Péter Csikvári

Probabilistic Methods

Lecture note

Contents

Preface

It was late afternoon. The last rays of the sun filtered through the curtain. My brother was about to leave after spending the whole day with us. Suddenly he turned to me: "I have a mathematical question. I'm writing a program that produces some log files. My plan is to generate a random integer between 1 and N, and give it as a name of the file. I'm doing this, because the program will run on a network, and I would like to avoid that two different computers produce different log files with the same name. I have an estimate on the number of log files. How large should I choose N to make the probability of the collision negligible?". The problem itself is not very hard. It became popular under the name birthday paradox. This paradox asserts that in case of 23 people the chance that two of them have birthdays on the same day is bigger than 50% assuming that none of them were born on February 29 and the rest 365 days have the same probability. I could instantly provide various bounds on the probability and how to choose N in terms of the number m of log files. My brother was really impressed by the fast answer. But honestly, I was even more impressed. One hundred years ago even mathematicians didn't think about random constructions, and now even programmers think to randomness as a possible solution. Under one hundred year randomness became tool, structure and property at the same time. This course is about taming randomness. I'm pretty sure that ten years later you won't remember much from the course. You will not only not remember what Janson's inequalities are, but you won't even know that they exist. Still you will have a clever feeling: *Maybe randomness can solve the problem.* This is good enough goal for a semester course.

⋆ ⋆ ⋆

So this is the Probabilistic method lecture note. Having it is only a good start, you have to carefully study to learn its content. I suggest you printing it out and carry with yourself to the lectures. If you carry it to the lectures then you don't have to make notes, you can add things to the margins. If you decide not to print it out then use at least a tablet and download the lecture note in advance. The wi-fi is very bad at the lecture halls, and figuring out difficult things (with notations on a different page) on a small smart phone is almost impossible.

Some words about the topics of this course. The primary goal is to learn how to use probabilistic tools to solve discrete mathematical problems. Please, make sure that you have a firm knowledge on probability theory at the beginning of the semester, otherwise you will be constantly lost. The first chapter may give you some impression what kind of things you need to revise.

Some words about the lecturer. Not too bad. If you have problems, please ask his help. He can save you a lot of time by answering questions. Note that even this marvelous lecture note is not capable of directly answering your questions. So, please, don't be shy or too proud to ask questions.

Good luck!

1. Basics

1.1 Notations

Let $\mathbb{Z}$ be the set of integers, and $\mathbb{Z}_+ = \{k \geq 0 | \ k \in \mathbb{Z}\}$ be the set of non-negative integers. Let $[n] = \{1, 2, \ldots, n\}$. The notation $\binom{[n]}{k}$ stands for the k-element subsets of $[n]$.

The expected value of a random variable X is

$$\int_\Omega X dP.$$

If X takes only non-negative integers then

$$\mathbb{E}X = \sum_{k=0}^{\infty} k\mathbb{P}(X = k).$$

The variance of a random variable is

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

The covariance of the random variables X and Y will be denoted by

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y.$$

If $X = X_1 + X_2 + \cdots + X_n$ then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

The conditional probability of an event A with respect to an event B with non-zero probability is defined as follows:

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

One can also the consider the conditional expectation of X with respect to the event B:

$$\mathbb{E}(X|B) = \sum_x x\mathbb{P}(X = x|B).$$

If X and Y are discrete random variables such that $\mathbb{P}(Y = y) \neq 0$ for any $y \in \mathbb{Z}_+$ then $E(X|Y)$ is a function on $\mathbb{Z}_+$ that at evaluated at y as follows:

$$\mathbb{E}(X|Y)(y) = \mathbb{E}(X|Y = y) = \sum_x x\mathbb{P}(X = x|Y = y).$$

This is an important special case of the general definition of the conditional expectation $\mathbb{E}(X|\mathcal{F})$, where $\mathcal{F}$ is a σ-algebra. Note that when we have a finite or countably finite set, say $\{1, 2, \ldots, n\}$ or $\mathbb{Z}$, then a σ-algebra is a particularly simple thing: we need to consider a partition $A_1 \cup \cdots \cup A_k$ of $\{1, 2, \ldots, n\}$ and a σ-algebra consists of those sets that are unions of some A_i's. Note that an $\mathcal{F}$-measurable function is simply a function that is constant on each A_i. The conditional expectation $\mathbb{E}(X|Y)$ corresponds to the case when $\mathcal{F}_Y$ is generated by the sets $A_y = \{Y = y\}$. (In this sense, the notation $\mathbb{E}(X|Y)(y)$ was a tiny cheating as $\mathbb{E}(X|Y)$ is defined on the whole set, say $\{1, 2, \ldots, n\}$, not on the elements A_i's, it is only constant on the A_i's.) So for a $k \in \{1, 2, \ldots, n\}$

$$\mathbb{E}(X|\mathcal{F})(k) = \sum_r r\mathbb{P}(X = r|r \text{ and } k \text{ are in the same set } A_i).$$

In words, we take the set A_i of the partition containing k, and we average according to the corresponding conditional probability $\mathbb{P}(X = r)/\mathbb{P}(X \in A_i)$.

1.2 Useful inequalities

Proposition 1.2.1. *For all $x \in \mathbb{R}$ we have $1 + x \leq e^x$.*

Proof. If $x > 0$ then

$$1 + x \leq \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x.$$

If $x \leq -1$ then the claim is trivial. If $-1 \leq x \leq 0$, then set $y = -x \geq 0$. Then

$$\frac{1}{1-y} = \sum_{k=0}^{\infty} y^k \geq \sum_{k=0}^{\infty} \frac{y^k}{k!} = e^y.$$

Hence $e^x = e^{-y} \geq 1 - y = 1 + x$. □

Proposition 1.2.2. *We have*

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$

Proof. Since $\binom{n}{k} \leq \frac{n^k}{k!}$, it is enough to prove that $k! \geq \left(\frac{k}{e}\right)^k$. This is indeed true:

$$e^k \geq \prod_{j=1}^{k-1} \left(1 + \frac{1}{j}\right)^j = \prod_{j=1}^{k-1} \frac{(j+1)^j}{j^j} = \frac{k^{k-1}}{(k-1)!} = \frac{k^k}{k!}.$$

□

1.3 Basic inequalities in probability theory

We recall some basic inequalities.

Proposition 1.3.1 (Union bound)**.**

$$\mathbb{P}\left(\bigcup_{i=1}^m A_i\right) \leq \sum_{i=1}^m \mathbb{P}(A_i).$$

Theorem 1.3.2 (Markov's inequality)**.** *Let X be a non-negative random variable with $\mathbb{E}X > 0$. Then for arbitrary positive λ we have*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}X}{\lambda}.$$

Proof.

$$\mathbb{E}X = \int X dP \geq \int_{\{X \geq \lambda\}} X dP \geq \int_{\{X \geq \lambda\}} \lambda dP = \lambda \mathbb{P}(X \geq \lambda).$$

□

A simple corollary of Markov's inequality is Chebyshev's inequality.

Theorem 1.3.3 (Chebyshev's inequality)**.** *Let X be a random variable with $\mathbb{E}X = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Then*

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Proof. Let us apply Markov's inequlity to the random variable $Y = (X - \mu)^2$. Then $\mathbb{E}Y = \mathrm{Var}(X) = \sigma^2$ by definition.

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) = \mathbb{P}(Y \geq \lambda^2\sigma^2) \leq \frac{\mathbb{E}Y}{\lambda^2\sigma^2} = \frac{1}{\lambda^2}.$$

□

$\star \quad \star \quad \star$

In this lecture note we would like to prove combinatorial theorems and so many times the studied random variable X takes only non-negative integer values. In fact, often one can translate the combinatorial statement to a statement about the probability that a random variable takes the value 0. This motivates us to collect some results on estimates of $\mathbb{P}(X = 0)$.

Theorem 1.3.4. *If X takes only non-negative integer values then*

$$\mathbb{P}(X = 0) \geq 1 - \mathbb{E}X.$$

Proof. We have

$$\mathbb{P}(X > 0) = \sum_{k=1}^{\infty} \mathbb{P}(X = k) \leq \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \mathbb{E}X,$$

or equivalently,

$$\mathbb{P}(X = 0) \geq 1 - \mathbb{E}X.$$

□

This implies that for instance, if the sequence of random variables X_n satisfy that $\lim_{n\to\infty} \mathbb{E}X_n = 0$ then

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0) = 1.$$

However, $\mathbb{E}X_n \to \infty$ does not guarantee that

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0) = 0.$$

To phrase such a statement we also need the variance of the random variables X_n.

Theorem 1.3.5.

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}.$$

Proof. Let us use Chebyshev-inequality.

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}X| \geq \mathbb{E}X) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}.$$

□

Remark 1.3.6. For non-negative random variables the above inequality can be improved as follows:

$$\mathbb{P}(X=0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X^2)}.$$

Since $\mathbb{E}(X^2) \geq (\mathbb{E}X)^2$ this is indeed an improvement. The proof of this inequality is a simple application of the Cauchy–Schwarz inequality: set $A = \{\omega \mid X(\omega) > 0\}$, then

$$\left(\int_A X dP\right)^2 \leq \left(\int_A 1 dP\right)\left(\int_A X^2 dP\right),$$

that is

$$(\mathbb{E}X)^2 \leq (1 - \mathbb{P}(X=0))(\mathbb{E}(X^2)).$$

After some algebraic manipulation we get that

$$\mathbb{P}(X=0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X^2)}.$$

Theorem 1.3.5 implies that if $\lim_{n\to\infty} \frac{\mathrm{Var}(X_n)}{(\mathbb{E}X_n)^2} = 0$ then

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0) = 0.$$

We can also see that if $\mathrm{Var}(X_n) = o((\mathbb{E}X_n)^2)$ then X_n is concentrated around $\mathbb{E}X_n$ which we can simply denote by $X_n \sim \mathbb{E}X_n$.

2. Existence results

In mathematics we often face with the problem that we need to prove that a certain structure S exists with a required property P. In most cases we simply prove the existence by constructing the required structure S. Unfortunately, sometimes this route does not work and we can only give an existence proof, a proof that does not give much besides the existence. A popular tool providing such proofs is for instance the pigeonhole principle.

In this whole course, we study another tool. This is the so-called probabilistic method. Using this method we show that in a certain probability space the required structure S exists with positive probability. In this chapter we give the most basic examples of this method where one only needs to use the union bound, Proposition 1.3.1.

