > 0$:*

$$\mathbb{P}[|f(x) - \mathbb{E}f(x)| > t] \leq \exp\left(-\frac{t^2}{4K^2mp + 2Kt}\right).$$

Set $\alpha = \varepsilon^2$. Let $(H, H') \sim \mathbb{N}(n, \alpha)$. Let $X = PC(H)$. We first calculate the expectation of X. This is the number of PCs in H' whose triangles are edge disjoint from all other edges in H'. Thus:

$$\mathbb{E}X = \left(1 - O\left(\frac{1}{n}\right)\right)\frac{n^6}{24}\left(\frac{\alpha}{n}\right)^4\left(1 - \frac{\alpha}{n}\right)^{4n(1-o(1))} \geq (1-5\alpha)\alpha^4\frac{n^2}{24}.$$

We might expect the next step to be an application of Lemma 3.9 to X. However, the Lipschitz constant would be on the order of n - far too large to be useful. We therefore define the random variable Z as the maximal size of a collection of *triangle disjoint* PCs in H. As adding a triangle to H' can increase Z by at most 1, and removing a triangle from H' can change Z by at most 3 (because removing a triangle from H' causes at most three other triangles to be added to H), Z is 3-Lipschitz. In order to bound $\mathbb{E}Z$ from below, we let Y be the number of pairs of PCs in H' that share a triangle. We may obtain a collection of triangle-disjoing PCs in H by taking the set of all PCs and removing one from each pair counted by Y. Therefore, $Z \geq X - Y, \mathbb{E}Z \geq \mathbb{E}X - \mathbb{E}Y$. We have:

$$\mathbb{E}Y \leq \frac{n^6}{24}\left(\frac{\alpha}{n}\right)^4\left(n^3\left(\frac{\alpha}{n}\right)^3 + n\left(\frac{\alpha}{n}\right)^2\right) \leq \alpha^7 n^2.$$

Thus:

$$\mathbb{E}Z \geq (1-5\alpha)\alpha^4\frac{n^2}{24} - \alpha^7 n^2 \geq (1-6\alpha)\alpha^4\frac{n^2}{24}.$$

Applying Lemma 3.9 with (for example) $t = \alpha^5 n^2$ proves the lemma. □

3.4. **Proof of the Approximation Theorem.** The main idea in the proof of the Approximation Theorem is that if we have a partial design S of size αN, and S satisfies a pseudorandomness condition, the number of completions to an STS doesn't depend too much on S. We also need a coupling of the Nibble with the the uniform distribution on pseudorandom partial designs of size αN, and a lemma about pseudorandom properties of random subsets of STSs.

It is our wish to avoid overloading with parameters and definitions, so we will postpone the precise definition of our pseudorandomness condition. Very roughly, a partial design is pseudorandom if in the *graph* of uncovered edges all degrees, codegress, etc., are approximately the same. We will denote this condition by Q.

We need the following lemmas to prove the Approximation Theorem.

For a partial design S, let $STS(S)$ denote the number of completions of S to an STS.

Lemma 3.10 (Completion Lemma)**.** *Let $S, S' \in Q$ be partial designs of size αN. Then:*

$$\frac{STS(S)}{STS(S')} \leq \exp\left(-o\left(n^2\right)\right).$$

This is proved by obtaining a good estimate on the number of completions of a pseudo-random partial design. The upper bound follows from the entropy method of Linial and Luria, while the lower bound follows from Keevash's techniques.

Denote by $\mathbb{R}(n, \alpha)$ the distribution on partial designs obtained by running the TRP for αN steps.

Lemma 3.11 (Coupling Lemma)**.** *Let P be a monotone increasing property of partial designs. Let $(H, H') \sim \mathbb{N}(n, \alpha)$ and let R be a uniformly random design of size αN. If $\mathbb{P}\left[H \in P\right] \leq \exp\left(-\Omega\left(n^2\right)\right)$ then $\mathbb{P}\left[R \in P | Q\right] \leq \exp\left(-\Omega\left(n^2\right)\right)$.*

Lemma 3.12 (Pseudorandomness Lemma)**.** *Let S be an order-n STS. Let $S_{\alpha N}$ be a uniformly random subset of S of size αN. A.a.s. $S_{\alpha N} \in Q$.*

Putting these together we can prove the Approximation Theorem.

Proof of Theorem 3.5. Let S be a uniformly random order-n STS and let $S_{\alpha N} \subseteq S$ be a uniformly random subset of size αN. We have:

$$\mathbb{P}\left[S_{\alpha N} \in P\right] \leq \mathbb{P}\left[S_{\alpha N} \in P | S_{\alpha N} \in Q\right] + \mathbb{P}\left[S_{\alpha N} \notin Q\right].$$

The second term tends to zero by the Pseudorandomness Lemma. As for the first term, by the Completion Lemma:

$$\mathbb{P}\left[S_{\alpha N} \in P | S_{\alpha N} \in Q\right] \leq \frac{\left|\left\{T \in P \cap Q : T \text{ is a partial design of size } \alpha N\right\}\right|}{\left|\left\{T \in Q : T \text{ is a partial design of size } \alpha N\right\}\right|} \exp\left(o\left(n^2\right)\right)$$

Finally, by the Coupling Lemma:

$$\frac{\left|\left\{T \in P \cap Q : T \text{ is a partial design of size } \alpha N\right\}\right|}{\left|\left\{T \in Q : T \text{ is a partial design of size } \alpha N\right\}\right|} \leq \exp\left(-\Omega\left(n^2\right)\right).$$

The claim follows. □

References

[1] László Babai, *Almost all steiner triple systems are asymmetric*, Annals of Discrete Mathematics, vol. 7, Elsevier, 1980, pp. 37–39.

[2] Tom Bohman, Alan Frieze, and Eyal Lubetzky, *Random triangle removal*, Advances in Mathematics **280** (2015), 379–438.

[3] Svante Janson and Andrzej Ruciński, *The infamous upper tail*, Random Structures & Algorithms **20** (2002), no. 3, 317–342.

[4] Peter Keevash, *The existence of designs*, arXiv preprint arXiv:1401.3665 (2014).

[5] ______, *Counting designs*, arXiv preprint arXiv:1504.02909 (2015).

[6] ______, *The existence of designs ii*, arXiv preprint arXiv:1802.05900 (2018).

[7] Matthew Kwan, *Almost all steiner triple systems have perfect matchings*, arXiv preprint arXiv:1611.02246 (2016).

[8] Matthew Kwan and Benny Sudakov, *Intercalates and discrepancy in random latin squares*, arXiv preprint arXiv:1607.04981 (2016).

[9] Nathan Linial and Michael Simkin, *Monotone subsequences in high-dimensional permutations*, Combinatorics, Probability and Computing **27** (2018), no. 1, 69–83.

[10] Brendan D McKay and Ian M Wanless, *On the number of latin squares*, Annals of combinatorics **9** (2005), no. 3, 335–344.

[11] Richard M Wilson, *Nonisomorphic steiner triple systems*, Mathematische Zeitschrift **135** (1974), no. 4, 303–313.

Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

E-mail address: menahem.simkin@mail.huji.ac.il