COMS W6998-5: Algorithms through Geometric Lens (Fall'17) Nov 27, 2017

Lecture 22 – Discrepancy

Instructors: *Alex Andoni, Ilya Razenshteyn* Scribes: *Malhar Thakkar*

1 Introduction

Today and in the next class, we'll talk about discrepancy which is the last topic for this course. There will be no geometry in today's lecture, but, it will be an introduction to Wednesday's class.

Setup We have some $\mathcal{U}$ of size n and a bunch of subsets $\mathcal{S} = \{S_1, S_2, \ldots, S_m\} \subseteq \mathcal{U}$.

Goal Color $\mathcal{U}$ so that each set $S \in \mathcal{S}$ is as equally colored as possible.
More formally, we would like to find a coloring $\chi : \mathcal{U} \to \{-1, 1\}$ to find the value of the following objective function.

$$\min \max_{S_i \in \mathcal{S}} |\sum_{s \in S_i} \chi(s)|$$

Let the entire objective function be denoted by $disc_{\mathcal{S}}(\chi)$ and $\chi(S_i) = \sum_{s \in S_i} \chi(s)$. If $\chi(S_i) = 0$ for some i, it means that the set S_i is balanced (in other words, equal number of 1's and -1's) and if it is non-zero, it means that there is some imbalance in that set and we're trying to minimize this imbalance.

Today, we'll primarily focus on the question that for any family of sets $\mathcal{S}$, does there exist a coloring χ such that some particular bounds hold?

So, what's the motivation behind studying discrepancy? In addition to giving nice theoretical results, it is also pragmatic in the sense that various state of the art approximation algorithms (for instance, an approximation algorithm for the bin packing problem) are based on discrepancy.

Today, we'll focus on a very specific case where $|\mathcal{S}| = n$ as it happens to be the most convenient think to look at. Hence, we just have a single parameter n. Now, in this case, we want to find a good way of coloring the universe $\mathcal{U}$ so that the sets are as balanced as possible.

We have the following trivial claim.

Claim 1. *$\exists\, \chi$ such that $disc_{\mathcal{S}}(\chi) \leq n$*

This is a very trivial claim in the sense that it holds for any coloring χ as the sizes of the sets in $\mathcal{S}$ is at most n and hence, $\chi(S_i)$ for any i is at most n which trivially proves the claim.

Now, the question is, how much can we improve it?

One way to improve it would be to try a random coloring of $\mathcal{U}$. This results in the following claim.

Claim 2. *$\exists\, \chi$ such that $disc_{\mathcal{S}}(\chi) \leq O(\sqrt{n \log n})$*

One important thing to note is that the above coloring χ is random and that the claim holds with high probability for this random coloring χ.

Proof. Choose χ to be random.
Then, $\forall S_i\ Pr[|\chi(S_i)| \geq t\sqrt{n}] \leq e^{-\Omega(t^2)}$.
This holds because $\chi(S_I)$ is a sum of random independent $\pm 1's$ and there are at most n items in S_i and so, we can apply Chernoff bound to this sum.

For $Pr[\forall S_i, |\chi(S_i)| \geq t\sqrt{n}] \leq 0.1$, we have to set $e^{-\Omega(t^2)} \leq \frac{1}{10\,n}$ and hence, set a proper value of t. We set $t = O(\sqrt{log\, n})$.

By union bound, with probability at least 0.9, $\forall S_i,\ |\chi(S_i)| \leq O(\sqrt{n \log n})$ □

Not only does this give an existential proof, it also gives an easy algorithm. Now, the question is, is $O(\sqrt{n \log n})$ the best we can do?

Towards the end of the lecture, we'll see a lower bound that states $\exists \mathcal{S}$ such that $\forall \chi$, $disc_{\mathcal{S}}(\chi) \geq \Omega(\sqrt{n})$. So, now the question is, what can we improve? The lower bound? Or the upper bound? This will be the main result of today's lecture that actually, the upper bound can be improved. The proof for this improved upper bound is given in the following theorem.

Theorem 3 (Spencer, 1985)**.** *$\forall \mathcal{S}\ \exists \chi$ such that $disc_{\mathcal{S}}(\chi) \leq O(\sqrt{n})$*

The proof of this theorem is very nice. As discrepancy is very important practically, it is important to understand this proof. An interesting aspect of this proof is that it is non-constructive. For random coloring too, the proof was non-constructive but at least we had an efficient polynomial time algorithm that gave us a coloring which satisfied the required bounds. In this case, we can show that a coloring exists such that the $disc_{\mathcal{S}}(\chi) \leq O(\sqrt{n})$ but it is not clear how to obtain this coloring χ. In fact, till 2010, no one knew how to obtain this coloring. Moreover, [**?**] gave the first randomized polynomial time algorithm for discrepancy minimization.

