Thresholds, Expectation Thresholds and Cloning

Tomasz Przybyłowski Oliver Riordan

Submitted: Feb 22, 2024; Accepted: Oct 24, 2024; Published: Dec 27, 2024

Abstract

Let p_{c} and q_{c} be the threshold and the expectation threshold, respectively, of an increasing family $\mathcal{F}$ of subsets of a finite set X, and let l be the size of a largest minimal element of $\mathcal{F}$. Recently, Park and Pham proved the Kahn–Kalai conjecture, which says that $p_{\mathrm{c}} \leqslant K q_{\mathrm{c}} \log_2 l$ for some universal constant K. Here, we slightly strengthen their result by showing that $p_c \leqslant 1 - \mathrm{e}^{-K q_{\mathrm{c}} \log_2 l}$. The idea is to apply the Park–Pham Theorem to an appropriate 'cloned' family $\mathcal{F}_k$, reducing the general case (of this and related results) to the case where the individual element probability p is small.

Mathematics Subject Classifications: 60C05

1 Introduction

Given a finite set X and $p \in [0,1]$, let X_p be the random subset where each element is included with probability p, independently of the others; we call this a *p-random subset of X* for short. A family $\mathcal{F} \subseteq \mathcal{P}(X)$ is *increasing* (or an *up-set*) if $A \in \mathcal{F}$ and $A \subseteq B \subseteq X$ implies $B \in \mathcal{F}$. Throughout, $\mathcal{F}$ will denote a non-trivial increasing family, so $\mathcal{F} \neq \varnothing$, $\mathcal{P}(X)$.

Let $\mathcal{G} \subseteq \mathcal{P}(X)$. Following (to some extent) the terminology in [5], for $q \in [0,1]$, we define the *q-cost* of $\mathcal{G}$ to be

$$\mathrm{cost}_q(\mathcal{G}) := \sum_{S \in \mathcal{G}} q^{|S|} = \mathbb{E}[|\{S \in \mathcal{G} \colon S \subseteq X_q\}|].$$

We say that $\mathcal{G}$ is *q-cheap* if $\mathrm{cost}_q(\mathcal{G}) \leqslant \frac{1}{2}$. We say that $\mathcal{G}$ *covers* $\mathcal{F}$ if

$$\mathcal{F} \subseteq \langle \mathcal{G} \rangle := \bigcup_{S \in \mathcal{G}} \{T \colon T \supseteq S\},$$

i.e., if $\mathcal{F}$ is contained in the up-set generated by $\mathcal{G}$, and that $\mathcal{G}$ is a *q-cheapest* cover of $\mathcal{F}$ if it has minimal q-cost among all covers of $\mathcal{F}$. Of course, such a $\mathcal{G}$ may not be unique.

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
(`przybylowski@maths.ox.ac.uk`, `riordan@maths.ox.ac.uk`).

The *threshold* of $\mathcal{F}$ is the unique $p_{\mathrm{c}} \in (0,1)$ such that $\mathbb{P}(X_{p_{\mathrm{c}}} \in \mathcal{F}) = \frac{1}{2}$. Following Kahn and Kalai [3], the *expectation threshold* of $\mathcal{F}$ is defined to be

$$q_{\mathrm{c}}(\mathcal{F}) := \sup\{q \in [0,1] \colon \text{there exists a } q\text{-cheap cover of } \mathcal{F}\}.$$

Kahn and Kalai noted that q_{c} easily (by applying the union bound) gives a lower bound on p_{c}. Strikingly, they conjectured, and Park and Pham [5] proved, that a not-too-large multiple of q_{c} provides an upper bound on p_{c}. Let $l(\mathcal{F})$ be the size of a largest minimal element of $\mathcal{F}$.

Theorem 1 ([5])**.** *There is a universal constant K such that for every finite set X and non-trivial increasing family $\mathcal{F} \subseteq \mathcal{P}(X)$ with $l(\mathcal{F}) \geqslant 2$,*

$$q_{\mathrm{c}}(\mathcal{F}) \leqslant p_{\mathrm{c}}(\mathcal{F}) \leqslant K q_{\mathrm{c}}(\mathcal{F}) \log_2 l(\mathcal{F}). \tag{1}$$

In [1], a slightly simpler version of the proof from [5] is given, which yields $K \leqslant 16$ (or $K \leqslant 11$ if $l(\mathcal{F}) \geqslant 4$). The arguments in [5, 1] easily give the stronger bound

$$p_{\mathrm{c}}(\mathcal{F}) \leqslant 1 - (1 - K q_{\mathrm{c}}(\mathcal{F}))^{\log_2 l(\mathcal{F})}, \tag{2}$$

whenever $q_{\mathrm{c}}(\mathcal{F}) \leqslant 1/K$. Our aim is to strengthen this bound even further.