2.1 Diagonal Ramsey numbers

Recall that the Ramsey-number $R(r, b)$ denotes the smallest n such that no matter how we color the edges of the complete graph K_n with red and blue colors it will either contain an induced red K_r or a blue K_b. Note that the definition implies that for $n = R(r, b) - 1$ there is a coloring of K_n without red K_r and blue K_b.

Theorem 2.1.1 (Erdős)**.** *Suppose that the positive integers n, k satisfy the inequality $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$. Then $R(k, k) > n$. In particular, $R(k, k) > \lfloor 2^{k/2} \rfloor$ if $k \geq 3$.*

Proof. We need to show that there exists a coloring of the edge set of K_n that does not contain either monochromatic red or blue clique K_k. Let us color each edges with color red or blue with probability $1/2$ independently of each other. Now let us estimate the probability that the coloring is bad, i. e., it contains a monochromatic red or blue K_k. For each $S \subset V(G)$ with $|S| = k$ let A_S be the event the induced

subgraph on S is monochromatic. Then

$$\mathbb{P}(\text{coloring is bad}) \leq \sum_{|S|=k} \mathbb{P}(A_S) = \binom{n}{k} \frac{2}{2^{\binom{k}{2}}}.$$

By the condition of the theorem $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$, so the probability that the coloring is good is positive.

Next we show that for $k \geq 3$ and $n = \lfloor 2^{k/2} \rfloor$ the condition of the theorem is satisfied. Indeed,

$$\binom{n}{k} 2^{1-\binom{k}{2}} < \frac{n^k}{k!} 2^{1-\binom{k}{2}} \leq \frac{2^{k^2/2}}{k!} 2^{1-\binom{k}{2}} = \frac{2^{(k+2)/2}}{k!} < 1.$$

if $k \geq 3$. □

2.2 Tournaments

Definition 2.2.1. A tournament is a complete directed graph. A tournament D is called k-dominated if for every k vertices $v_1, \ldots, v_k$ there exists a vertex u such that $(u, v_i) \in E(D)$ for $i = 1, \ldots, k$.

Theorem 2.2.2 (Erdős [10])**.** *If n is large enough then there exists a k-dominated tournament on n vertices.*

Proof. Let us orient each edge with probability $1/2-1/2$ independently of each other. Then the probability that for a given set of vertices $v_1, \ldots, v_k$ there is no u such that all $(u, v_i) \in E(D)$ is $(1-1/2^k)^{n-k}$. Hence the probability that the orientation is bad is at most

$$\binom{n}{k} \left(1 - \frac{1}{2^k}\right)^{n-k}.$$

A little computation shows that if $\frac{n}{\ln n} > k2^k$, then this is less than 1. For large k this is satisfied if $n > k^2 2^k$. Hence with positive probability there exists a k-dominated tournament. □

Remark 2.2.3. The computation can be carried out using the bound $1 + x < e^x$:

$$\begin{aligned}\binom{n}{k} \left(1 - \frac{1}{2^k}\right)^{n-k} &\leq \frac{n^k}{k!} \exp\left(-\frac{1}{2^k}(n-k)\right) \\ &\leq \frac{1}{k!} \exp\left(\frac{k}{2^k}\right) \cdot \exp\left(k \ln n - \frac{n}{2^k}\right)\end{aligned}$$

$$\leq \frac{e}{k!} \exp\left(k \ln n - \frac{n}{2^k}\right)$$
$$\leq \frac{e}{k!} < 1$$

if $\frac{n}{\ln n} > k2^k$.

3. First Moment Method

In the previous chapter we have seen some very simple ideas how to find a certain structure S by proving that it exists with positive probability just by using union bound. Here we study another very simple technique. This is the so-called first moment method. In many cases the structure S that we need to find is defined through some parameter $f(S)$. For instance, we need to prove that there exists a structure S for which some parameter $f(S)$ satisfies $f(S) \geq \rho$. If we find a probability space in which the expected value of $f(S)$ is bigger or equal to ρ then we can immediately conclude that $f(S) \geq \rho$ with positive probability.

3.1 Warm-up: large bipartite subgraphs

Theorem 3.1.1 ([4]). *Let G be a graph with n vertices and $e(G)$ edges. Then G has a bipartite subgraph with at least $e(G)/2$ edges.*

Proof. One can rephrase the statement of the theorem as follows: there exists a cut $(A, V \setminus A)$ of G such that the number of edges $(e(A, V \setminus A))$ contained in the cut is at least $e(G)/2$.

Let us consider the random set A which contains every $v \in V(G)$ with probability $1/2$ independently of each other. (This way we have defined a probability space.) Let us consider the random variable $X = e(A, V \setminus A)$. We have to show that with positive probability $X \geq e(G)/2$. To this end it is enough to show that $\mathbb{E}X = e(G)/2$. This is indeed true. For every edge $f \in E(G)$ let us introduce the indicator random variable X_f which takes value 1 if f is in the cut $(A, V \setminus A)$, and 0 otherwise. Then

$$\mathbb{E}X = \mathbb{E}\left(\sum_{f \in E(G)} X_f\right) = \sum_{f \in E(G)} \mathbb{E}X_f.$$

(Note that the random variables X_f are not necessarily independent, but the linearity of expectation holds true even with non-independent random variables.) For all

$f \in E(G)$ we have $\mathbb{E}X_f = 1/2$ since the end points of f are in the same set with probability $1/2$ and they are in different sets with probability $1/2$. Hence

$$\mathbb{E}X = \sum_{f \in E(G)} \mathbb{E}X_f = \sum_{f \in E(G)} \frac{1}{2} = \frac{1}{2} e(G).$$

We are done! □

3.2 Independent sets

Theorem 3.2.1 (Caro; Wei)**.** *Let G be a graph with vertex degrees $d_1, \ldots, d_n$. Let $\alpha(G)$ be the size of the largest independent set of the graph G. Then*

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i + 1}.$$

Proof. Consider a random permutation of the vertices. Let us encircle all the vertices that precede all their neighbors in the given order. Let $X(\pi)$ be the random variable that counts the number of encircled vertices. For a given vertex $v \in V(G)$ let X_v be the indicator variable that the the vertex v is encircled or not. Then $X = \sum_{v \in V(G)} X_v$, consequently

$$\mathbb{E}X = \sum_{v \in V(G)} \mathbb{E}X_v.$$

Note that for a vertex v we have $\mathbb{E}X_v = \frac{1}{d_v+1}$ since the probability that v precedes its neighbors is the same as saying that v is the first among $d_v + 1$ vertices in a random permutation, and this probability is clearly $\frac{1}{d_v+1}$. Hence

$$\mathbb{E}X = \sum_{v \in V(G)} \mathbb{E}X_v = \sum_{i=1}^{n} \frac{1}{d_i + 1}.$$

With positive probability X is at least as large as this expected value. On the other hand, in an arbitrary order the encircled vertices form an independent set since if two of them were adjacent, then the second of the two vertices in the order would not be encircled. Hence

$$\alpha(G) \geq \mathbb{E}X = \sum_{i=1}^{n} \frac{1}{d_i + 1}$$

as required. □

Remark 3.2.2. From the above proof one can easily deduce Turán's theorem.

3.3 Crossing number

Theorem 3.3.1 (Ajtai-Chvátal-Newborn-Szemerédi [2]; Leighton)**.** *Let G be a graph with n vertices and e edges. Let $X(G)$ be the crossing number of the graph G. If $e \geq 4n$, then*

$$X(G) \geq \frac{e^3}{64n^2}.$$

Proof. Recall that any planar graph with n vertices has at most $3n-6$ edges. Consequently, if G is a graph with n vertices and e edges, then the crossing number is at least $e-(3n-6)$ (why?). So

$$X(G) \geq e(G) - 3v(G).$$

(The +6 won't be important for us.) This is of course a weaker statement than what we want to prove. The key idea of the better bound is to apply this weak inequality to a random subgraph of G. Set $0 \leq p \leq 1$ and consider the random subgraph of G where we keep each vertex with probability p and delete it with probability $1-p$. Let G_p be the obtained graph. Then

$$\mathbb{E}v(G_p) = pv(G) \quad \text{and} \quad \mathbb{E}e(G_p) = p^2 e(G),$$

since the probability that we keep an edge is p^2, the probability that we keep both end points of the edge. We need to be a bit more careful with $\mathbb{E}X(G_p)$. Starting from an optimal drawing of G, the probability that a crossing remains is p^4 since all four vertices determining the crossing should remain. This means that starting from an optimal drawing of G the expected value of the crossing number of G_p is $p^4 X(G)$. However, it may happen that G_p has a better drawing with smaller number of crossings. So all we can say is that

$$\mathbb{E}X(G_p) \leq p^4 X(G).$$

Hence

$$p^4 X(G) - p^2 e(G) + 3pv(G) \geq \mathbb{E}X(G_p) - \mathbb{E}e(G_p) + 3\mathbb{E}v(G_p) =$$
$$= \mathbb{E}(X(G_p) - e(G_p) + 3v(G_p)) \geq 0.$$

Whence $p^4 X(G) - p^2 e(G) + 3pv(G) \geq 0$ for all $0 \leq p \leq 1$. Now let us choose p to be $\frac{4v(G)}{e(G)}$. This is at most 1 according to the assumption of the theorem. Then

$$X(G) \geq p^{-2} e(G) - 3p^{-3} v(G) = \frac{e(G)^3}{64 v(G)^2}.$$

This is exactly what we wanted to prove. □

Since it is not clear how we can use such a statement let us consider a corollary of this theorem. Then later we even consider a corollary of this corollary.

Suppose that we are given some points and lines on the plane. Let $\mathcal{P}$ be the set of points, and $\mathcal{L}$ be the set of lines. The number of point-line incidences is exactly what we expect:

$$I(\mathcal{P}, \mathcal{L}) = |\{(P, L) \in \mathcal{P} \times \mathcal{L} \mid P \in L\}|.$$

Let $I(n, m)$ be the maximal number of incidences given n points and m lines:

$$I(n, m) = \max_{|\mathcal{P}|=n, |\mathcal{L}|=m} I(\mathcal{P}, \mathcal{L}).$$

The following theorem gives a good bound on $I(n, m)$.

Theorem 3.3.2 (Szemerédi-Trotter [17])**.**

$$I(n, m) \leq 4(m^{2/3}n^{2/3} + m + n).$$

Proof. Let us consider the graph G whose vertices are the elements of the set $\mathcal{P}$, i. e., the points, and two points are adjacent if there is a line $\ell \in \mathcal{L}$ that contains the two points next to each other.

