E0 206: Theorist's Toolkit October 20, 2020

Lecture 6-7: Probabilistic Methods

Instructor: Arindam Khan *Scribe: Arka Ray, Prasanna Srikar Regati*

These two lectures are the last in series of lectures covering probabilistic methods. These start with the Chernoff bounds and discuss one of its applications in combinatorial discrepancy. Further in the lectures, martingales are defined, using which Azuma-Hoeffding inequalities are derived. These can be applied even if underlying random variables are not independent. Such a martingale sequence can be constructed from an arbitrary sequence of random variables and hence more generically applicable. Lastly the Lovasz Local Lemma improves upon the union bound for events with limited dependency.

1 Chernoff Bounds

Recall that in the probability refresher the Chernoff bounds were briefly touched upon. There are many variants of the Chernoff bounds. Chernoff bounds are used to analyse tails of random variables which are themselves sum of independent random variables taking two values.

1.1 Two Bounds

Here we look at two variants, one with 0-1 random variables other with -1/1 random variables and will go on to use the -1/1 variant to prove a bound in combinatorial discrepancy.

Theorem 1. *Let $X_1, X_2, ..., X_n$ be n mutually independent random variables taking the values 0 and 1. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}\, X$. If $\Pr[X_i = 1] = p_i$ then $\forall \delta \in (0, 1)$,*

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

Theorem 2. *Let $X_1, X_2, ..., X_n$ be n mutually independent random variables and $X = \sum_{i=1}^{n} X_i$.*
If $\Pr[X_i = 1] = \Pr[X_i = -1] = \frac{1}{2}$ then,

$$\Pr[X > a] \leq e^{-a^2/2n}$$

1.2 Discrepancy

Discrepancy is in essence the study of gaps in approximating the continuous by the discrete. This has many applications in numerous areas including computational geometry, approximation algorithms, complexity theory, machine learning, Monte Carlo methods, etc. Consider approximation algorithms where many a times problems are solved using linear programs. After obtaining a solution to an appropriate linear program the values are rounded to give a solution to the original problem. One can notice the connection between discrepancy the gaps in rounding. For bin packing an improvement to $\mathrm{OPT} + O(\log \mathrm{OPT})$ was obtained using this method.

To understand the concept we look at an example,

Example 1. *Given n arbitrary points in an unit square, color them with red or blue such that each rectangular region is colored as evenly as possible. In this scenario we measure the discrepancy as,*

$$\textit{Discrepancy} = \max_{\textit{rectangles } r \subset [0,1]\times[0,1]} ||r \cap R| - |r \cap B||$$

where the R, B are the set of red, blue points respectively.

Take the instantiation of this problem described by the figure 1. The coloring shown has a discrepancy 3. If continuous values were allowed then each point could have been colored half blue and half red. In the discrete case random coloring gives $O(\sqrt{n \ln n})$ discrepancy.

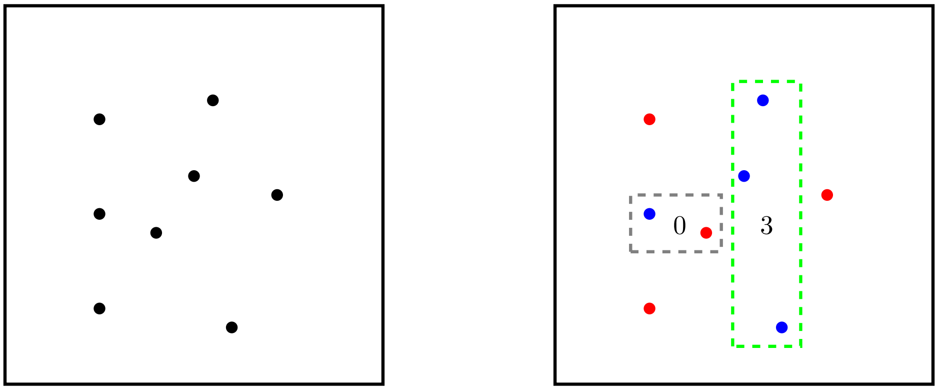


Figure 1: An example of geometric discrepancy

Now we turn to the notion of combinatorial discrepancy in contrast to geometric discrepancy which was presented above. We generalize the above problem can be generalized to the following by considering arbitrary family of subsets instead of rectangles.

Definition 1. *Given a family of subset $\mathcal{A} \subset 2^{\Omega}$ of a set Ω. Consider a coloring $\chi : \Omega \to \{-1, 1\}$ then for any $A \subset \Omega$, $\chi(A) = \sum_{a \in A} \chi(a)$. In this case the discrepancy of $\mathcal{A}$ with respect to χ is,*

$$disc(\mathcal{A}, \chi) = \max_{A \in \mathcal{A}} |\chi(A)|$$

and the discrepancy of $\mathcal{A}$ is,

$$disc(\mathcal{A}) = \min_{\chi} disc(\mathcal{A}, \chi)$$

An alternate way of looking at combinatorial discrepancy is by using matrices. In this formulation we restrict ourselves to $\Omega = [n]$ but this is a moot point since it is possible to map any finite set to $[n]$ and this allows a simpler presentation. Except for $\Omega = [n]$ the two definitions are equivalent.