Theorem 2. *Let K be the universal constant from Theorem 1. Then for every finite set X and non-trivial increasing family $\mathcal{F} \subseteq \mathcal{P}(X)$ with $l(\mathcal{F}) \geqslant 2$,*

$$p_{\mathrm{c}}(\mathcal{F}) \leqslant 1 - \mathrm{e}^{-K q_{\mathrm{c}}(\mathcal{F}) \log_2 l(\mathcal{F})}. \tag{3}$$

Since $\mathrm{e}^{-xy} \geqslant (1-x)^y \geqslant 1 - xy$, the bound (3) does indeed imply (2) and (1). The strategy of the proof is simply to apply Theorem 1 to an appropriate transformed family $\mathcal{F}_k$. The main ingredient is a 'cloning' lemma showing that the expectation threshold changes as expected under this transformation.

The improvement in (3) comparing with (2) is most significant when $K q_{\mathrm{c}}$ is relatively large. In particular, (2) gives no information when $K q_{\mathrm{c}} \geqslant 1$. However, as pointed out by Keith Frankston[1], in this case neither does (3), because of the trivial bound $p_{\mathrm{c}} \leqslant 2^{-1/l}$ obtained by considering a single minimal element of $\mathcal{F}$. This trivial bound is strictly stronger than (3) already when $l \geqslant 3$ and $K q_{\mathrm{c}} \geqslant 1$. Thus it is far from clear that Theorem 2 will have any applications. Nevertheless, we hope that the method of proof is interesting, namely using a 'cloning' lemma to transfer results from the small p regime to general p.

2 The k-cloned model

Let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a non-trivial increasing family, and let k be a positive integer. Informally, we define the k-cloned version of $\mathcal{F}$ as follows: instead of performing a single

[1]Personal communication.

experiment for each $x \in X$ to determine whether $x \in X_p$, we perform k experiments, declaring $x \in X_p$ if and only if at least one of $(x, 1), \ldots, (x, k)$ is present in the corresponding random subset $Y_{p'}$ of $Y = X \times [k]$. Then $\mathcal{F}_k$ will be the set of subsets of Y corresponding to $\mathcal{F}$.

Formally, we define the *k-cloned* family $\mathcal{F}_k \subseteq \mathcal{P}(X \times [k])$ to be

$$\mathcal{F}_k := \{S \subseteq X \times [k] : \pi(S) \in \mathcal{F}\},$$

where $\pi \colon X \times [k] \to X$ is the projection onto the first coordinate, and $\pi(S)$ denotes the image of a set S under π. Since $\mathcal{F}$ is increasing and non-trivial, so is $\mathcal{F}_k$.

If Y_p is a p-random subset of $X \times [k]$, then $\pi(Y_p)$ is a $(1 - (1 - p)^k)$-random subset of X. It is thus immediate that

$$p_{\mathrm{c}}(\mathcal{F}) = 1 - (1 - p_{\mathrm{c}}(\mathcal{F}_k))^k. \tag{4}$$

If every cheapest cover $\mathcal{H}$ of $\mathcal{F}_k$ arose in the natural way (see below) from a cheapest cover of $\mathcal{F}$, there would be a simple relationship between the expectation thresholds of $\mathcal{F}$ and $\mathcal{F}_k$. Although the former does not hold in general, the latter does.

Lemma 3. *Let k be a positive integer and let $\mathcal{F}$ be a non-trivial increasing family of subsets of a finite set X. Then*

$$q_{\mathrm{c}}(\mathcal{F}_k) = \frac{1}{k} q_{\mathrm{c}}(\mathcal{F}).$$

Proof. For the lower bound, let $\mathcal{G}$ be a q-cheap cover of $\mathcal{F}$. It suffices to construct a (q/k)-cheap cover of $\mathcal{F}_k$.

