Randomized Algorithms 2023A – Lecture 13
Concentration Bounds and Edge-Sparsification of Hypergraphs*

Robert Krauthgamer

1 Concentration Bounds

Chernoff-Hoeffding bound: Let $X = \sum_{i\in[n]} X_i$ where $X_i \in [0,1]$ for $i \in [n]$ are independently distributed random variables. Then

$$\begin{aligned}
\forall t > 0, &\qquad \Pr[|X - \mathbb{E}[X]| \ge t] \le 2e^{-2t^2/n}. \\
\forall 0 < \varepsilon \le 1, &\qquad \Pr[X \le (1-\varepsilon)\,\mathbb{E}[X]] \le e^{-\varepsilon^2\,\mathbb{E}[X]/2}. \\
\forall 0 < \varepsilon \le 1, &\qquad \Pr[X \ge (1+\varepsilon)\,\mathbb{E}[X]] \le e^{-\varepsilon^2\,\mathbb{E}[X]/3}. \\
\forall t \ge 2e\,\mathbb{E}[X], &\qquad \Pr[X \ge t] \le 2^{-t}.
\end{aligned}$$

Exer: Let X be binomial $B(n, 1/3)$. What is the probability that X deviates from its expectation additively by $r > 1$ standard deviations? Think of r being $10, \log n, \sqrt{n}$, and compare the different bounds.

Exer: Let $a_1, \ldots, a_n$ be an array of numbers in the range $[0,1]$. Design a randomized algorithm that estimates their average within $\pm\varepsilon$ (i.e., additive error ε) by reading only $O(1/\varepsilon^2)$ elements. The algorithm should succeed with probability at least 90%.

Exer: Let $S_1, \ldots, S_n$ be subsets of $[n]$. Design an algorithm for 2-coloring the elements $[n]$, such that in every set S_i the balance, defined as $|\#\text{black} - \#\text{white}|$, is at most $O(\sqrt{n \log n})$.

2 Edge-Sparsification of Hypergraphs (via Importance Sampling)

Cuts in Hypergraphs: Let $H = (V, E, w)$ be a hypergraph with edge weights $w : E \mapsto \mathbb{R}_+$. Every (nontrivial) $S \subset V$ defines a cut

$$\delta_H(S) := \{e \in E : \text{both } e \cap S \ne \emptyset, e \cap \bar{S} \ne \emptyset\}.$$

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. The exercises are for self-practice and need not be handed in. In the interest of brevity, most references and credits were omitted.

For every subset of edges $E' \subset E$ define $w(E') = \sum_{e\in E'} w(e)$, which in particular defines the weight of a cut S as $w(\delta_H(S))$.

Cut sparsifier: Let $H = (V, E, w)$ be a hypergraph, and let $\varepsilon \in (0, 1)$. A hypergraph $H' = (V, E', w')$ (on same vertex set) is a $(1 + \varepsilon)$-cut-sparsifier if

$$\forall S \subset V, \qquad w'(\delta'_H(S)) \in (1 \pm \varepsilon) w(\delta_H(S)).$$

Theorem 1 [Kogan and Krauthgamer, 2015]: For every n-vertex hypergraph $H = (V, E, w)$ and every $\varepsilon \in (0, 1/2)$, there is a $(1 + \varepsilon)$-cut-sparsifier H' with $m = O(n^2/\varepsilon^2)$ hyperedges.

The original proof was an extension of an earlier result, by [Benczur and Karger, 1996], which introduced this sparsification concept for graphs (all hyperedges are of size 2) and proved a (better) bound of $O(\varepsilon^{-2} n \log n)$ edges and also gave an algorithm with near-linear running time. We will see a different proof based on Importance Sampling. The sparsification bound for hypergraphs was recently improved to $O(\varepsilon^{-2} n \log n)$ by [Chen, Khanna, and Nagda, 2020].

Idea: We will construct H' by sampling m edges, where each edge is drawn according to probabilities $\{p(e)\}_{e\in E}$, and a sampled edge e is given new weight $w'(e) = \frac{w(e)}{mp(e)}$. It follows immediately that the expected weight of a (every) cut S in H' equals the weight of the same cut S in H. Viewing this as importance sampling, it will be easy to reduce the variance. But this holds only for any one cut S, and a sparsifier H' requires a guarantee for *all cuts*, and thus we will prove a concentration bound and then apply a union bound.

Construction of sparsifier: For each edge $e \in E$ define its *sensitivity*

$$s(e) := \max_{S\subset V:\ e\in\delta_H(S)} \frac{w(e)}{w(\delta_H(S))},$$

and define the total sensitivity to be $s(E) = \sum_{e\in E} s(e)$.