But, today, we'll see a non-constructive proof. On Wednesday, we'll see a constructive proof of this theorem.

The proof of this theorem uses entropy of a random variable. We'll not need to know much about it apart from a couple of properties, but we'll see its definition.

2 Entropy

Definition 4. *Entropy H of a discrete random variable X is defined as*

$$H(X) = \sum_i Pr[X = i] \log_2 \frac{1}{Pr[X = i]}$$

Now, a natural question arises, what values can entropy take?

If X is uniformly distributed over $\{1, 2, \ldots, n\}$. Then, the entropy of X is equal to $\log_2 n$. And, in fact, you can show that, for a random variable which can take n values, this is the maximum possible value of entropy. Another extreme is when X is constant. This is a special case when $n = 1$. In this case, the entropy is 0.

Entropy gives information about the random variable due to the following theorem proved by Shannon(maybe).

Setup Let $x_1, x_2, \ldots,$ be independently drawn samples from some known distribution X. Then, what is the number of bits that are required to encode theses samples such that we can uniquely decode all these samples uniquely and the expected length of the encoding is as small as possible?

The answer is we can do it in $(\alpha + H(X)) \times (\#\,\text{samples})$ bits where α is some constant and $H(X)$ is the entropy of the random variable X. This can be proved by Huffman encoding. We'll not need this property for our proof but we'll require a couple of other simple properties namely the following.

Claim 5. *Entropy is sub-additive.*

$$H(X, Y) \leq H(X) + H(Y)$$

Moreover, the equality holds only when X and Y are independent random variables. Roughly speaking, for the joint distribution, the probabilities in the entropy formula will be product of probabilities for X and Y and taking a logarithm of this quantity will result in addition of these terms. This sub-additivity is the key to the theorem that we'll be proving.

Another claim that we will need is also very easy.

Claim 6. *If $\forall i$, $Pr[X = i] \leq \delta$ then*

$$H(X) \geq \log_2 \frac{1}{\delta}$$

For instance, if X is uniformly distributed over a large set, it is easy to see that the above claim holds for some δ. We know that the entropy can be at most $\log_2 n$, so this claim gives us the lower bound on the entropy. The proof is left as an exercise for the reader.

That's all we need to know about entropy to prove the main theorem.

2.1 Asymptotic Formula for the Volume on a Hypercube

The proof of the asymptotic formula for the volume of hypercubes uses the concept of entropy. Let us formalize the problem.
Suppose that we a hypercube $\{-1, 1\}^d$. Let's define a ball $B(x, \alpha\, d)$ such that

$$B(x, \alpha\, d) = \{y| \,\#(i's \,\text{such that}\, y_i \neq x_i) \leq \alpha\, d\}$$

. Basically, we have a Hamming ball of radius $\alpha\, d$. The question is, how many points are in this ball? We'd be interested in the regime where $0 < \alpha < 1/2$.

Actually, we can prove that and if you're interested in combinatorics or information theory, it is highly recommended that you try to work out the proof once by yourself.
The answer is

$$|B(x, \alpha\, d)| \approx 2^{H(\alpha)\, d}$$

where

$$H(\alpha) = \alpha\, \log_2\, \frac{1}{\alpha} + (1 - \alpha)\, \log_2\, \frac{1}{1 - \alpha}$$

Basically, for any fixed α, the size of the ball will be an exponential function, but this exponent depends on α. Hence, $H(\alpha)$ is equal to the entropy of a biased coin with the probability of heads being α.

Why does the formula hold?
Rough idea of the proof:
The number of points on a d-dimensional hypercube which are at a distance k from some point x on the hypercube is exactly equal to $\binom{d}{k}$. This implies

$$|B(x, \alpha\, d)| = \sum_{k \leq \alpha\, d} \binom{d}{k} \approx \binom{d}{\alpha\, d} = \frac{d!}{(\alpha\, d)!\ ((1 - \alpha)\, d)!}$$

where we use Stirling's formula to approximate the factorial of a number n as

$$n! \approx \sqrt{2\, \pi\, n}\, \left(\frac{n}{e}\right)^n$$

Under the regime, $0 < \alpha < 1/2$, the last term of the summation of $\binom{d}{k}$ dominates and hence, we can approximate it using just the last term and so, we analyze what the value of $\binom{d}{\alpha\, d}$ will be when $d \to \infty$.