For a family of subsets $\mathcal{A} = \{S_1, S_2, ..., S_m\}$ where $S_i \subset \Omega$, we call a $m \times n$ matrix $B = [b_{ij}]$ the *incidence matrix* for $\mathcal{A}$ if

$$b_{ij} = \begin{cases} 1 & \text{if } j \in S_i \\ 0 & \text{otherwise} \end{cases}$$

Now if we take $u = (\chi(1), \chi(2), ...)$ to represent the coloring then, $Bu^T = (\chi(S_1), \chi(S_2), ...)^T$. With this notation in place we define discrepancy as follows.

Definition 2. *Given a family of subsets $\mathcal{A} \subset 2^{\Omega}$ of a set $\Omega = [n]$. If B is the incidence matrix for $\mathcal{A}$ and $u \in \{-1, 1\}^n$ represent a coloring then the discrepancy of $\mathcal{A}$ with respect to u is,*

$$disc(\mathcal{A}, u) = \|Bu^T\|_\infty$$

and the discrepancy of $\mathcal{A}$ is,

$$disc(\mathcal{A}) = \min_{u \in \{-1,1\}^n} disc(\mathcal{A}, u) = \min_{u \in \{-1,1\}^n} \|Bu^T\|_\infty$$

where $\|\cdot\|_\infty$ is the L_∞ norm.

A final remark before showing a bound on the value of discrepancy is if we take v_j to be the j-th column of B then, $\mathrm{disc}(\mathcal{A}) = \min\|\pm v_1 \pm v_2 ... \pm v_n\|_\infty$ where we minimize over choice of signs in the sum.

Now we show a bound on discrepancy with the help of Chernoff bound's second variant.

Theorem 3. *Let $\mathcal{A} \subset 2^\Omega$ be a family of subsets of a set Ω. If $|\mathcal{A}| = n$ and $|\Omega| = m$ then, $disc(\mathcal{A}) \leq \sqrt{2m\ln(2n)}$*

The main idea in proving this is to take a random coloring and consider the events where we violate the bound for a given set and show the expected number of such sets is less than one, thereby showing a set which satisfies the bound exists.

Proof. Take random coloring $\chi : \Omega \to \{-1, 1\}$. For $A \subset \Omega$, let X_A be the indicator random variable for the event $\chi(A) > \alpha$ where $\alpha = \sqrt{2m\ln(2n)}$. If $|A| = a$ then using second variant of Chernoff bound (and union bound),

$$\mathbb{E}[X_A] = \Pr[\chi(A) > \alpha] < 2e^{-\alpha^2/2a} < 2e^{-\alpha^2/2m} = 2e^{-\frac{2m\ln(2n)}{2m}} = 2e^{-\ln(2n)} = \frac{1}{n}$$

Therefore,

$$\mathbb{E}\left[\sum_{A\in\mathcal{A}} X_A\right] = \sum_{A\in\mathcal{A}} \mathbb{E}[X_A] < |\mathcal{A}| \cdot \frac{1}{n} = 1$$

Now, it can be concluded that for some χ, $\sum_{A\in\mathcal{A}} X_A = 0$. Hence, $\mathrm{disc}(\mathcal{A}, \chi) \leq \alpha$, i.e, $\mathrm{disc}(\mathcal{A}) \leq \alpha$. □

A corollary of the theorem is if $m = n$ then $\mathrm{disc}(\mathcal{A}) = O(\sqrt{n\ln n})$. This result was improved by Spencer in 1985 by showing six standard deviations suffice (note: the mean is 0 and the standard deviation is $\sqrt{n}$) using partial coloring. There are a lot of open questions in this area.

One might ask (at least should ask) that the result is existential, the naive constructive version is not efficient (requiring exponential time), but can we get a constructive proof (which is efficient). In fact, Bansal in 2010 gave an algorithm using semi definite programming and Lovett and Meka in 2012 gave a simpler (randomized algorithm) using a restricted version of random walks.

2 Martingales

The Chernoff/Hoeffding bounds that we saw can be applied only if the random variables are independent. However, such independence does not hold in most of the practical scenarios. So we need much generic bounds that work even for dependent random variables. In this section we study a setting in which such bounds can be obtained, namely martingales.

Martingales are sequences of random variables satisfying certain conditions that arise in numerous applications, such as random walks and gambling problems.Here we give a brief introduction to martingales and then discuss several concepts related to them: Doob Martingales, the martingale stopping theorem and the Azuma–Hoeffding inequality.

Definition 4 (Martingales)**.** *A sequence of random variables $Z_0, Z_1, ...$ is a martingale with respect to the sequence $X_0, X_1, ...$ if, for all $n \geq 0$, the following conditions hold:*

(i) Z_n *is a function of* $X_0, X_1, ..., X_n$ *;*

(ii) $\mathbb{E}[|Z_n|] < \infty$*;*

(iii) $\mathbb{E}[Z_{n+1}|X_0, ..., X_n] = Z_n$ *.*

Also, a sequence of random variables $X_0, X_1, ...$ is called a martingale if ,for all $n \geq 0$ it holds that:

(i) $\mathbb{E}[|X_n|] < \infty$*;*

(ii) $\mathbb{E}[X_{n+1}|X_0, ..., X_n] = X_n$ *.*

2.1 Example of a Martingale Sequence

Consider a gambler who plays a sequence of fair games(i.e., the expected money that he can win/lose in each game is zero).The gambler plays the game multiple times; neither his stakes, nor the outcome of the games need be independent, but each play is fair. Let X_i be the amount the gambler wins on the i^{th} game (X_i is negative if the gambler loses), and let Z_i be the gambler's total winnings at the end of the i^{th} game(Z_i can also be negative if the gambler loses more games than he wins). Because each game is fair, $\mathbb{E}[X_i] = 0$ and

$$\begin{aligned}\mathbb{E}[Z_{i+1}|X_1, X_2, ..., X_i] &= Z_i + \mathbb{E}[X_{i+1}]\\ &= Z_i\end{aligned}$$

Thus, $Z_1, Z_2, ..., Z_n$ is a martingale with respect to the sequence $X_1, X_2, ..., X_n$. The interesting part is that the sequence is a martingale regardless of the amount bet on each game, even if these amounts are dependent upon previous results.

2.2 Construction of Martingales

But how often does such Martingales sequences occur? In many scenarios it might not be clear if there is a Martingale sequence associated with it. But we can create a martingale sequence from essentially any random variables. Such explicitly constructed martingales are referred to as Doob martingales and it can be done using the following general approach:

Let $X_0, X_1, ...$ is a sequence of random variables and Y be a random variable such that $\mathbb{E}[|Y|] \leq \infty$ and is a function of $X_0, X_1, ...$ Then $Z_i = \mathbb{E}[Y|X_0, X_1, ..., X_i], i = 0, 1, 2.., n$, gives a martingale with respect to $X_0, X_1, ..., X_n$, since

$$\begin{aligned}\mathbb{E}[Z_{i+1}|X_0, ..., X_i] &= \mathbb{E}[\mathbb{E}[Y|X_0, ..., X_{i+1}]|X_0, ..., X_i]\\ &= \mathbb{E}[Y|X_0, ..., X_i]\\ &= Z_i\end{aligned}$$

The second equality follows from the tower property of conditional expectations: if F $\subseteq$ G then
$\mathbb{E}[\mathbb{E}[X|G]|F] = \mathbb{E}[X|F]$.

Frequently in applications we will have $A = f(X_1, ..., X_n)$, i.e., A is determined by the random variables X_i. In this case, $Z_0 = \mathbb{E}[A]$ and $Z_n = \mathbb{E}[A|X_1, ..., X_n] = A$. We can think of the martingale as the estimates obtained from progressively more information about the random variable A. We begin with no information about A, and the value of the martingale is just the expectation $\mathbb{E}[A]$. At the end of the sequence we have specified all of the X_i so we have complete information about A and the martingale has the deterministic value $A(X_1, ..., X_n)$.

2.2.1 Examples on construction of Doob Martingales

We now consider two examples of Doob martingales that arise in evaluating the properties of random graphs.
Example-1 : Edge Exposure Martingale:
Let G be a random undirected simple graph $G_{n,p}$. There could possibly be $m = \binom{n}{2}$ possible edges and each such edge exists with probability p independent of other edges. Let,

$$X_j = \begin{cases} 1 & \text{if there is an edge in } j^{th} \text{ slot} \\ 0 & \text{otherwise} \end{cases}$$

Consider any finite-valued function F defined over graphs; for example, let $F(G)$ be the size of the largest independent set in G. The value of F depends on all the edges present i.e $F(G) = f(X_1, X_2, ..., X_m)$ Now let $Z_0 = E[F(G)]$ and $Z_i = E[F(G)|X_1, ..., X_i], i = 1, ..., m$. The sequence $Z_0, Z_1, ..., Z_m$ is a Doob martingale that represents the conditional expectations of $F(G)$ as we reveal whether each edge is in the graph, one edge at a time. This process of revealing edges gives a martingale that is commonly called the edge exposure martingale.

Example 2-Vertex Exposure Martingale:
Similarly, instead of revealing edges one at a time, we could also reveal the set of edges connected to a given vertex, one vertex at a time. Fix an arbitrary numbering of the vertices 1 through n, and let G_i be the subgraph of G induced by the first i vertices. Then, setting $Z_0 = \mathbb{E}[F(G)]$ and $Z_i = \mathbb{E}[F(G)|G_1, ..., G_i], i = 1, ..., n$, gives a Doob martingale that is commonly called the vertex exposure martingale.

2.3 Stopping Times

Consider again the gambler's problem that we discussed in previous section. Just to make sure not to lose all his money , the gambler might want to strategize on when he should stop playing. One such naive strategy could be to quit after exactly k(fixed number) games. Then what would be the gambler's expected winnings?

Lemma 5. *If the sequence $Z_0, Z_1, ..., Z_n$ is a martingale with respect to $X_0, X_1, ..., X_n$, then $\mathbb{E}[Z_n] = \mathbb{E}[Z_0]$*

Proof. Since $Z_0, Z_1, ...$ is a martingale with respect to $X_0, X_1, ..., X_n$, it follows that $Z_i = \mathbb{E}[Z_{i+1}|X_0, ..., X_i]$. Taking the expectation of both sides and using the definition of conditional expectation, we have

$$\mathbb{E}[Z_{i+1}] = \mathbb{E}[\mathbb{E}[Z_{i+1}|X_0, ..., X_i]] = \mathbb{E}[Z_i]. \tag{1}$$

Repeating this argument yields $\mathbb{E}[Z_n] = \mathbb{E}[Z_0]$. □

Thus, if the number of games played is initially fixed then the expected gain from the sequence of games is zero.So a better strategy would be not play fixed number of games instead decide to continue playing or to stop based on the outcomes of the games already played. For example, the gambler could decide to keep playing until his winnings total at least a hundred dollars or until his losses are less than hundred dollars. Hence the following notion of stopping times is quite powerful.