Construct H' at random by picking m edges from H, each chosen independently according to the distribution on edges given by $p(e) = \frac{s(e)}{S(E)}$, and every edge $e \in E$ that is chosen is given a new weight $w'(e) = \frac{w(e)}{m\ p(e)}$.

Remark: This construction may create parallel edges, because the same edge may be picked multiple times (up to m). We can always merge parallel edges at the end, which does not change the weight of any cut, but it will be easier for us to analyze the hypergraph before such merges.

Expectation:

$$\forall S \subset V, \qquad \mathbb{E}[w'(\delta_{H'}(S))] = m \sum_{e\in\delta_H(S)} p(e) \frac{w(e)}{m\ p(e)} = \sum_{e\in\delta_H(S)} w(e) = w(\delta_H(S)).$$

Lemma 2: $s(E) \le n^2$.

Proof: Was seen in class by "charging" the sensitivity of each hyperedge to the minimum cut between a pair of vertices.

Lemma 2': $s(E) \le n - 1$.

Exer: Prove this bound by repeatedly removing from G a minimum cut whose removal increases the number of connected components by 1 (this is a global minimum cut in one of the components).

Hint: It then suffices to show that in each of the $n-1$ iterations, the sensitivity of the removed edges sums up to at most 1.

Lemma 3 (Importance Sampling): Let $S \subset V$ and $\lambda = s(E)$. Then

$$\forall e \in \delta_H(S), \qquad p(e) \geq \frac{1}{\lambda} \frac{w(e)}{w(\delta_H(S))}.$$

Proof: Given our S and e,

$$p(e) = \frac{s(e)}{s(E)} \geq \frac{1}{s(E)} \frac{w(e)}{w(\delta_H(S))}.$$

QED

Corollary 4: If $m \geq c \cdot s(E)$ for a suitable $c = c(\varepsilon) > 0$, then

$$\forall S \subset V, \qquad \Pr[w'(\delta_{H'}(S)) \in (1 \pm \varepsilon) w(\delta_H(S))] \geq 3/4.$$

Proof: Using the importance sampling theorem we saw in previous classes, and the bound in Lemma 3, $\mathrm{Var}(w'(\delta_{H'}(S))) \leq m \cdot \frac{1}{m^2} \cdot \lambda \cdot (w(\delta_H(S)))^2 = (\frac{1}{2}\varepsilon w(\delta_H(S)))^2$. The corollary now follows by Chebyshev's inequality.

QED

Proof of Theorem 1: Was seen in class, by proving a concentration bound for each cut, and then applying a union bound over all 2^n cuts.

Exer: Show that with high probability the total weight of all edges of H' is approximately equal to that in H, i.e., $w'(E') = \Theta(w(E))$.

Exer: Analyze a variant of this algorithm, where the sampling is different: Independently for each edge $e \in E$, with probability $q(e) = \min\{1, O(\varepsilon^{-2} n) \cdot s(e)\}$ add this edge to H' with new weight $w'(e) = \frac{w(e)}{q(e)}$, and otherwise do not add to H'. Note that now the number of edges is random (and has to be analyzed).

3 Coresets for Clustering (cont'd)

We finished the proof of a theorem from an earlier class (stated next for completeness).

Theorem 7: Let Y be a multiset of $m \geq L' d \varepsilon^{-2} \log \frac{1}{\varepsilon}$ points from X, each sampled iid according to distribution $q(.)$ and reweighted by $w(x) = \frac{1}{m q(x)}$, for a suitable constant $L' > 0$. Then with high probability, Y is a strong coreset for the 1-median of X.

Proof of Theorem 7 (sketch): Discretize the set of possible centers by considering a cover of the ball $B^* = B(c^*, \frac{1}{\varepsilon} \frac{\mathrm{OPT}}{n})$ by balls of radius $\varepsilon \frac{\mathrm{OPT}}{n}$, and letting N be the set of all their center points. Applying Lemma 5 to each $c' \in N$ with $\delta = \frac{1}{4|N|}$ and taking a union bound over all $c' \in N$, we get that with probability at least $1 - |N|\delta \geq 3/4$,

$$\forall c' \in N, \qquad f(Y, c') \in (1 \pm \varepsilon) f(X, c'). \tag{1}$$

Assuming this event holds, we derive a slightly weaker bound for all possible centers, namely,

$$\forall c \in \mathbb{R}^d \qquad f(Y, c) \in (1 \pm 8\varepsilon) f(X, c).$$

The proof was seen in class, by distinguishing two cases, $c \in B^*$ and $c \notin B^*$. In the former, we "replace" c with a nearby $c' \in N$, and in the latter we "appproximate" both $f(X, c)$ and $f(Y, c)$ by $n\|c - c^*\|$).