Now, we have enough tools to prove our theorem.

3 Proof of Spencer's Theorem

We'll start proving by using a random coloring. Random coloring is bad. But, we'll show that it can "guide" us in some sense.
Instead of proving the theorem directly, we'll see proof of a key lemma and then we'll see how this lemma implies the theorem.

Let us formulate the lemma.

Lemma 7. *Partial coloring.*

$$\exists \chi : \mathcal{U} \to \{-1, 0, 1\}$$

such that

$$disc_{\mathcal{S}}(\chi) \leq O(\sqrt{n})$$

Basically, we'll get the desired bound by partial coloring by coloring not the entire universe $\mathcal{U}$ but almost the entire universe $\mathcal{U}$. But, of course, a trivial coloring that satisfies the bound is coloring the entire universe with the color 0. Hence, it is clear to see that we'll have to restrict the number of 0 colored items to in order to prove the main theorem.

So, we'll require that χ is 0 on at most $\frac{3n}{4}$ elements. So, we color at least a non-trivial fraction of the whole universe with $\pm 1's$.

So, why does partial coloring imply the theorem?
Because, we can apply partial coloring recursively. Basically, we apply the partial coloring lemma and get some partial coloring χ. It gives us good discrepancy and it leaves uncolored significantly fewer elements. Then we restrict everything to this subset of the universe and repeat.

So, we can't literally repeat this lemma because it assumes that the number of sets and sizes of the sets are exactly the same. But, there is a natural way to relax this lemma for all the subsequent steps. We are not going to see the proof in class, but you can look it up. The proof is exactly the same with an additional letter that corresponds to this imbalance. But, we will just prove this partial coloring.

It's actually interesting that the partial coloring approach works, i.e., if you give up some elements, you can then make it work and it's interesting to compare it with random coloring. If you think how the random coloring works, we sample coloring at random and then, for a very small fraction of the sets, the imbalance is much larger than $O(\sqrt{n})$. Hence, we can handle almost all sets. But, handling almost all the elements turns out to be much more efficient than handling all the sets somehow.

We'll use entropy to prove the partial coloring lemma. To compute entropy, we need some random variable. Let's consider the coloring as our random variable.

Proof.

Definition 8. *Let χ be a random coloring.*

We apply this random coloring and hence, for each set, we get some imbalance. The imbalance is a number between $-n$ and n.
Let us compute these imbalances and let us round them a little bit.

Definition 9. *Let b_i be the nearest integer to $\frac{\chi(S_i)}{100\sqrt{n}}$.*

If $b_i = 0$,

$$-50\sqrt{n} \leq \chi(S_i) \leq 50\sqrt{n}$$

Hence, if $b_i = 0$, we're doing pretty well in the sense that we almost have a perfectly balanced set.
As χ is a random variable, $b_i's$ are also random variables.

Claim 10. *$H(b_i) \leq \epsilon$ where ϵ is explicit but a very small constant.*

Proof. The proof of this claim is quite simple, actually. It is because, $b_i's$ are almost always 0. With Chernoff bound, we can show that with very high probability, we'll be in the desired range. Basically, all the probability mass is almost concentrated on one point.
We can use the Chernoff bound to state that

$$Pr[b_i \geq t] \leq e^{-\frac{100^2\, t^2}{2}}$$

. Hence,

$$Pr[b_i = 0] \geq 1 - 2\, e^{-\frac{100^2\, t^2}{2}}$$

which is almost equal to 1.

$$Pr[b_i = 1] \approx e^{-\frac{100^2\, t^2}{2}}$$

$$Pr[b_i = -1] \approx e^{-\frac{100^2\, t^2}{2}}$$

$$Pr[b_i = 2] \approx e^{-\frac{100^2 \times 4\, t^2}{2}}$$

$$Pr[b_i = -2] \approx e^{-\frac{100^2 \times 4\, t^2}{2}}$$

Hence, entropy of the random variable b is given by

$$H(b) = \sum_j Pr[b_i = j]\, \log_2 \frac{1}{Pr[b_i = j]}$$

The first term in the above formula will be very small as $Pr[b_i = j] \approx 1$. In fact, the biggest term will be the second term. So,

$$H(b_i) \approx e^{-\frac{100^2\, t^2}{2}} \times \frac{100^2}{2}$$

Hence, $H(b_i)$ is very small.

The basic intuition is that when we apply a random coloring, almost all the sets will be balanced hence, the entropy of b is very small.