Definition 6. *A non-negative integer-valued random variable T is a stopping time for the sequence $\{Z_i, i \geq 0\}$ if the probability of the event $T = n$ is independent of the variables $\{Z_{n+j}|Z_1, ..., Z_n, j \geq 1\}$ (i.e. the variables $Z_{n+1}, Z_{n+2}, ...$ conditioned on the values of $Z_1, ..., Z_n$).*

Thus, a stopping time corresponds to a strategy for determining when to stop a sequence based only on the outcomes seen so far.
Examples:

1. The first time the gambler wins five games in a row is a stopping time. This can be determined by looking at outcomes of the games played.

2. The first time the gambler has won at least a hundred dollars is also a stopping time.

3. The last time the gambler wins five games in a row is not a stopping times since it cant be determined by looking at already played games.

But consider the case where the gambler's stopping time is the first T such that $Z_T > B$, where B is a fixed constant greater than 0. In this case, the expected gain when the gambler quits playing is greater than 0. The subtle problem with this stopping time is that it might not be finite, so the gambler may never finish playing. The martingale stopping theorem shows that, under certain conditions and in particular when the stopping time is bounded or has bounded expectation, the expected value of the martingale at the stopping time is equal to E$[Z_0]$.

Theorem 7. *If $Z_0, Z_1, ...$ is a martingale with respect to $X_1, X_2, ...$ and if T is a stopping time for $X_1, X_2, ...$, then*

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$$

whenever one of the following holds:

(i) The Z_i are bounded, so there is a constant c such that, for all i, $|Z_i| \leq c$;

(ii) T is bounded;

(iii) $\mathbb{E}[T] < \infty$, and there is a constant c such that $\mathbb{E}[|Z_{i+1} - Z_i||X_1, ..., X_i] < c$.

2.3.1 Applying Stopping Theorem on Gambler's ruin Problem

Coming back to the gambling problem we discussed already, say a player wins a dollar with probability 1/2 or loses a dollar with probability 1/2 in each game. Let X_i be the amount won on the i^{th} game, and let Z_i be the total amount won by the player after i games (again, X_i and Z_i can negative if the player loses). Assume that the player quits the game when she either loses l_1 dollars or wins l_2 dollars.We would like to answer the question:

*W*hat is the probability that the player wins l_2 dollars before losing l_1 dollars?

The players stops playing at time T (which is the stopping time), then T is the first instance of time that the player has won l_2 or lost l_1. The sequence $\{Z_0, Z_1, ...\}$ is a martingale with respect to $\{X_1, X_2, ...\}$ and Z_0=0.As the extreme values that Z_i can take are l_2 and $-l_1$, they are clearly bounded. So we can apply the martingale stopping theorem. We therefore have $\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0$. Let q be the probability that the gambler quits playing after winning l_2 dollars. Then

$$\mathbb{E}[Z_T] = l_2 * q - l_1 * (1 - q) = 0$$
$$q = \frac{l_1}{l_1 + l_2}$$

2.4 Azuma-Hoeffding Inequalities

Perhaps the most useful property of martingales for the analysis of algorithms is that Chernoff-like tail inequalities can apply, even when the underlying random variables are not independent. The main results in this area are Azuma's inequality and Hoeffding's inequality. They are quite similar, so they are often together referred to as the Azuma–Hoeffding inequality.

Theorem 8. *Let $X_0, ..., X_n$ be a martingale such that $|X_k - X_{k-1}| \leq c_k$, then for all $t \geq 1$ and any $\lambda \geq 0$*

$$Pr(|X_t - X_0| \geq \lambda) \leq 2e^{-\lambda^2/(2\sum_{k=1}^{t} {c_k}^2)} \tag{2}$$

Below is a more general form of the Azuma–Hoeffding inequality that yields slightly tighter bounds in our applications.

Theorem 9. *Let $X_0, ..., X_n$ be a martingale such that*

$$B_k \leq X_k - X_{k-1} \leq B_k + d_k \tag{3}$$

for some constants d_k and for some random variables B_k that may be functions of $X_0, X_1, ..., X_{k-1}$. Then, for all $t \geq 0$ and any $\lambda \geq 0$,

$$Pr(|X_t - X_0| \geq \lambda) \leq 2e^{-\lambda^2/(2\sum_{k=1}^{t} {d_k}^2)} \tag{4}$$

2.4.1 Application:Chromatic Number

Given a random graph G in $G_{n,p}$, the chromatic number χ(G) is the minimum number of colors needed in order to color all vertices of the graph so that no adjacent vertices have the same color. We use the vertex exposure martingale defined already to obtain a concentration result for χ(G).