First let us determine the number of edges of the graph G. If a line contains k points then it determines $k - 1$ edges. Hence the number of edges is $I(\mathcal{P}, \mathcal{L}) - m$. Next let us give an upper bound on $X(G)$. Two edges intersect each other if two lines intersect each other. Hence $X(G)$ is at most $\binom{m}{2}$. If $e(G) < 4n$ then

$$I(\mathcal{P}, \mathcal{L}) < 4n + m < 4(m^{2/3}n^{2/3} + m + n).$$

If $e(G) \geq 4n$, then we can use the previous theorem:

$$\binom{m}{2} \geq X(G) \geq \frac{e(G)^3}{64n^2} = \frac{(I(\mathcal{P}, \mathcal{L}) - m)^3}{64n^2}.$$

Thus

$$I(\mathcal{P}, \mathcal{L}) \leq (32m^2n^2)^{1/3} + m < 4(m^{2/3}n^{2/3} + m + n).$$

Hence

$$I(n, m) \leq 4(m^{2/3}n^{2/3} + m + n).$$

□

Remark 3.3.3. We used very little information about the lines. We simply used that two lines have at most one intersection. We could have considered circles or arbitrary curves of degree at most d, these curves have also bounded number of intersections. Naturally, the constants in the theorem would have been worse, but still we would have received a bound of type $O_d(n^{2/3}m^{2/3} + n + m)$ for the number of incidences.

In what follows we consider a nice application of the Szemerédi-Trotter theorem. Let $A \subset \mathbb{R}$ be a finite set, and let

$$A + A = \{a + a' \mid a, a' \in A\}$$

and

$$A \cdot A = \{a \cdot a' \mid a, a' \in A\}.$$

If $A = \{1, 2, \dots, n\}$, then $A + A = \{2, \dots, 2n\}$, and so $|A + A| = 2n - 1$. However, in this case we have $|A \cdot A| = \Omega\left(\frac{n^2}{(\log n)^\alpha}\right)$. If $A = \{1, 2, 2^2, \dots, 2^{n-1}\}$ then $|A \cdot A| = 2n - 1$, but then we have $|A + A| = \binom{n}{2}$. After checking several examples one will have the feeling that one of the sets should be large. This is a well-known conjecture:

Conjecture 3.3.4 (Erdős-Szemerédi)**.** For all $\varepsilon > 0$ there exists a constant $c(\varepsilon)$ such that for all finite set $A \subset \mathbb{R}$ we have

$$|A + A| + |A \cdot A| \geq c(\varepsilon)|A|^{2-\varepsilon}.$$

We are very far from proving this conjecture. The following result of György Elekes was a real breakthrough in 1997, and it opened the way of geometric arguments in additive combinatorics.

Theorem 3.3.5 (Elekes [8])**.** *Let $A \subset \mathbb{R}$ be a finite set. Then*

$$|A + A| \cdot |A \cdot A| \geq c|A|^{5/2}.$$

In particular,

$$|A + A| + |A \cdot A| \geq c'|A|^{5/4}.$$

Proof. Let $P = \{(a, b) \ | a \in A + A, b \in A \cdot A\}$. This is a point set on the plane of size $|A + A||A \cdot A|$.

Let us consider the lines of following type:

$$\ell_{a,b} = \{(x, y) \mid y = a(x - b)\},$$

where $a, b \in A$. Let L be the set of these lines. Then $|L| = |A|^2$. Every such line contains $|A|$ points form P: $(b + c, ac) \in \ell_{a,b}$ if $c \in A$. Whence $I(P, L) \geq |A|^3$. According to Szemerédi-Trotter theorem we have

$$|A|^3 \leq 4((|A + A| \cdot |A \cdot A|)^{2/3}(|A|^2)^{2/3} + |A + A| \cdot |A \cdot A| + |A|^2).$$

From this the statement of the theorem follows after a little computation. □

Remark 3.3.6. Currently, the best-known result is due to József Solymosi [16]:

$$|A + A| + |A \cdot A| \geq c(\varepsilon)|A|^{4/3-\varepsilon}.$$

More precisely, Solymosi showed that

$$|A \cdot A| \cdot |A + A|^2 \geq \frac{|A|^4}{4\lceil \ln |A| \rceil},$$

consequently,

$$\max(|A \cdot A|, |A + A|) \geq \frac{|A|^{4/3}}{2\lceil \ln |A| \rceil^{1/3}}.$$

4. Alteration

In this chapter we study a method called the altered first moment method. It is a slightly bit more tricky than the first moment method. Here the randomly chosen structure S will not be immediately good, but will be bad just a little bit so that we can fix the bad part of the structure. In practice, there will be a parameter $f(.)$ that measures the badness of the structure (or if there is a given parameter $f(.)$ already, then we prepare a new parameter $f'(.)$ measuring $f(.)$ and the badness at the same time). If the expected value of this badness parameter is small, then with positive probability we can find a random structure that we can fix later. After the examples it will be clear how this method works.

4.1 Independent sets in graphs and hypergraphs

Theorem 4.1.1. *Let H be an r-uniform hypergraph with n vertices and $e(H)$ edges. Suppose that $n \leq 2e$. Then there exists a set $S \subseteq V(H)$ inducing no edge such that*

$$|S| \geq \frac{1}{2}\left(\frac{n}{2e(H)}\right)^{1/(r-1)} n.$$

Proof. Let T be a random subset of the vertex set chosen as follows: we choose each element of V to be in T with probability p. We will choose p later. Then $\mathbb{E}|T| = pn$ and for the number of edges induced by T we have $\mathbb{E}(e(T)) = p^r e(H)$. Then

$$\mathbb{E}(|T| - e(T)) = pn - p^r e(H).$$

Let

$$p = p_0 = \left(\frac{n}{2e(H)}\right)^{1/(r-1)}.$$

Then $p_0 n - p_0^r e(H) = p_0 n/2$. Therefore,

$$\mathbb{E}(|T| - e(T)) = \frac{1}{2} p_0 n.$$

Hence there must be a set T for which $|T| - e(T) \geq \frac{1}{2}p_0 n$. Let $S \subseteq T$ be a set obtained from T by deleting one vertex of each edge of T. Then S induces no edge and

$$|S| \geq |T| - e(T) \geq \frac{1}{2}p_0 n = \frac{1}{2}\left(\frac{n}{2e(H)}\right)^{1/(r-1)} n.$$

□

Remark 4.1.2. In the case of graphs, that is $r = 2$, this theorem says that

$$\alpha(G) \geq \frac{n^2}{4e(G)}.$$

This is always weaker than the bound

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i + 1}$$

obtained earlier.

We could have chosen p in a bit better way by simply choosing it such a way that it maximizes $pn - p^r e(H)$. This would have yielded the bound

$$\alpha(H) \geq \frac{r-1}{r}\left(\frac{n}{re(H)}\right)^{1/(r-1)} n.$$

4.2 Ramsey-numbers revisited

Theorem 4.2.1 ([4])**.** *For all n and k we have $R(k,k) > n - \binom{n}{k}2^{1-\binom{k}{2}}$.*

Proof. Let us color the edges of a complete graph K_n with red and blue. Let X be the number of monochromatic K_k. Then

$$\mathbb{E}X = \binom{n}{k}2^{1-\binom{k}{2}}.$$

So there must be a coloring with at most as many monochromatic K_k. Now let us delete one vertex from each monochromatic K_k. Then the number of vertices is at least $n - \binom{n}{k}2^{1-\binom{k}{2}}$ and the resulting graph has no monochromatic K_k. Hence $R(k,k) > n - \binom{n}{k}2^{1-\binom{k}{2}}$. □

Remark 4.2.2. A careful analysis shows that this bound implies that

$$R(k,k) \geq \frac{1}{e}(1 + o(1))k2^{k/2}$$

while our previous argument only gave

$$R(k,k) \geq \frac{1}{\sqrt{2}e}(1+o(1))k2^{k/2}.$$

Later we will show by Lovász local lemma that

$$R(k,k) \geq \frac{\sqrt{2}}{e}(1+o(1))k2^{k/2}.$$

4.3 Dominating sets in graphs

Theorem 4.3.1 ([4]). *Let $G = (V, E)$ be a graph with n vertices and minimum degree $\delta > 1$. Then it has a dominating set of size at most*

$$n\frac{1+\ln(\delta+1)}{\delta+1}.$$

(A set U is called a dominating set of G if all $v \in V \setminus U$ has some neighbor u in U.)

Proof. The strategy is the following: we choose a random subset S and let $T = T(S)$ be the set of vertices v such that neither v, nor any of the neighbors of v are in the set S. Then $S \cup T$ is a dominating set. Let us choose S as follows: we choose each vertex v into S with probability p. We will choose p later. Then for any vertex $v \in V$ we have

$$\mathbb{P}(v \in T) = (1-p)^{1+d(v)} \leq (1-p)^{1+\delta} \leq e^{-p(\delta+1)}$$

since neither v, nor any of the neighbors of v are in the set S. Hence

$$\mathbb{E}(|S| + |T|) = \mathbb{E}|S| + \mathbb{E}|T| \leq n(p + e^{-p(\delta+1)}).$$

Let

$$p = \frac{\ln(\delta+1)}{\delta+1}.$$

Then

$$\mathbb{E}(|S| + |T|) \leq n(p + e^{-p(\delta+1)}) = \frac{n(1+\ln(\delta+1))}{\delta+1}.$$

Hence with positive probability there must be a dominating set of at most this size. □

4.4 Graphs with large chromatic number and girth

Theorem 4.4.1 (Erdős [9]). *For arbitrary (k, ℓ) there exists a graph G whose chromatic number is at least k and the length of its shortest cycle is at least ℓ.*

Proof. Let $G(n,p)$ be the random graph with n vertices such that we draw all edges with probability $p = p(n)$ independently of each other. In this proof we will set $p = n^{-\alpha}$, where $\alpha \geq 0$ is a parameter chosen later. First we estimate the number of cycles shorter than ℓ . Given vertices $v_1 v_2 \dots v_r$ form a cycle if $v_i v_{i+1}$ $(r+1=1)$ are all edges, the probability of this event is p^r. Naturally, we can choose the sequence $v_1 v_2 \dots v_r$ in $n(n-1)\dots(n-r+1)$ ways, we only have to take take into account that we counted the same cycle $2r$ ways (rotated and reflected copies). Let X be the random variable counting the number of cycles of length at most $\ell - 1$. Furthermore, let $X(v_1 \dots v_r)$ $(r \leq \ell - 1)$ be the indicator random variable that the vertices $v_1 \dots v_r$ form a cycle in this order. Then

$$X = \sum_{r, v_1 \dots v_r} X(v_1 \dots v_r).$$

Hence

$$\mathbb{E}X = \sum_{r, v_1 \dots v_r} \mathbb{E}X(v_1 \dots v_r) = \sum_{r=3}^{\ell-1} \frac{n(n-1)\dots(n-r+1)}{2r} p^r \leq \sum_{r=3}^{\ell-1} \frac{(np)^r}{2r}.$$

Set $M = \sum_{r=3}^{\ell-1} \frac{(np)^r}{2r}$. Suppose that with some choice of p we can ensure that M is small, then with positive probability the number of cycles of length at most $\ell - 1$ will be at most M and by throwing out one point from each cycle we get a graph on at least $n - M$ vertices that does not contain a cycle of length at most $\ell - 1$. In fact, we need to be a little bit more careful as we need that the number of short cycles is small with large probability. Fortunately, we get it immediately: with probability at least $1/2$ the number of cycles of length at most $\ell - 1$ is at most $2M$. Otherwise the expected value would be bigger than M.