□

Now, let's look at all the $b_i's$ together. We consider the joint entropy as follows.

$$H(b_1, b_2, \ldots, b_n)$$

It's not easy to compute the above quantity as $b_i's$ are not independent because if two sets S_i and S_j are exactly the same, then b_i and b_j will also be the same and hence, are not independent of each other. But, all we care about is that

$$H(b_1, b_2, \ldots, b_n) \leq \epsilon\, n$$

just by sub-additivity and the claim that we just proved.

Earlier, we showed the claim that if all the probabilities are small, the entropy must be large. Here, we have the opposite. The entropy is small, and so, there is some large enough probability. Hence, we can conclude that,

$$\exists b_i^*, b_2^*, \ldots, b_n^* \in \mathbb{Z}$$

such that

$$Pr[b_i = b_i^*\, \forall i] \geq 2^{-\epsilon\, n}$$

because if there are no such points then entropy must be larger than $\epsilon\, n$.
In our algorithm, we first choose some random coloring and then map the coloring χ to $b_i's$ ($\chi \rightarrow (b_1, b_2, \ldots, b_n)$).
As there are 2^n possible colorings,

$$\#\chi's \text{ such that } b_i = b_i^* \forall i \geq 2^{(1-\epsilon)\, n}$$

Now, we'll use the bounds for volume of a ball on a hypercube.
$\chi's$ live on $\{-1, 1\}^n$ and we have a large set of $\chi's$ which satisfy the above property.

Claim 11. *$\mathcal{U} \subseteq \{-1, 1\}^d$ such that $|S| \geq 2^{(1-\epsilon)\, n}$*
$\exists u_1, u_2 \in \mathcal{U}$ such that
 # i's where $(u_1)_i \neq (u_2)_i$ is at least $\left(\frac{1}{2} - \delta\right) n$ where $\delta = \delta(\epsilon)$

Proof. Just take a set $\mathcal{U}$ and find any point in it and call it u_1. Now, consider a ball around u_1 of the radius equal to $\left(\frac{1}{2} - \delta\right) n$. Now, we'll use the volume bound to find an upper bound of this ball centred around u_1.

$$\text{Volume of the ball} \leq 2^{H\left(\frac{1}{2} - \delta\right) n}$$

We can choose δ such that the volume becomes upper bounded by $2^{(1-\epsilon)\, n}$ □

□

Summary
Let χ_1 and χ_2 be two colorings such that

1. χ_1 and χ_2 are different in $\geq 0.49\, n$ places.
2. $\forall S_i\ |\chi_1(S_i) - \chi_2(S_i)| \leq O(\sqrt{n})$

Now, we can define our partial coloring as

$$\chi = \frac{\chi_1 - \chi_2}{2}$$

You can check that χ is indeed a partial coloring. χ will be 0 in at most $0.51\,n$ places.
Hence, it can be seen that the discrepancy of this coloring is at most $O(\sqrt{n})$. Let's now see the results of the lemma for the general case where the number of sets may not be exactly equal to n.

Lemma 12. *$\mathcal{U} = n$ and $S_1, S_2, \ldots, S_m \subseteq U$ such that $m \geq n$*

$$\exists \chi : \mathcal{U} \to \{-1, 0, 1\} \text{ such that}$$

1. *# of 0's is $\geq \frac{3n}{4}$*
2. *$|\chi(S)| \leq O(\sqrt{n \log \frac{m}{n}})$*

Proof of this lemma is exactly the same as before with some additional notation introduced.

The following theorem gives the lower bound on the discrepancy.

Theorem 13 (Spencer, 1985)**.** *$S_1, S_2, \ldots S_n \subseteq \mathcal{U}$ and $|\mathcal{U}| = n$ and the $S_i's$ are formed randomly such that*

$$\text{whp } (\forall \chi \, \exists S_i \text{ such that } |\chi(S_i)| \geq \Omega(\sqrt{n}))$$

Proof. Let's take a fixed coloring. For simplicity, let us assume that this coloring has equal number of 1's and -1's. This is not the case in general but the bounds are even better when this is not the case.

χ is fixed and #i's such that $\chi(i) = 1$ is $n/2$.

Let's consider the probability of the complement of the event mentioned in the theorem.

$$Pr[\forall S_i \ |\chi(S_i)| \leq c\sqrt{n}] = Pr[|\chi(S_i)| \leq c\sqrt{n}]^n \leq \left(\frac{1}{10}\right)^n$$

due to central limit theorem. as all S_i's are independent. □