Let G_i be the random subgraph of G induced by the set of vertices $1, ..., i$, let $Z_0 = \mathbb{E}[\chi(G)]$, and let $Z_i = \mathbb{E}[\chi(G)|G_1, ..., G_i]$.Since a vertex uses no more than one new color, we have that the gap between Z_i

and Z_{i-1} is at most 1, so we can apply the general framework of the Azuma–Hoeffding inequality from Theorem 13.6. We conclude that

$$Pr(|\chi(G) - \mathbb{E}[\chi(G)]| \geq \lambda\sqrt{n}) \leq 2e^{-2\lambda^2} \tag{5}$$

This result holds even without knowing the value of $\mathbb{E}[\chi(G)]$.

2.5 McDiarmid's Inequality

Another variation of Azuma-Hoeffding inequality can be derived which operates on functions of random variables. To formalize that we need to understand the Lipschitz condition.

Definition 10. *A function $f(\overline{X}) = f(X_1, X_2, ..., X_n)$ satisfies Lipschitz condition with bound c, if for any i and for any set of values $x_1, x_2, ..., x_n$ and y_i,*

$$|f(x_1, ..., x_i, ..., x_n) - f(x_1, ..., y_i, ..., x_n)| \leq c \tag{6}$$

That is, changing the value of any single coordinate can change the function value by at most c.

Theorem 11 (McDiarmid's Inequality)**.** *Let f be a function on n variables that satisfies the above Lipschitz condition with bound c. Let $X_1, ..., X_n$ be independent random variables such that $f(X_1, ..., X_n)$ is in the domain of f . Then*

$$Pr(|f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)]| \geq \lambda) \leq 2e^{(-2\lambda^2/nc^2)} \tag{7}$$

2.5.1 Application : Balls and Bins

Suppose there are m balls and n bins and we are throwing each ball into a bin at random. Let X_i is the random variable that tells us the bin in which i^{th} ball falls i.e. X_i ranges from 1 to n. Let the F be the number of empty bins after all the m balls are thrown. We can see that F is a function of the random variables $X_1, X_2, ..., X_m$ i.e $F = f(X_1, X_2, ..., X_n)$.

To check for the Lipschitz condition, we have to find a bound c such that by changing one of $X_1, X_2, ..., X_n$, F does not change by a value more than c. Say, X_i changes, i.e the i^{th} falls in a different bin. There are four cases here: (let X_i value changes from k to l)

1. The k^{th} bin has only i^{th} ball and l^{th} bin has no balls. After the change, the value of f remains same.
2. The k^{th} bin has only i^{th} ball and l^{th} bin has more than zero balls. After the change, the value of f increases by 1.
3. The k^{th} bin has balls other than i^{th} ball and l^{th} bin has no ball balls. After the change, the value of f decrease by 1.
4. The k^{th} bin has balls other than i^{th} ball and l^{th} bin has more than zero balls. After the change, the value of f remains same.

Therefore the function f satisfies the Lispchitz condition with bound $c = 1$. Hence we obtain:

$$Pr(|F - \mathbb{E}[F]| \geq \varepsilon) \geq 2e^{-2\varepsilon^2/n} \tag{8}$$

3 The Lovász Local Lemma

The is the final topic of probabilistic methods. Suppose we want to show that no bad event happens with some non-zero probability. If they are mutually independent it is simple. Say $E_1, ..., E_n$ are the bad events then $\overline{E}_1, ..., \overline{E}_n$ are also mutually independent. Hence the probability that no bad events happen is given by the expression,

$$\Pr\left[\bigcap_{i\in[n]} \overline{E}_i\right] = \prod_{i\in[n]} \Pr\left[\overline{E}_i\right]$$

which has a non-zero probability given $\Pr[E_i] < 1$ for $i \in [n]$.

In some situations mutual independence might be too much to ask for. One option is to use union bound but it will give us loose bounds. The Lovász local lemma(LLL) can be used to get better bounds for situations with limited dependency.

3.1 The Result

Before stating the result we first need the following definitions.

Definition 3. *An event E_{n+1} is said to be independent of the events $E_1, E_2, ..., E_n$ if for every subset $I \subset [n]$,*

$$\Pr\left[E_{n+1} \middle| \bigcap_{i\in I} E_i\right] = \Pr[E_{n+1}]$$

Definition 4 (Dependency di-graph)**.** *For a set of events $E_1, E_2, ..., E_n$ a graph $G = (V, E)$ with $V = [n]$ such that for all $i \in [n]$, E_i is independent of $\{E_j | (i, j) \notin E\}$.*

The dependency (di)graph is defined in terms of a negative which might make it a bit opaque. So as an example consider the following graph: In this case A_1 is independent of A_4, A_5, A_7. In general dependency graph is directed but we'll work mostly with undirected version. Also note that there can be multiple choices for dependency graph for a particular set of events on a probability space.

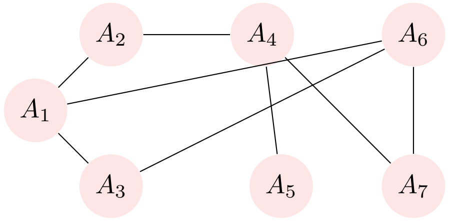


Figure 2: Example of a dependency graph

The general setup for Lovász local lemma is,

- A collection of independent random variables $x_1, x_2, ..., x_n$.
- Each event A_i only depends on $\{x_j | j \in S_i\}$ for $S_i \subset [n]$.