Before we try to chose p appropriately let us see how we can bound the chromatic number $\chi(G)$ of G. Here we use the simple fact that

$$\chi(G) \geq \frac{n}{\alpha(G)}.$$

This is true since all color class induces an independent set so its size is at most $\alpha(G)$, so we need at least $\frac{n}{\alpha(G)}$ colors to color G. So to make $\chi(G)$ large, it is enough

to ensure that $\alpha(G)$ is small. Let us bound the probability that $\alpha(G) \geq s$. For a set S of size s let A_S be the event that S does not induce any edge. Then

$$\mathbb{P}(\alpha(G) \geq s) \leq \sum_{|S|=s} \mathbb{P}(A_S) = \binom{n}{s}(1-p)^{\binom{s}{2}} \leq n^s(1-p)^{\binom{s}{2}} \leq (ne^{-p(s-1)/2})^s.$$

(In the last step we used the fact that $1 + x \leq e^x$ is satisfied for all x. This is a rather standard bound that is quite good if x is small.)

Now it is clear what we have to keep in mind: let M be small, so we need a small p, but we also need that s is not too large and so we need that $ne^{p(s-1)/2} < 1$. We can easily achieve it as follows: set $p = n^{\theta-1}$ where $\theta = \frac{1}{2(\ell-1)}$ and $s = \lceil \frac{3}{p} \log n \rceil$. Then

$$M = \sum_{r=3}^{\ell-1} \frac{(np)^r}{2r} \leq n^{\theta(\ell-1)} \sum_{r=3}^{\ell-1} \frac{1}{2r} \leq n^{1/2} \log n \leq \frac{n}{4}$$

if n is large enough. On the other hand,

$$\mathbb{P}(\alpha(G) \geq s) \leq (ne^{-p(s-1)/2})^s \leq 1/4$$

if n is large enough. Since $\mathbb{P}(X \geq 2M) \leq 1/2$ and $\mathbb{P}(\alpha(G) \geq s) \leq 1/4$, with positive probability there exists a graph where the number of short cycles is at most $n/2$ and $\alpha(G) \leq s$. Now from all cycles of length at most $\ell - 1$ let us throw out 1 vertex and let G^* be the obtained graph. Then G^* has at least $n/2$ vertices and it does not contain a cycle of length at most $\ell - 1$. Furthermore, $\alpha(G^*) \leq \alpha(G)$ since G^* is an induced subgraph of G. Then

$$\chi(G^*) \geq \frac{|V(G^*)|}{\alpha(G^*)} \geq \frac{n/2}{3n^{1-\theta} \log n} = \frac{n^\theta}{6 \log n}.$$

If n is large enough this is bigger than k. We are done! □

5. Second Moment method

In the previous chapters we used the union bounds and the first moment method. These techniques are very powerful due to the fact that they do not require any information on the dependence of the random variables.

In this section we see some applications of the second moment method which roughly means that we use Chebyshev's inequality as a new ingredient in our proofs. We will see that at least we need some partial information about the dependence of the random variables, but not too much. Generally quite crude bounds will be enough to achieve our goals.

This section is based on the corresponding chapter of the book The Probabilistic Method by Noga Alon and Joel Spencer.

In this section we study the threshold function of random graphs. This topic was initiated in the seminal paper [11] of Erdős és Rényi: *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. 5 (1960), 17-61. In fact all results of this section can be found in this paper. This paper is on the internet in a scanned form.

5.1 General approach

From Chapter 1 we know that for a non-negative random variable X taking only non-negative integers we have

$$1 - \mathbb{E}X \leq \mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}.$$

These two inequalities will play a major role in this chapter. We will often encounter the situation that having some property is equivalent with some random variable taking value 0. Hence if $\mathbb{E}X$ is small then $\mathbb{P}(X = 0)$ is large, and so the random structure has the desired property with large probability. On the other hand, if

$\frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}$ is small then the random structure doesn't have the desired property with large probability.

Often we will encounter with a sequence of structures, notably a sequence of random graphs $G(n,p)$. In this case X will be some X_n in a sequence. As we will see it is also worth considering separately the case when $X_n = X_1^{(n)} + X_2^{(n)} + \cdots + X_m^{(n)}$ where $X_i^{(n)}$ are indicator random variables. Let $X_i^{(n)}$ be the indicator random variable of the event $A_i^{(n)}$. Let us introduce the notation $i \sim j$ if $A_i^{(n)}$ and $A_j^{(n)}$ are not independent. Then it is also worth introducing the following sum:

$$\Delta_n = \sum_{i \sim j} \mathbb{P}(A_i^{(n)} \cap A_j^{(n)}).$$

(In this sum both (i,j) and (j,i) appear.) If $\mathbb{P}(A_i^{(n)}) = p_i^{(n)}$ then

$$\mathrm{Var}(X_i^{(n)}) = \mathbb{E}(X_i^{(n)})^2 - (\mathbb{E}X_i^{(n)})^2 = p_i^{(n)} - (p_i^{(n)})^2 \leq p_i^{(n)} = \mathbb{E}X_i^{(n)}.$$

Furthermore,

$$\mathrm{Cov}(X_i^{(n)}, X_j^{(n)}) = \mathbb{E}(X_i^{(n)} X_j^{(n)}) - \mathbb{E}X_i^{(n)} \cdot \mathbb{E}X_j^{(n)} \leq \mathbb{E}(X_i^{(n)} X_j^{(n)}) = \mathbb{P}(A_i^{(n)} \cap A_j^{(n)}).$$

Using these inequalities we get that

$$\begin{aligned}
\mathrm{Var}(X_n) &= \sum_{i=1}^{n} \mathrm{Var}(X_i^{(n)}) + 2 \sum_{i<j} \mathrm{Cov}(X_i^{(n)}, X_j^{(n)}) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i^{(n)}) + \sum_{i \sim j} \mathrm{Cov}(X_i^{(n)}, X_j^{(n)}) \\
&\leq \sum_{i=1}^{n} \mathbb{E}X_i^{(n)} + \sum_{i \sim j} \mathbb{P}(A_i^{(n)} \cap A_j^{(n)}) \\
&= \mathbb{E}X_n + \Delta_n.
\end{aligned}$$

Here we used the fact that if $i \not\sim j$, equivalently $A_i^{(n)}$ and $A_j^{(n)}$ are independent, then $\mathrm{Cov}(X_i^{(n)}, X_j^{(n)}) = 0$. Hence

$$\mathrm{Var}(X_n) \leq \mathbb{E}X_n + \Delta_n.$$

Hence Theorem 1.3.5 implies the following statement.

Theorem 5.1.1. *Suppose that* $\mathbb{E}X_n \to \infty$ *and* $\Delta_n = o((\mathbb{E}X_n)^2)$.
Then $\mathbb{P}(X_n > 0) \to 1$.

It is worth doing some extra work with Δ_n. Many times the indicator random variables $X_1^{(n)}, \dots, X_m^{(n)}$ have a symmetric role, in other words, for all i and j there is an automorphism of the underlying space that takes $A_i^{(n)}$ to $A_j^{(n)}$. Then

$$\Delta_n = \sum_{i \sim j} \mathbb{P}(A_i^{(n)} \cap A_j^{(n)}) = \sum_i \mathbb{P}(A_i^{(n)}) \sum_{j \sim i} \mathbb{P}(A_j^{(n)} \mid A_i^{(n)}).$$

The inner sum is independent of i, because of the symmetry:

$$\Delta_n^* = \sum_{j \sim i} \mathbb{P}(A_j^{(n)} \mid A_i^{(n)}).$$

Hence

$$\Delta_n = \sum_i \mathbb{P}(A_i^{(n)}) \Delta_n^* = \Delta_n^* \sum_i \mathbb{P}(A_i^{(n)}) = \Delta_n^* \mathbb{E} X_n.$$

So in this case we get the following theorem

Theorem 5.1.2. *Suppose that* $\mathbb{E}X_n \to \infty$ *and* $\Delta_n^* = o(\mathbb{E}X_n)$. *Then* $\mathbb{P}(X_n > 0) \to 1$.

Remark 5.1.3. (**Important!**) It is rather inconvenient to write out the $.^{(n)}$ every time: $X_i^{(n)}, A_i^{(n)}, p_i^{(n)}$... So in what follows we hide the notation n and for instance the last claim will read as follows: "Suppose that $\mathbb{E}X \to \infty$ and $\Delta^* = o(\mathbb{E}X)$. Then $\mathbb{P}(X > 0) \to 1$." This is of course completely stupid if we forget that there is a hidden parameter n. Nevertheless, the parameter n will always be clear from the context. For instance, if we study the random graph $G(n, p(n))$ and X is the number of K_4 in the graph then it is clear that actually $X = X_n$ belongs to $G(n, p(n))$.

5.2 Threshold functions of graph appearance

Let $G(n, p)$ be the random graph on n vertices whose edges appear with probability p independently of each other. The probability p may depend on n, for instance, it can be $p = p(n) = n^{-1/2}$.

Surprisingly, one can see "all" graphs $G(n, p)$ at the same time as p runs from 0 to 1. For all edges let us pick a random number from the interval $[0, 1]$, then just as we rotate the frequency finder of a radio we start to increase p. At some point t the edges with a number less than t will lit up. As we increase t more and more edges will lit up. At point $t = 0$ the whole graph is dark (with probability 1, while at $t = 1$ the whole graph is lit up. At point p we can see $G(n, p)$. This process is called the evolution.