Given $S \subseteq X$ let

$$\Psi(S) = \{S' \subseteq X \times [k] \colon \pi(S') = S \text{ and } |S'| = |S|\} \tag{5}$$

be the set of *minimal* pre-images of S under π. Note that $|\Psi(S)| = k^{|S|}$, since we must choose one $(x, i) \in X \times [k]$ for each $x \in S$. Let

$$\mathcal{H} := \bigcup_{S \in \mathcal{G}} \Psi(S). \tag{6}$$

It is immediate that $\mathcal{H} \subseteq \mathcal{P}(X \times [k])$ is a cover of $\mathcal{F}_k$. Indeed, if $A \in \mathcal{F}_k$ then $\pi(A) \in \mathcal{F}$ and hence there is $B \in \mathcal{G}$ such that $\pi(A) \supseteq B$. Then $A \in \langle \Psi(B) \rangle \subseteq \langle \mathcal{H} \rangle$. Furthermore,

$$\mathrm{cost}_{q/k}(\mathcal{H}) = \sum_{T \in \mathcal{H}} \left(\frac{q}{k}\right)^{|T|} = \sum_{S \in \mathcal{G}} \sum_{T \in \Psi(S)} \left(\frac{q}{k}\right)^{|T|} = \sum_{S \in \mathcal{G}} k^{|S|} \left(\frac{q}{k}\right)^{|S|} = \mathrm{cost}_q(\mathcal{G}).$$

Since $\mathcal{G}$ is q-cheap, $\mathcal{H}$ is (q/k)-cheap. Hence $q_{\mathrm{c}}(\mathcal{F}_k) \geqslant \frac{1}{k} q_{\mathrm{c}}(\mathcal{F})$.

Turning to the upper bound, let $\mathcal{H}$ be an arbitrary q-cheap cover of $\mathcal{F}_k$. It suffices to deduce the existence of a kq-cheap cover of $\mathcal{F}$. For this we use an averaging argument, over all $k^{|X|}$ copies of X living inside $X \times [k]$.

Formally, for $X' \in \Psi(X)$ let $\mathcal{H}_{X'} = \mathcal{P}(X') \cap \mathcal{H}$. Observe that $\mathcal{H}_{X'}$ is a cover of $\mathcal{P}(X') \cap \mathcal{F}_k$. The projection π induces an inclusion- and size-preserving bijection from $\mathcal{P}(X')$ to $\mathcal{P}(X)$. Hence, $\pi(\mathcal{H}_{X'})$ is a cover of $\mathcal{F}$ and $\mathrm{cost}_r(\mathcal{H}_{X'}) = \mathrm{cost}_r(\pi(\mathcal{H}_{X'}))$ for any $r \in [0,1]$.

A set $S \subseteq X \times [k]$ is contained in *some* $X' \in \Psi(X)$ if and only if S contains no 'duplicates' (two elements from the same fibre $\{x\} \times [k]$). In that case, S is contained in exactly $k^{|X|-|S|}$ sets $X' \in \Psi(X)$, since in constructing X' we have a free choice from each $\{x\} \times [k]$ where $x \in X \setminus \pi(S)$. Thus

$$\begin{aligned}\sum_{X' \in \Psi(X)} \mathrm{cost}_{kq}(\mathcal{H}_{X'}) &= \sum_{X' \in \Psi(X)} \sum_{S \in \mathcal{H}_{X'}} (kq)^{|S|} = \sum_{S \in \mathcal{H}} (kq)^{|S|} \sum_{X' \in \Psi(X)} \mathbb{1}(S \subseteq X') \\ &\leqslant \sum_{S \in \mathcal{H}} (kq)^{|S|} k^{|X|-|S|} = k^{|X|} \sum_{S \in \mathcal{H}} q^{|S|} = k^{|X|} \mathrm{cost}_q(\mathcal{H}).\end{aligned}$$

Therefore there is some $X' \in \Psi(X)$ such that $\mathrm{cost}_{kq}(\mathcal{H}_{X'}) \leqslant \mathrm{cost}_q(\mathcal{H})$. Then $\pi(\mathcal{H}_{X'})$ is a kq-cheap cover of $\mathcal{F}$, as desired. □

We are now ready to prove Theorem 2.