A valid dependency graph can be formed by placing $i \sim j$ whenever $S_i \cap S_j \neq \phi$.

For example, we'll analyze satisfiability formulas later in the lecture. Consider the formula $(x_1 \vee \overline{x}_2 \vee x_3) \wedge (\overline{x}_1 \vee x_2 \vee x_3)$, it has two clauses with 3 literals each. When we analyze the problem we'll define $E_i =$ event that i-th clause of the formula. Also let X_k be the assignment of x_k. Clearly, E_i depends on the variables in the i-th clause. Therefore, we add an edge $E_i \sim E_j$ whenever they share a variable.

Let us take a final example of dependency graph construction before going into the result. Say, $x_1, x_2, x_3 \in 0, 1$ are independent random variables distributed uniformly (they take the value 1 with probability 1/2 and the value 0 with probability 1/2). Notice they constitute a set of independent and identically distributed (i.i.d) random variables. Now,consider the following events:

1. $A_1 : x_1 +_2 x_2 = 0$
2. $A_2 : x_2 +_2 x_3 = 0$
3. $A_3 : x_3 +_2 x_1 = 0$

where $+_2$ denotes the addition modulo 2. We see that,

$$\Pr[A_1] = \Pr[A_2] = \Pr[A_3] = \frac{1}{2}$$

$$\Pr[A_1 \cap A_2] = \frac{1}{4} = \Pr[A_1]\Pr[A_2]$$

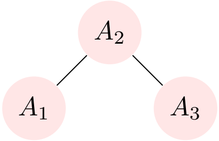
$$\Pr[A_2 \cap A_3] = \frac{1}{4} = \Pr[A_2]\Pr[A_3]$$

$$\Pr[A_3 \cap A_1] = \frac{1}{4} = \Pr[A_3]\Pr[A_1]$$

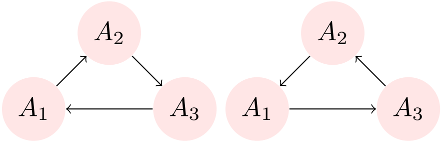
But,

$$\Pr[A_1 \cap A_2 \cap A_3] = \frac{1}{4} \neq \Pr[A_1]\Pr[A_2]\Pr[A_3]$$

Therefore these events are pairwise independent but not mutually independent. Therefore, the empty graph is not a valid dependency graph. Following are some of the possible dependency graph.



On the left side we have a valid dependency graph. This is correct as the only edge missing is between A_1, A_3 and as we showed above A_1 is independent of A_3 and vice versa. Notice, we can not remove any of the edges present. Say we were to remove $A_1 \sim A_2$, then it would mean A_1 is independent of A_3, A_2 which we showed is not correct.



On the right we give two more examples of valid dependency (di)graph. Note that $A_1 \not\sim A_2$ in the first figure and $A_1 \not\sim A_3$ in the second figure are both consistent.

Now, we state the Lovász local lemma in the general form.

Theorem 12 (Lovász Local Lemma)**.** *Let $E_1, E_2, ..., E_n$ be a set of events in an arbitrary probability space, and $G = (V, E)$ be their dependency di-graph. Suppose $\exists x_1, x_2, ..., x_n \in [0, 1)$ such that $\forall i \in [n]$*

$$\Pr[E_i] \leq x_i \prod_{i,j \in E} (1 - x_j)$$

then,

$$\Pr\left[\bigcap_{i \in [n]} \overline{E}_i\right] \geq \prod_{i \in [n]} (1 - x_i)$$

In particular with non-zero probability none of the events happen.

Proof. Let $S \subset [n]$ and $|S| = s < n$. Using induction on s we show that $\forall k \notin S$,

$$\Pr\left[E_k \middle| \bigcap_{j\in S} \overline{E}_j\right] \leq \prod_{j\in S} x_j$$

This intuitively means that E_k is a low probability event even if other events don't occur.

For base case we consider $s = 0$ for which by from the precondition of the theorem,

$$\Pr[E_k] \leq x_k \prod_{(k,j)\in E} (1 - x_j) \leq x_k$$

For the inductive step assume the statement holds for $s' < s$. Partition S into

$$S_1 = \{j \in S | (k, j) \in E\}$$

and

$$S_2 = S - S_1$$

Before proceeding recall that,

$$\Pr\left[\bigcap_{i\in[n]} A_i\right] = \prod_{i\in[n]} \Pr\left[A_i \middle| \bigcap_{j\in[i-1]} A_j\right]$$

and

$$\Pr[A|B \cap C] = \frac{\Pr[A \cap B|C]}{\Pr[B|C]}$$

Using the second identity,

$$\Pr\left[E_k \middle| \bigcap_{j\in S} \overline{E}_j\right] = \frac{\Pr\left[E_k \cap \bigcap_{j\in S_1} \overline{E}_j \middle| \bigcap_{j\in S_2} \overline{E}_j\right]}{\Pr\left[\bigcap_{j\in S_1} \overline{E}_j \middle| \bigcap_{j\in S_2} \overline{E}_j\right]}$$