What kind of questions can we study? We can for instance ask the probability that $G(n,p)$ contains a Hamiltonian-cycle or we can seek for the probability that the graph is a planar graph or the probability that its chromatic number is at most 100. For a fixed n these questions might be very difficult to answer and answers might be very ugly. In general, we only wish to know the answer as the number of vertices tends to infinity. In other words, we are seeking $\lim \mathbb{P}(G(n,p) \in P)$ for some property P like containing Hamiltonian-cycle or not. Actually, we will be even less ambitious as we only try to determine the so-called threshold function of the property P.

For a property P a function $p_t(n)$ is the threshold function if

$$\lim_{n\to\infty} \mathbb{P}(G(n,p(n)) \in P) = \begin{cases} 1 & \text{if } \frac{p(n)}{p_t(n)} \to \infty, \\ 0 & \text{if } \frac{p(n)}{p_t(n)} \to 0. \end{cases}$$

If the probability of a (sequence of) events converge to 1 then we simply say that the considered event asymptotically almost surely happens. From the definition of the threshold function it is clear that being a threshold function is not a uniquely determined function. For instance, if $p_t(n)$ is a threshold function then for any positive constant c the function $cp_t(n)$ is also a threshold function. Another observation is that the definition suggests that we only consider a threshold function if increasing $p(n)$ also increases $\mathbb{P}(G(n,p(n)) \in P)$. This happens if the property P is monotone increasing, this means that if G has property P then adding edges to G won't lead out from P. For instance, if G has a Hamiltonian-cycle and we add some edges then the obtained graph will have a Hamiltonian-cycle too. If the chromatic number is at least 100, then no matter how many edges we add the chromatic number will be at least 100. But for instance, if we consider the planarity of G then we should study the property that at which p will $G(n,p)$ likely to loose the planarity. This is a monotone decreasing property.

$\star \quad \star \quad \star$

Now let us consider a concrete example: at which p the graph K_4 will appear in $G(n,p)$?

Theorem 5.2.1. *The threshold function of the appearance of K_4 is $n^{-2/3}$.*

Remark 5.2.2. We can rephrase the claim as follows: the threshold function of the property $\omega(G) \geq 4$ is $n^{-2/3}$.

Proof. Let S be a subset of size 4 of $V(G)$, where $G = G(n,p)$ is a random graph. Let A_S be the event that S induces a K_4 in G, and let X_S be the indicator random variable of A_S. Let X be the number of K_4 in G. Then

$$X = \sum_{\substack{S \subseteq V(G) \\ |S|=4}} X_S.$$

Whence

$$\mathbb{E}X = \sum_{\substack{S \subseteq V(G) \\ |S|=4}} \mathbb{E}X_S = \binom{n}{4} p^6 \leq \frac{(pn^{2/3})^6}{24}.$$

If $p(n)n^{2/3} \to_{n\to\infty} 0$ then $\mathbb{E}X \to 0$ as $n \to \infty$. Hence

$$\lim_{n\to\infty} \mathbb{P}(\omega(G) \geq 4) = 0.$$

Now suppose that $p(n)n^{2/3} \to_{n\to\infty} \infty$. Then $\mathbb{E}X \to \infty$ as $n \to \infty$. We will use Theorem 5.1.1; since all set of size 4 looks the same way the random variables X_S are symmetric. Note that $S \sim T$ if $|S \cap T| \geq 2$, otherwise the events A_S and A_T are independent since they don't have a common edge. Let us fix a set S. Then then there are $6\binom{n-2}{2} = O(n^2)$ sets T that intersects S in 2 elements and there are $4\binom{n-3}{1} = O(n)$ sets T intersecting S in 3 vertices. In the former case we have $\mathbb{P}(A_T|A_S) = p^5$, while in the latter case $\mathbb{P}(A_T|A_S) = p^3$. Then

$$\Delta^* = O(n^2p^5) + O(np^3) = o(n^4p^6) = o(\mathbb{E}X)$$

since $p(n)n^{-2/3} \to \infty$. Hence by Theorem 5.1.1 the graph K_4 appears asymptotically almost surely. □

Now let us consider the bit more general problem of determining the threshold function of the appearance of a given graph H. After a quick check of the proof concerning K_4 we see that the value $2/3$ comes from the ratio of the vertices and edges of K_4. This may prompt us to believe that this is also the answer for the general question, i. e., for any graph H the threshold function is $n^{-v(H)/e(H)}$. There is a minor problem with this idea: in order to make sure that H appears one needs that all subgraphs H' also appears, and it might very easily occur that for some H' the value $n^{-v(H')/e(H')}$ is bigger than $n^{-v(H)/e(H)}$. This motivates the following definition.

Definition 5.2.3. Let H be a graph with v vertices and e edges. We call the quantity $\rho(H) = \frac{e}{v}$ the density of H. We say that a graph H is balanced if for all subgraph H' we have $\rho(H') \leq \rho(H)$. The graph H is said to be strictly balanced if for all proper subgraph H' we have $\rho(H') < \rho(H)$.

The proof of the next theorem practically does not require any new idea.

Theorem 5.2.4. *Let H be a balanced graph with n vertices and e edges. Let P_H be the property that H is a (not necessarily induced) subgraph of a graph G. Then the threshold function of P_H is $p = n^{-v/e}$.*

Proof. For all subsets S of size v let A_S be the event that H is a subgraph of $G[S]$. Then

$$p^e \leq \mathbb{P}(A_S) \leq v! p^e.$$

Let X_S be the indicator random variable of A_S. Furthermore, set $X = \sum_{|S|=v} X_S$. Hence the event that G contains H occurs if and only if $X > 0$. By the linearity of expectation we get that

$$\mathbb{E}X = \sum_{|S|=v} \mathbb{E}X_S = \binom{n}{v} \mathbb{P}(A_S) = \Theta(n^v p^e).$$

Hence if $p(n)n^{e/v} \to 0$ then $\mathbb{E}X = o(1)$, thus $X = 0$ asymptotically almost surely.

Now suppose that $p(n)n^{e/v} \to \infty$. Then $\mathbb{E}X \to \infty$. Let us consider Δ^*. (We can do it as the events A_S are symmetric.). If $S \sim T$ then $2 \leq |S \cap T| \leq v - 1$. Then

$$\Delta^* = \sum_{T \sim S} \mathbb{P}(A_T | A_S) = \sum_{i=2}^{v} \sum_{|T \cap S| = i} \mathbb{P}(A_T | A_S).$$

Let i be fixed. Then there are $\binom{v}{i}\binom{n-v}{v-i} = O(n^{v-i})$ ways to choose a set T intersecting S in exactly i vertices. The subgraph induced by $S \cap T$ has i vertices and since H was balanced, the intersection contains at most $i\frac{e}{v}$ edges. So there are at least $e - i\frac{e}{v}$ edges of T not in the intersection with S. Whence

$$\mathbb{P}(A_T | A_S) = O(p^{e - i\frac{e}{v}}).$$

Hence

$$\Delta^* = \sum_{i=2}^{v-1} O(n^{v-i} p^{e - i\frac{e}{v}}) = \sum_{i=2}^{v-1} O((n^v p^e)^{1 - i/v}) = \sum_{i=2}^{v-1} o(n^v p^e) = o(\mathbb{E}X)$$

since $n^v p^e \to \infty$. By Theorem 5.1.1 we get that H appears in G asymptotically almost surely. □

Next we study the isolated vertices and connectedness of $G(n,p)$.

Theorem 5.2.5. *Let $\omega(n) \to \infty$. Furthermore, let $p_\ell(n) = (\log n - \omega(n))/n$ and $p_u(n) = (\log n + \omega(n))/n$. Then $G(n, p_\ell(n))$ contains an isolated vertex asymptotically almost surely while $G(n, p_u(n))$ does not contain isolated vertex vertex asymptotically almost surely.*

Proof. First we prove that $G(n, p_u(n))$ does not contain an isolated vertex asymptotically almost surely. From now on let $p = p_u(n)$. Let X be the number of isolated vertices, and X_v be the indicator random variable of v being an isolated vertex. Then

$$X = \sum_{v \in V} X_v.$$

Observe that $\mathbb{P}(X_v = 1) = (1-p)^{n-1}$. We can assume that $p \leq 1/2$ (why?). Then

$$\mathbb{E}X = \sum_{v \in V} \mathbb{E}X_v = n(1-p)^{n-1} = \frac{1}{1-p} n(1-p)^n \leq 2ne^{-pn} = 2ne^{-\log n + \omega(n)} = 2e^{-\omega(n)} \to 0.$$

as $n \to \infty$. Then

$$\mathbb{P}(X = 0) \geq 1 - \mathbb{E}X \to 1.$$

Next we show that $G(n, p_\ell(n))$ contains an isolated vertex asymptotically almost surely. From now on let $p = p_\ell(n)$. As before, let X be the number of isolated vertices, and X_v be the indicator random variable of v being an isolated vertex. Then $X = \sum_{v \in V} X_v$, and $\mathbb{P}(X_v = 1) = (1-p)^{n-1}$. Hence

$$\mathbb{E}X = \sum_{v \in V} \mathbb{E}X_v = n(1-p)^{n-1} \sim ne^{-\log n + \omega(n)} = e^{\omega(n)} \to \infty.$$

Let us determine $\mathbb{E}X^2$.

$$\mathbb{E}X^2 = \sum_{v \in V} \mathbb{E}X_v^2 + 2 \sum_{u,v \in V} \mathbb{E}X_u X_v = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}.$$

Whence

$$\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} - n^2(1-p)^{2(n-1)} \leq$$

$$\leq n(1-p)^{n-1} + n^2(1-p)^{2n-3} - n^2(1-p)^{2(n-1)} = n(1-p)^{n-1} + pn^2(1-p)^{2n-3} = \mathbb{E}X + \frac{p}{1-p}(\mathbb{E}X)^2$$

Thus

$$\mathbb{P}(X=0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2} \leq \frac{1}{\mathbb{E}X} + \frac{p}{1-p} \to 0$$

since

$$\frac{p}{1-p} \leq 2p \leq \frac{2\log n}{n},$$

if n is large enough. Hence $G(n, p_a(n))$ contains an isolated vertex asymptotically almost surely.