Proof of Theorem 2. Consider the k-clone $\mathcal{F}_k$ of $\mathcal{F}$ for $k > \frac{1}{K q_c(\mathcal{F}) \log_2 l(\mathcal{F})}$. Applying the Park–Pham result, Theorem 1, to $\mathcal{F}_k$ and then using Lemma 3 we have

$$p_c(\mathcal{F}_k) \leqslant K q_c(\mathcal{F}_k) \log_2 l(\mathcal{F}_k) = K \frac{q_c(\mathcal{F})}{k} \log_2 l(\mathcal{F}),$$

since $l(\mathcal{F}_k) = l(\mathcal{F})$. Applying (4) we deduce that

$$p_c(\mathcal{F}) \leqslant 1 - \left(1 - K \frac{q_c(\mathcal{F})}{k} \log_2 l(\mathcal{F})\right)^k.$$

Taking the $k \to \infty$ limit yields the result. □

3 Discussion

The proof of Theorem 2 is based on 'scaling' the bound (1) up from the sparse case. Since we can take our cloning factor k to be arbitrarily large, only the asymptotic behaviour of (1) near zero matters. In particular, we obtain exactly the same final result if we start from (2) instead. More generally, to deduce (3) we only need the bound $p_c(\mathcal{F}) \leqslant (K + o(1)) q_c(\mathcal{F}) \log_2 l(\mathcal{F})$ as $q_c(\mathcal{F}) \to 0$ with $l(\mathcal{F})$ fixed.

At first sight, it might seem 'obvious' that $q_c(\mathcal{F}_k) = q_c(\mathcal{F})/k$, since one might expect every q-cheapest cover of $\mathcal{F}_k$ arises from a q-cheapest cover of $\mathcal{F}$ via the cloning procedure in (5) and (6). However, this is not true in the general case. For $\mathcal{F} = \{\{1,2,3\},\{1,2\},\{1,3\}\}$ we have $q_c(\mathcal{F}_2) = 1/4$ and $\mathcal{F}_2$ admits a 1/4-cheapest cover not arising in this way.

Related to this, one might expect that if $\mathcal{F}$ is symmetric under the action of some group G on the ground-set X, then for every $q \in [0,1]$ it has a q-cheapest cover which has the same symmetry. This again is not true in general. Let $\mathcal{F}$ consist of all subsets of a 5-element set having size at least 4. Then one can check that all q_{c}-cheapest covers take the form $\{\{i,j,k,l\},\{i,j,m\},\{k,l,m\}\}$, which does not have the full symmetry of $\mathcal{F}$.

Finally, we note that obvious modifications to the proof of Lemma 3 show that the *fractional expectation threshold* q_{f} defined in [2, 6], which behaves better with respect to symmetry, scales under cloning in the same way that q_{c} does: we have $q_{\mathrm{f}}(\mathcal{F}_k) = q_{\mathrm{f}}(\mathcal{F})/k$.

Since the first version of the current paper was written, B. Park and J. Vondrák [4] have proved a non-uniform version of the Park–Pham theorem. They suggest that the 'cloning' procedure described above gives an alternative proof of their result, but with worse constants. However, *starting* from their Theorem 1, one can apply the cloning procedure (adapted in the natural way to the non-uniform setting) to deduce the (very slightly) stronger bound $p_x = 1 - \mathrm{e}^{-q_{\mathrm{c}}(4\lfloor \log_2(2l) \rfloor + 7)}$. In other words, our reduction can also be combined with their result.

Acknowledgements

First author was supported by the Additional Funding Programme for Mathematical Sciences, delivered by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research.

References

[1] T. Bell, *The Park–Pham Theorem with Optimal Convergence Rate*, Electron. J. Comb. 30, 2023, #P2.25.

[2] K. Frankston, J. Kahn, B. Narayanan, and J. Park, *Thresholds versus fractional expectation-thresholds*, Ann. of Math. (2) 194 (2021), no. 2, 475–495.

[3] J. Kahn and G. Kalai, *Thresholds and Expectation Thresholds*, Combin., Probab. and Comput. (16) 495–502, 2007.

[4] B. Park and J. Vondrák, *A simple proof of the non-uniform Kahn–Kalai conjecture*, SIAM J. Discrete Math. (38), no. 3, 2089–2094, 2024.

[5] J. Park and H. T. Pham, *A proof of the Kahn–Kalai conjecture*, J. Amer. Math. Soc., 37, 235–243, 2024.

[6] M. Talagrand, *Are many small sets explicitly small?*, STOC'10 – Proceedings of the 2010 ACM International Symposium on Theory of Computing, 13–35, ACM, New York, 2010.