The above quantity can be bounded from above by obtaining a upper bound and lower bound for the numerator and denominator respectively. The upper bound for numerator is,

$$\begin{aligned}\Pr\left[E_k \cap \bigcap_{j\in S_1} \overline{E}_j \middle| \bigcap_{j\in S_2} \overline{E}_j\right] &\leq \Pr\left[E_k \middle| \bigcap_{j\in S_2} \overline{E}_j\right] && (\because \Pr[A \cap B] \leq \Pr[A]) \\ &= \Pr[E_k] \leq x_k \prod_{(k,j)\in E} (1 - x_j)\end{aligned}$$

and let $S_1 = \{j_1, j_2, ..., j_r\}$ then lower bound for the denominator is,

$$\begin{aligned}\Pr\left[\bigcap_{j\in S_1} \overline{E}_j \middle| \bigcap_{j\in S_2} \overline{E}_j\right] &= \prod_{i\in[r]} \Pr\left[\overline{E}_{j_i} \middle| \bigcap_{m\in[i-1]} \overline{E}_{j_m} \cap \bigcap_{j\in S_2} \overline{E}_j\right] \\ &= \prod_{i\in[r]} \left(1 - \Pr\left[E_{j_i} \middle| \bigcap_{m\in[i-1]} \overline{E}_{j_m} \cap \bigcap_{j\in S_2} \overline{E}_j\right]\right) \\ &\geq \prod_{i\in[r]} (1 - x_{j_i}) && \text{(Induction Hypothesis)} \\ &= \prod_{(k,j)\in E} (1 - x_j)\end{aligned}$$

Using the two bounds we obtain,

$$\Pr\left[E_k \middle| \bigcap_{j\in S} \overline{E}_j\right] \leq \frac{x_k \cancel{\prod_{(k,j)\in E}(1-x_j)}}{\cancel{\prod_{(k,j)\in E}(1-x_j)}} = x_k$$

Therefore we have shown the required claim. Now to finish the proof,

$$\begin{aligned}\Pr[\bigcap_{i\in[n]} \overline{E}_i] &= \prod_{i\in[n]} \Pr\left[\overline{E}_i \middle| \bigcap_{j\in[i-1]} E_j\right] \\ &= \prod_{i\in[n]} \left(1 - \Pr\left[E_i \middle| \bigcap_{j\in[i-1]} E_j\right]\right) \\ &\geq \prod_{i\in[n]} (1-x_i)\end{aligned}$$

□

The above form may be a bit cumbersome to use so, generally the following form is used.

Theorem 13 (Symmetric LLL)**.** *Let $E_1, E_2, ..., E_n$ be events satisfying:*

1. $\forall i \in [n], \Pr[E_i] \leq p$
2. *Maximum degree of some dependency graph of E_i be less than d*
3. *$ep(d+1) \leq 1$ where e is the base of natural logarithm.*

Then,

$$\Pr\left[\bigcap_{i\in[n]} \overline{E}_i\right] > 0$$

Proof. For $d = 0$ it is trivial as the events are mutually independent. For $d \geq 1$, assume the dependency graph has edges E. For $i \in [n]$ take $x_i = \frac{1}{d+1} < 1$, Now,

$$\begin{aligned} x_i \prod_{(i,j)\in E} (1-x_j) &\geq \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d \\ &\geq \frac{1}{e(d+1)} && \left(\because \forall d \geq 1, \left(1-\frac{1}{d+1}\right)^{d+1} > \left(1-\frac{1}{d+1}\right)^{d} > \frac{1}{e}\right) \\ &\geq p \geq \Pr[E_i]\end{aligned}$$

As the x_i along with the events and the dependency (di)graph meet the conditions of LLL,

$$\Pr\left[\bigcap_{i\in[n]} \overline{E}_i\right] \geq \prod_{i\in[n]} (1-x_i) > 0$$

□

The question now is can we have a smaller constant as compared to e in the statement. Let us look at an example, take events $A_1, A_2, ..., A_{d+1}$ with probabilities equal to $\frac{1}{d+1}$. In that case $p(d+1) = 1$ and $\Pr\left[\bigcap_{i\in[n]} \overline{E}_i\right] = 0$. So, 1 clearly does not work. In fact this constant has been shown to be tight by Shearer.

Notice that the necessary condition for LLL to yield $\Pr\left[\bigcap_{i\in[n]} \overline{E}_i\right] > 0$ is $p \leq \frac{1}{e(d+1)} = \Theta(d)$ while for union bound we need $p < \frac{1}{n}$.

The proof of LLL shown is non-constructive. In a breakthrough, Robin Moser gave a proof of constructive version Lovász Local Lemma.

Theorem 14. *Let $E_1, E_2, ..., E_n$ be a set of events in an arbitrary probability space that are determined by mutually independent random variable $y_1, y_2, ..., y_l$, and let $G = (V, E)$ be the dependency graph for these events. Suppose the following hold for d and p:*

1. *Each event E_i is adjacent to at most d other events.*
2. $\Pr[E_i] \leq p$
3. $ep(d+1) \leq 1$

Then there is an assignment of the y_i so that $\bigcap_{i\in[n]} \overline{E}_i$ occurs, and a resampling algorithm that in expectation the number of times that the algorithms resamples E_i is at most $1/d$. Hence the expected running time of the algorithm is at most n/d.