□

Theorem 5.2.6. *Let $\omega(n) \to \infty$. Furthermore, let $p_\ell(n) = (\log n - \omega(n))/n$ and $p_u(n) = (\log n + \omega(n))/n$. Then $G(n, p_\ell(n))$ is disconnected asymptotically almost surely, while $G(n, p_u(n))$ is connected asymptotically almost surely.*

Proof. It is clear from the previous theorem that $G(n, p_\ell(n))$ is disconnected asymptotically almost surely since it contains an isolated vertex with high probability. So we only need to prove that $G(n, p_u(n))$ is connected asymptotically almost surely. This is stronger than what we proved earlier, namely that it does not contain an isolated vertex. From now on let $p = p_u(n)$, and let X_k denote the number of connected components of size k. Furthermore, let

$$X = \sum_{k=1}^{\lfloor n/2 \rfloor} X_k.$$

This is the number of connected components of size at most $\lfloor n/2 \rfloor$. Note that if G is connected, then $X = 0$, and if G is disconnected, then $X \geq 1$ non-negative integer. Hence

$$\mathbb{P}(X=0) \geq 1 - \mathbb{E}X.$$

So we only need to prove that $\mathbb{E}X \to 0$ as $n \to \infty$. Let $f(k,p)$ be the probability that a random graph $G(k,p)$ is connected. For a set S let X_S be indicator random variable that the graph induced by the set S is a connected component of $G(n,p)$. Then

$$\mathbb{E}X_S = \mathbb{P}(X_S = 1) = f(|S|, p)(1-p)^{|S|(n-|S|)}$$

since there must be no edge between S and $V \setminus S$ and the induced subgraph must be connected. Then

$$\mathbb{E}X = \mathbb{E}\left(\sum_{1 \leq |S| \leq \lfloor n/2 \rfloor} X_S \right) = \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} f(k,p)(1-p)^{k(n-k)}.$$

Since $f(k,p) \leq 1$ we have

$$\mathbb{E}X \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)}.$$

We have

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)} \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \left(\frac{en}{k}\right)^k e^{-pk(n-k)}.$$

Here one term can be bounded as follows:

$$\begin{aligned}
\left(\frac{en}{k}\right)^k e^{-pk(n-k)} &= \exp\left(k(1+\log n - \log k) - k(n-k)\frac{\log n + \omega(n)}{n}\right) \\
&= \exp\left(-\omega(n)\frac{k(n-k)}{n}\right) \cdot \exp\left(k\left(1 + \frac{k}{n}\log n - \log k\right)\right) \\
&\leq \exp\left(-\omega(n)\frac{n-1}{n}\right) \cdot \exp\left(k\left(1 + \frac{k}{n}\log n - \log k\right)\right) \\
&\leq \exp\left(-\omega(n)\frac{n-1}{n}\right) e^{-k}.
\end{aligned}$$

if $300 \leq k \leq n/2$, and less than some constant $C \exp\left(-\omega(n)\frac{n-1}{n}\right)$ for $299 \leq k \leq 5$. Indeed, if $x = \frac{k}{n}$ then

$$2 + \frac{k}{n}\log n = 2 + x \log \frac{k}{x} = 2 + x \log \frac{1}{x} + x \log k \leq 2 + \frac{1}{2}\log 2 + \frac{1}{2}\log k \leq \log k$$

for $k \geq 300$. Then

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)} \leq \exp\left(-\omega(n)\frac{n-1}{n}\right) \cdot \left(\sum_{k=1}^{299} C + \sum_{k=300}^{\infty} e^{-k}\right) = C' \exp\left(-\omega(n)\frac{n-1}{n}\right).$$

This last expression goes to 0 as $n \to \infty$. Hence

$$\mathbb{P}(G(n,p) \text{ is not connected}) \to 0.$$

We are done. □

Remark 5.2.7. Another way to estimate the sum

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)}$$

is the following.

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)} \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \left(\frac{en}{k}\right)^k e^{-pk(n-k)} = \sum_{k=1}^{\lfloor n/2 \rfloor} \left(\frac{en}{k} e^{pk} e^{-pn}\right)^k = \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{1-\omega(n)} \frac{e^{pk}}{k}\right)^k.$$

The function e^{px}/x is convex on the interval $[0, \infty)$ for arbitrary p. In particular, it takes its maximum at one of the end points on the interval $[1, n/2]$. At 1 this function is at most e. At $n/2$ we can assume that $\omega(n) \leq \log n$ and we get that the value of the function is at most 2. So on the whole interval it is at most e. Hence

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{1-\omega(n)} \frac{e^{pk}}{k} \right)^k \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{2-\omega(n)} \right)^k \leq \frac{e^{2-\omega(n)}}{1 - e^{2-\omega(n)}} \to 0.$$

6. Lovász Local Lemma

In the previous chapter about the second moment method we have seen that some crude information about the dependence of the random variables can help a lot. In combinatorial problems it almost never happens that we have completely independent random variables, but in many cases it turns out that each random variable is independent of all others, but a small number of exceptions. This turns out to be almost as good as complete independence. At least we can prove statements where certain probabilities are exponentially small, but luckily they are positive. In this chapter we will study such a tool, this is the so called Lovász local lemma.

6.1 Lovász local lemma

Theorem 6.1.1. *(Lovász local lemma, general version) Let $B_1, \dots, B_n$ be events in an arbitrary probability space. The directed graph $D = (V, E)$ with vertex set $V = \{1, 2, \dots, n\}$ is the dependence graph of the events $B_1, \dots, B_n$ if for all $1 \leq i \leq n$ the event B_i is mutually independent from the events $\{B_j \mid (i, j) \notin E\}$. Assume that the directed graph $D = (V, E)$ is the dependence graph of the events $B_1, \dots, B_n$ and there exist real numbers $x_1, \dots, x_n$ satisfying $0 \leq x_i < 1$ and*

$$\mathbb{P}(B_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j)$$

for all $1 \leq i \leq n$. Then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{B_i}\right) \geq \prod_{i=1}^{n} (1 - x_i) > 0.$$

Remark 6.1.2. We can think of the events $B_1, \dots, B_n$ as bad events. We would like to avoid all of them, so we need that with positive probability none of them occurs, that is,

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{B_i}\right) > 0.$$

Proof. Let $S \subseteq \{1, 2, \ldots, n\}$, $|S| = s$. First we show by induction on s that for arbitrary $i \notin S$ we have

$$\mathbb{P}\left(B_i \;\middle|\; \bigcap_{j\in S} \overline{B_j}\right) \leq x_i.$$

This is trivial for $s = 0$. Assume that we have proved the statement for all $s' < s$, we show that the statement is also true for s. Let $S_1 = \{j \in S \mid (i,j) \in E\}$ and $S_2 = \{j \in S \mid (i,j) \notin E\}$. Then

$$\mathbb{P}\left(B_i \;\middle|\; \bigcap_{j\in S} \overline{B_j}\right) = \frac{\mathbb{P}\left(B_i \cap \left(\bigcap_{j\in S_1} \overline{B_j}\right) \mid \bigcap_{t\in S_2} \overline{B_t}\right)}{\mathbb{P}\left(\bigcap_{j\in S_1} \overline{B_j} \mid \bigcap_{t\in S_2} \overline{B_t}\right)}.$$

First we estimate the enumerator of the fraction using that B_i is mutually independent from the events $\{B_t \mid t \in S_2\}$.

$$\mathbb{P}\left(B_i \cap \left(\bigcap_{j\in S_1} \overline{B_j}\right) \;\middle|\; \bigcap_{t\in S_2} \overline{B_t}\right) \leq \mathbb{P}\left(B_i \;\middle|\; \bigcap_{t\in S_2} \overline{B_t}\right) = \mathbb{P}(B_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j.)$$

To estimate the denominator we use the induction hypothesis. Let $S_1 = \{j_1, \ldots, j_r\}$. If $r = 0$ then the denominator is 1 and the statement immediately follows. If $r \geq 1$ then we can use the inductive hypothesis

$$\mathbb{P}\left(\overline{B_{j_1}} \cap \overline{B_{j_2}} \cap \cdots \cap \overline{B_{j_r}} \;\middle|\; \bigcap_{t\in S_2} \overline{B_t}\right) =$$

$$= \left(1 - \mathbb{P}\left(B_{j_1} \;\middle|\; \bigcap_{t\in S_2} \overline{B_t}\right)\right) \cdot \left(1 - \mathbb{P}\left(B_{j_2} \;\middle|\; \overline{B_{j_1}} \cap \bigcap_{t\in S_2} \overline{B_t}\right)\right) \cdots$$

$$\left(1 - \mathbb{P}\left(B_{j_r} \;\middle|\; \overline{B_{j_1}} \cap \cdots \cap \overline{B_{j_{r-1}}} \cap \left(\bigcap_{t\in S_2} \overline{B_t}\right)\right)\right) \geq$$

$$\geq (1 - x_{j_1})(1 - x_{j_2}) \ldots (1 - x_{j_r}).$$

Putting together the estimates of the enumerator and the denominator we get the required statement. From this the statement of the theorem immediately follows:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{B_i}\right) = (1 - \mathbb{P}(B_1)) \left(1 - \mathbb{P}\left(B_2 \mid \overline{B_1}\right)\right) \cdots \left(1 - \mathbb{P}\left(B_n \;\middle|\; \bigcap_{i=1}^{n-1} \overline{B_i}\right)\right) \geq \prod_{i=1}^{n} (1 - x_i).$$

Hence we have proved the theorem. □

Theorem 6.1.3. *(Lovász local lemma, symmetric form) Let $B_1, \ldots, B_n$ be events in an arbitrary probability space. Assume that for all i the event B_i is mutually independent from all other, but at most d events. Assume that $\mathbb{P}(B_i) \leq p$ and*

$$p \leq \frac{1}{e(d+1)},$$

where $e = 2,71...$ is the base of the natural logarithm. Then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{B_i}\right) > 0.$$

Proof. If $d = 0$ then the statement is trivial. If $d \geq 1$ we can apply the statement of the previous theorem for the dependence graph of the events $B_1, \ldots, B_n$, where for all i we have $|\{j \mid (i,j) \in E\}| \leq d$. Let $x_i = \frac{1}{d+1} < 1$. Since

$$\left(1 - \frac{1}{d+1}\right)^d \geq \frac{1}{e}$$

we have

$$\mathbb{P}(B_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j).$$

Therefore,

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{B_i}\right) \geq \prod_{i=1}^{n} (1 - x_i) > 0.$$

□

6.2 2-colorings of hypergraphs

Theorem 6.2.1. *Suppose that all edge of the hypergraph $H = (V, E)$ has at least k vertices and all of them intersects at most d others. If $e(d+1) \leq 2^{k-1}$, then it is possible to color the vertices of the hypergraph with two colours without resulting monochromatic edge.*

Proof. Let us color each vertex with blue and red with probability $1/2-1/2$ independently of each other. Let B_f be the event that the edge f is monochromatic. Clearly, $\mathbb{P}(B_f) = 2/2^k = 1/2^{k-1}$. For all $f \in E(H)$ the event B_f is mutually independent of all events $B_{f'}$ if f' does not intersect f. Hence all event is mutually independent of all others, but d events. Since $\mathbb{P}(B_f) = \frac{1}{2^{k-1}} \leq \frac{1}{e(d+1)}$ we have $\mathbb{P}(\bigcap_{i=1}^{n} \overline{B_i}) > 0$ by the Lovász local lemma. Hence with positive probability there exists a 2-coloring without monochromatic edge. □

6.3 Ramsey-numbers revisited

Theorem 6.3.1. *(a) If*

$$e\left(\binom{k}{2}\binom{n-2}{k-2}+1\right)2^{1-\binom{k}{2}}<1$$

then we have $R(k,k)>n$.