Theorem 15. *Let $E_1, E_2, ..., E_n$ be a set of events in an arbitrary probability space that are determined by mutually independent random variable $y_1, y_2, ..., y_l$, and let $G = (V, E)$ be the dependency graph for these events. Assume there exist $x_1, x_2, ..., x_n \in [0, 1]$ such that, $\forall i \in [n]$,*

$$\Pr[E_i] \leq x_i \prod_{(i,j)\in E} (1 - x_j)$$

Then there is an assignment of the y_i so that $\bigcap_{i\in[n]} \overline{E}_i$ occurs, and a resampling algorithm that in expectation the number of times that the algorithms resamples E_i is at most $x_i/(1 - x_i)$. Hence the expected running time of the algorithm is at most $\sum_{i\in[n]} x_i/(1 - x_i)$.

The high level idea of the algorithms is to start with some assignments y_i. If an event is not satisfied then resample random variables on which the event depends on. The algorithm ensures progress and quick termination given limited dependency.

3.2 Applications of Lovasz Local Lemma

3.2.1 Satisfiability

In the first example, we look at how the lemma can be applied to the satisfiability (SAT) problem. In this problem, there are several clauses. Each clause is a disjunction of literals (a boolean variable or its negation). The goal is to find if there is a satisfying assignment of values to the boolean variables such that all clauses can be satisfied (i.e resulting value of each clause is $True$). In particular, a $k - SAT$ problem has exactly k literals in each of its clauses. Note that any particular clause cannot contain both a variable and its negation. Using the lemma we show that as long as no variable in the $k - SAT$ appears in too many clauses,there is a satisying assignment.

Theorem 16. *If no variable in a k-SAT formula appears in more than $T = 2^k/4k$ clauses, then the formula has a satisfying assignment.*

Proof. Let there are m-clauses in our $k-SAT$ problem and for each i in $i = 1, 2, ..m$ E_i denote the event that i^{th} clause is not satisfied. As a clause is a disjunction operation between k literals, for it to be not satisfied none of the literals can be true.So out of two values $(True, False)$ a literal can take, it has to take $False$. Note that this does not mean that each variable is $False$. Therefore,

$$\Pr(E_i) = 1/2^k \tag{9}$$

And each such event E_i is independent of the events related to clauses,which do not share any variables with clause i. Also note that each of the variable in a clause cannot appear in more than $T = 2^k/4k$ clauses. Now if we construct a dependency graph for these set of events, it can be observed that the degree of each vertex in the dependency graph is bounded by

$$d \leq k(T-1) \leq k.T - k \leq 2^{k-2} - k \tag{10}$$

as each literal appears in at most T clauses,each literal in a particular clause appears in at most $(T-1)$ other clauses.

$$ep(d+1) \leq e.2^{-k}(2^{k-2} - k + 1) \leq e.2^{-k}2^{k-2} \leq \frac{e}{4} \leq 1 \tag{11}$$

The second inequality holds because $k \geq 1$ and hence $1 - k \leq 0$. As all the conditions in the Theorem 14 hold, we can conclude that

$$\Pr\left(\bigcap_{i=1}^{m} \overline{E_i}\right) \geq 0 \tag{12}$$

i.e there exists a satisfying assignment. □

3.2.2 Coloring in HyperGraphs

Definition 17. *A Hypergraph G can be defined as a pair (V, E), where V is a set of vertices, and E is a set of hyperedges between the vertices. Each hyperedge is a set of vertices: $E \subseteq \{\{u, v, ...\} \in 2^V\}$. (Hyperedges are undirected.)*

Here we look at 2-coloring of such a hyper graph. A 2-coloring is an assignment of one of the two colors to each vertex such that no edge is monochromatic i.e. none of the edges have all its vertices assigned to one color. But when would such 2-coloring be possible in a hyper graph? More precisely ,let each edge has at least k vertices and each edge intersects with at most d other edges.
For what values of k, d can a hyper graph have a 2-coloring?

Step 1: Define the probability Space Consider the probability space obtained from randomly assigning one of the two colors to each vertex independently.
Step 2: Define bad events In our example , bad events are the colorings that assign same color to all vertices in an edge.For each edge e,let the event A_e = Event that edge e is monochromatic. Then,

$$\Pr(A_e) = \frac{1}{2^k} \tag{13}$$

Step 3: Construct the Dependency Graph In the dependency graph, let each vertex corresponds to an edge in the hyper graph. And there exists an edge between the two vertices in the dependency graph if the edges in the hyper graph corresponding to these vertices share a vertex. As each edge in hyper graph intersects with at most d other edges, the degree of any vertex in dependency graph is bounded by d.

Step 4: Check for the conditions

1. As each edge in hyper graph intersects with at most d other edges, the degree of any vertex in dependency graph is bounded by d.

2. $p = \Pr(A_e) = \frac{2}{2^k}$

3. $ep(d+1) \leq 1$

To apply the Lovász Local Lemma and conclude that

$$\Pr\left(\bigcap_e \overline{A_e}\right) \geq 0 \tag{14}$$

the third condition must satisfy. Therefore,

$$ep(d+1) \leq e.2^{1-k}(d+1) \leq 1$$

$$d \leq \frac{2^k - 1}{e} - 1 \tag{15}$$

For such values of d,k that satisfy the above condition, a hypergraph is 2-colorable.