(b) We have

$$R(k,k)>\frac{\sqrt{2}}{e}(1+o(1))k2^{k/2}.$$

Proof. Let us colour each edge of K_n to red or blue with probability $1/2-1/2$. Let $S\subset\{1,2,\dots,n\}$ for which $|S|=k$, and let B_S be the event that each edge induced by the set S get the same colour. Then

$$\mathbb{P}(B_S)=\frac{2}{2^{\binom{k}{2}}}=\frac{1}{2^{\binom{k}{2}-1}}.$$

The event B_S is mutually independent from all event $B_{S'}$ for which the sets S and S' have at most one common vertex. Thus the degree of B_S in the dependence digraph is at most

$$\binom{k}{2}\binom{n-2}{k-2}.$$

(Note that if $k\geq 4$ then we overcounted by counting those sets S' more times that have intersection of size at least 3, but this is not a problem as we only need an upper bound.) According to the condition of the theorem we have

$$\mathbb{P}(B_S)\leq\frac{1}{e(d+1)}.$$

Hence, by the Lovász local lemma we have

$$\mathbb{P}\left(\bigcap_{|S|=k}\overline{B_S}\right)>0.$$

In other words, with positive probability there exists a coloring such that there is no monochromatic subset of size k. Hence $R(k,k)>n$.

The analysis in part (b) is left to the Reader. □

6.4 Cycles in directed graphs

Theorem 6.4.1 (Alon and Linial [3]). *Let $D = (V, E)$ be a simple directed graph in which the minimal out-degree is δ and the maximal in-degree is Δ. If*

$$e(\Delta(\delta + 1) + 1)\left(1 - \frac{1}{k}\right)^{\delta} < 1,$$

then D contains a directed cycle whose length is divisible by k.

Proof. We can assume that all out-degree are δ, otherwise we consider a subgraph of D. Let $f : V \to \{0, 1, \ldots, k-1\}$ be a random coloring of the vertices where we choose each $f(v)$ independently from each other. For all $v \in V$ let B_v be the event that there exists no $u \in V$ such that $(v, u) \in E$ and $f(u) \equiv f(v) + 1 \pmod{k}$. Then $\mathbb{P}(B_v) = (1 - 1/k)^{\delta}$. Note that B_v is mutually independent from all events, but those events B_u for which

$$N^+(v) \cap (\{u\} \cup N^+(u)) \neq \emptyset,$$

where $N^+(u) = \{w \in V |\ (u, w) \in E\}$. The number of such events is at most $\Delta(\delta+1)$. By the condition $e(\Delta(\delta+1)+1)(1-\frac{1}{k})^{\delta} < 1$ we can apply the Lovász local lemma. Hence $\mathbb{P}(\cap_{v \in V} \overline{B_v}) > 0$. Therefore there exists an $f : V \to \{0, 1, \ldots, k-1\}$ coloring such that for all $v \in V$ there exists a $u \in V$ such that $(v, u) \in E$ and $f(u) \equiv f(v) + 1 \pmod{k}$.

Now let us pick a vertex v and start a walk from v by always stepping to a new vertex that has 1-bigger f-value modulo k then the previous vertex. At some point we have to arrive to an already visited vertex. Then we have found a cycle whose length is divisible by k. □

7. Correlation Inequalities

In politics it is a common wisdom that if you cannot stop something then stand to the lead of it. Mathematicians would say that if you can't prevent something then use it to your own purposes. In the previous chapters we have seen how we can conquer more and more of independence. But, alas, dependence happens. On the other hand, one might secretly hope that the arising correlation can be used efficiently. In this chapter we will see how to prove correlation inequalities and use them.

7.1 Positive correlation

Definition 7.1.1. For $\underline{x}, \underline{y} \in \{0,1\}^n$ let $\underline{x} \vee \underline{y}$ be the vector for which $(\underline{x} \vee \underline{y})_i = \max(x_i, y_i)$, and let $\underline{x} \wedge \underline{y}$ be the vector for which $(\underline{x} \wedge \underline{y})_i = \min(x_i, y_i)$.

Theorem 7.1.2 (Ahlswede and Daykin [1])**.** *Let $f_1, f_2, f_3, f_4 : \{0,1\}^n \to \mathbb{R}_+$ satisfying the inequality*

$$f_1(\underline{x}) f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y}) f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. Let

$$F_i = \sum_{\underline{x} \in \{0,1\}^n} f_i(\underline{x})$$

for $i = 1, 2, 3, 4$. Then

$$F_1 \cdot F_2 \leq F_3 \cdot F_4.$$

Proof. We prove the statement by induction on n. For $n = 1$ the condition of the theorem gives that

$$f_1(0) f_2(0) \leq f_3(0) f_4(0).$$

$$f_1(0) f_2(1) \leq f_3(1) f_4(0).$$

$$f_1(1) f_2(0) \leq f_3(1) f_4(0).$$

$$f_1(1)f_2(1) \leq f_3(1)f_4(1).$$

We need to prove that

$$(f_1(0) + f_1(1))(f_2(0) + f_2(1)) \leq (f_3(0) + f_3(1))(f_4(0) + f_4(1)).$$

If $f_3(1) = 0$ or $f_4(0) = 0$ then $f_3(1)f_4(0) \leq f_3(0)f_4(1)$ and the claim is trivially true:

$$(f_1(0)+f_1(1))(f_2(0)+f_2(1)) \leq f_3(0)f_4(0)+2f_3(1)f_4(0)+f_3(1)f_4(1) \leq (f_3(0)+f_3(1))(f_4(0)+f_4(1)).$$

So we can assume that $f_3(1) \neq 0$ and $f_4(0) \neq 0$. Then

$$(f_3(0) + f_3(1))(f_4(0) + f_4(1)) \geq \left(\frac{f_1(0)f_2(0)}{f_4(0)} + f_3(1)\right)\left(f_4(0) + \frac{f_1(1)f_2(1)}{f_3(1)}\right).$$

So it would be enough to prove that

$$\left(\frac{f_1(0)f_2(0)}{f_4(0)} + f_3(1)\right)\left(f_4(0) + \frac{f_1(1)f_2(1)}{f_3(1)}\right) \geq (f_1(0) + f_1(1))(f_2(0) + f_2(1)).$$

This is equivalent with

$$(f_1(0)f_2(0)+f_3(1)f_4(0))(f_3(1)f_4(0)+f_1(1)f_2(1)) \geq f_3(1)f_4(0)(f_1(0)+f_1(1))(f_2(0)+f_2(1)).$$

This is in turn equivalent with

$$(f_3(1)f_4(0) - f_1(0)f_2(1))(f_3(1)f_4(0) - f_1(1)f_2(0)) \geq 0$$

which is true by the assumptions of the theorem. This proves the case $n = 1$.

Now suppose that the claim is true till $n - 1$ and we wish to prove it for n. Set $f_i'(\underline{x}) : \{0,1\}^{n-1} \to \mathbb{R}_+$ for $i = 1, 2, 3, 4$ as follows:

$$f_i'(\underline{x}) = f_i(\underline{x}, 0) + f_i(\underline{x}, 1).$$

First we show that f_i' satisfies the inequality

$$f_1'(\underline{x})f_2'(\underline{y}) \leq f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^{n-1}$. This is of course true: for a fixed $\underline{x}, \underline{y} \in \{0,1\}^{n-1}$ let us apply the case $n = 1$ to the functions

$$g_1(u) = f_1(\underline{x}, u) \;\; g_2(u) = f_2(\underline{y}, u) \;\; g_3(u) = f_3(\underline{x} \vee \underline{y}, u) \;\; g_3(u) = f_4(\underline{x} \wedge \underline{y}, u),$$

where $u \in \{0,1\}$. Then the functions g_i satisfy

$$g_1(u_1)g_2(u_2) \leq g_3(u_1 \vee u_2)g_4(u_1 \wedge u_2)$$

for all $u_1, u_2 \in \{0,1\}$ by the assumption on f. By the case $n = 1$ we know that

$$(g_1(0) + g_1(1))(g_2(0) + g_2(1)) \leq (g_3(0) + g_3(1))(g_4(0) + g_4(1)).$$

In other words,

$$f_1'(\underline{x})f_2'(\underline{y}) \leq f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^{n-1}$. Then by induction we get that for $F_i' = \sum_{\underline{x} \in \{0,1\}^{n-1}} f_i(\underline{x})$ we have

$$F_1' \cdot F_2' \leq F_3' \cdot F_4'.$$

But of course $F_i' = F_i$ whence

$$F_1 \cdot F_2 \leq F_3 \cdot F_4.$$

□

Theorem 7.1.3. *Let $f_1, f_2, f_3, f_4 : \{0,1\}^n \to \mathbb{R}_+$ satisfying the inequality*

$$f_1(\underline{x})f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y})f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. Let $f_1', f_2', f_3', f_4' : \{0,1\}^k \to \mathbb{R}_+$ be defined by

$$f_i'(\underline{x}) = \sum_{\underline{u} \in \{0,1\}^{n-k}} f_i(\underline{x}, \underline{u}).$$

Then for all $\underline{x}, \underline{y} \in \{0,1\}^k$ we have

$$f_1'(\underline{x})f_2'(\underline{y}) \leq f_3'(\underline{x} \vee \underline{y})f_4'(\underline{x} \wedge \underline{y})$$

Proof. This immediately follows from Theorem 7.1.2. For fixed $\underline{x}, \underline{y} \in \{0,1\}^k$ define $g_1, g_2, g_3, g_4 : \{0,1\}^{n-k} \to \mathbb{R}_+$

$$g_1(\underline{u}) = f_1(\underline{x}, \underline{u}), \quad g_2(\underline{u}) = f_2(\underline{y}, \underline{u}), \quad g_3(\underline{u}) = f_3(\underline{x} \vee \underline{y}, \underline{u}), \quad g_4(\underline{u}) = f_4(\underline{x} \wedge \underline{y}, \underline{u}).$$

Then for any $\underline{u}, \underline{v} \in \{0,1\}^{n-k}$ we have

$$g_1(\underline{u})g_2(\underline{v}) \leq g_3(\underline{u} \vee \underline{v})g_4(\underline{u} \wedge \underline{v})$$

by the assumption on the functions f_1, f_2, f_3, f_4. Then for

$$f_i'(\underline{x}) = G_i = \sum_{\underline{u} \in \{0,1\}^{n-k}} g_i(\underline{u}) = \sum_{\underline{u} \in \{0,1\}^{n-k}} f_i(\underline{x}, \underline{u})$$

we have

$$f_1'(\underline{x}) f_2'(\underline{y}) = G_1 G_2 \leq G_3 G_4 = f_3'(\underline{x} \vee \underline{y}) f_4'(\underline{x} \wedge \underline{y}).$$

□

Definition 7.1.4. For $\underline{x}, \underline{y} \in \{0,1\}^n$ we say that $\underline{x} \geq \underline{y}$ if for all $i \in [n]$ we have $x_i \geq y_i$.

A function $f : \{0,1\}^n \to \mathbb{R}^+$ is monotone increasing if $f(\underline{x}) \geq f(\underline{y})$ for all $\underline{x} \geq \underline{y}$ and it is monotone decreasing if $f(\underline{x}) \leq f(\underline{y})$ for all $\underline{x} \geq \underline{y}$.

In general, for a poset (or lattice) L a function $f : L \to \mathbb{R}^+$ is monotone increasing if $f(x) \geq f(x)$ for all $x \geq_L y$ and it is monotone decreasing if $f(x) \leq f(y)$ for all $x \geq_L y$.

Theorem 7.1.5 (Fortuin, Kasteleyn, Ginibre [12])**.** *A function $\mu : \{0,1\}^n \to \mathbb{R}^+$ is log-supermodular if*

$$\mu(\underline{x})\mu(\underline{y}) \leq \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. Then for a log-supermodular $\mu : \{0,1\}^n \to \mathbb{R}^+$ and monotone increasing (decreasing) functions $f, g : \{0,1\}^n \to \mathbb{R}^+$ we have

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) \right) \left(\sum_{x \in \{0,1\}^n} \mu(\underline{x}) g(\underline{x}) \right) \leq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) g(\underline{x}) \right) \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) \right).$$

Furthermore, if $f : \{0,1\}^n \to \mathbb{R}^+$ is monotone increasing and $g : \{0,1\}^n \to \mathbb{R}^+$ is monotone decreasing then

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) \right) \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) g(\underline{x}) \right) \geq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) g(\underline{x}) \right) \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) \right).$$

Proof. First suppose that both f and g are monotone increasing. Let us apply Theorem 7.1.2 for the following theorems:

$$f_1(\underline{x}) = \mu(\underline{x}) f(\underline{x}), \quad f_2(\underline{x}) = \mu(\underline{x}) g(\underline{x}), \quad f_3(\underline{x}) = \mu(\underline{x}) f(\underline{x}) g(\underline{x}), \quad f_4(\underline{x}) = \mu(\underline{x}).$$

We need to check that

$$f_1(\underline{x}) f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y}) f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. This is indeed true:

$$\begin{aligned} f_1(\underline{x}) f_2(\underline{y}) &= \mu(\underline{x}) f(\underline{x}) \mu(\underline{y}) g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y}) \mu(\underline{x} \wedge \underline{y}) f(\underline{x}) g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y}) \mu(\underline{x} \wedge \underline{y}) f(\underline{x} \vee \underline{y}) g(\underline{x} \vee \underline{y}) \\ &= f_3(\underline{x} \vee \underline{y}) f_4(\underline{x} \wedge \underline{y}). \end{aligned}$$

In the first inequality we used the log-supermodularity of μ, and in the second inequality we used that both f and g are monotone increasing. Then by Theorem 7.1.2 we have $F_1 \cdot F_2 \leq F_3 \leq F_4$, i. e.,

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) \right) \left(\sum_{x \in \{0,1\}^n} \mu(\underline{x}) g(\underline{x}) \right) \leq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) g(\underline{x}) \right) \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) \right).$$

If f and g are both monotone decreasing then set

$$f_1(\underline{x}) = \mu(\underline{x}) f(\underline{x}), \quad f_2(\underline{x}) = \mu(\underline{x}) g(\underline{x}), \quad f_3(\underline{x}) = \mu(\underline{x}), \quad f_4(\underline{x}) = \mu(\underline{x}) f(\underline{x}) g(\underline{x}).$$

Again we need to check that

$$f_1(\underline{x}) f_2(\underline{y}) \leq f_3(\underline{x} \vee \underline{y}) f_4(\underline{x} \wedge \underline{y})$$

for all $\underline{x}, \underline{y} \in \{0,1\}^n$. This is indeed true:

$$\begin{aligned} f_1(\underline{x}) f_2(\underline{y}) &= \mu(\underline{x}) f(\underline{x}) \mu(\underline{y}) g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y}) \mu(\underline{x} \wedge \underline{y}) f(\underline{x}) g(\underline{y}) \\ &\leq \mu(\underline{x} \vee \underline{y}) \mu(\underline{x} \wedge \underline{y}) f(\underline{x} \wedge \underline{y}) g(\underline{x} \wedge \underline{y}) \\ &= f_3(\underline{x} \vee \underline{y}) f_4(\underline{x} \wedge \underline{y}). \end{aligned}$$

From this we can conclude again that

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) \right) \left(\sum_{x \in \{0,1\}^n} \mu(\underline{x}) g(\underline{x}) \right) \leq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) g(\underline{x}) \right) \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) \right).$$

If f is monotone increasing, and g is monotone decreasing then let $M = \max_{\underline{x} \in \{0,1\}^n} g(\underline{x})$, and consider the function $g'(\underline{x}) = M - g(\underline{x})$. Then $g'(\underline{x}) \geq 0$ and monotone increasing. Whence

$$\left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) \right) \left(\sum_{x \in \{0,1\}^n} \mu(\underline{x}) g'(\underline{x}) \right) \leq \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) f(\underline{x}) g'(\underline{x}) \right) \left(\sum_{\underline{x} \in \{0,1\}^n} \mu(\underline{x}) \right).$$

By writing the definition of $g(\underline{x}) = M - g'(\underline{x})$ into it we get that

$$\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{x\in\{0,1\}^n} \mu(\underline{x})(M-g(\underline{x}))\right) \leq \left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})(M-g(\underline{x}))\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})\right).$$

After subtracting $M(\sum \mu(\underline{x}))(\sum \mu(\underline{x})f(\underline{x}))$ and multiplying with -1 we get that

$$\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})g(\underline{x})\right) \geq \left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})f(\underline{x})g(\underline{x})\right)\left(\sum_{\underline{x}\in\{0,1\}^n} \mu(\underline{x})\right).$$

□

Remark 7.1.6. It is worth considering μ as a measure, and suppose that it is a log-supermodular probability measure. Then the above theorem says that

$$\mathbb{E}_\mu f \mathbb{E}_\mu g \leq \mathbb{E}_\mu fg$$

for monotone increasing functions f and g.

Theorem 7.1.7. *Let L be a distributive lattice. A function $\mu : L \to \mathbb{R}^+$ is log-supermodular if*

$$\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y)$$

for all $x, y \in L$. For a log-supermodular $\mu : L \to \mathbb{R}^+$ and monotone increasing (decreasing) functions $f, g : L \to \mathbb{R}^+$ we have

$$\left(\sum_{x\in L} \mu(x)f(x)\right)\left(\sum_{x\in L} \mu(x)g(x)\right) \leq \left(\sum_{x\in L} \mu(x)f(x)g(x)\right)\left(\sum_{x\in L} \mu(x)\right).$$

Furthermore, if $f : L \to \mathbb{R}^+$ is monotone increasing and $g : L \to \mathbb{R}^+$ is monotone decreasing then

$$\left(\sum_{x\in L} \mu(x)f(x)\right)\left(\sum_{x\in L} \mu(x)g(x)\right) \geq \left(\sum_{x\in L} \mu(x)f(x)g(x)\right)\left(\sum_{x\in L} \mu(x)\right).$$

Proof. This theorem follows from Theorem 7.1.5 since every distributive lattice L is a sublattice of some $\{0,1\}^n$. So all we need to do is to define μ on $\{0,1\}^n \setminus L$ to be 0, and to extend f and g in a monotone increasing way. (This last step is only needed formally since $\mu(\underline{x})f(\underline{x}), \mu(\underline{x})g(\underline{x}), \mu(\underline{x})f(\underline{x})g(\underline{x})$ are all 0 anyway for $\underline{x} \in \{0,1\}^n \setminus L$.) The extended μ will remain log-supermodular since $\mu(x)\mu(y) \neq 0$ then $x, y \in L$ and then $x \vee y, x \wedge y \in L$ so $\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y)$, and if $\mu(x)\mu(y) = 0$ then the inequality holds true trivially. □

In the next few results we give examples of various log-supermodular measures.

Theorem 7.1.8. *Assume that the function $\mu : \{0,1\}^n \to \mathbb{R}_+$ is log-supermodular. Then the function $\mu' : \{0,1\}^k \to \mathbb{R}_+$ defined by*

$$\mu'(\underline{x}) = \sum_{\underline{u} \in \{0,1\}^{n-k}} \mu(\underline{x}, \underline{u})$$

is also log-supermodular.

Proof. This theorem is an immediate application of Theorem 7.1.3 applied to $f_1 = f_2 = f_3 = f_4 = \mu$.

□

Theorem 7.1.9. *For probabilities $p_1, \ldots, p_n$ let*

$$\mathbb{P}_p(A) = \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j).$$

Let $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$ be monotone increasing, and $\mathcal{C}, \mathcal{D} \subseteq 2^{[n]}$ be monotone decreasing set families. For a set family $\mathcal{S}$ set

$$\mathbb{P}_p(\mathcal{S}) = \sum_{S \in \mathcal{S}} \mathbb{P}_p(S).$$

Then we have

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geq \mathbb{P}_p(\mathcal{A}) \cdot \mathbb{P}_p(\mathcal{B}),$$

$$\mathbb{P}_p(\mathcal{C} \cap \mathcal{D}) \geq \mathbb{P}_p(\mathcal{C}) \cdot \mathbb{P}_p(\mathcal{D}),$$

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{C}) \leq \mathbb{P}_p(\mathcal{A}) \cdot \mathbb{P}_p(\mathcal{C}).$$

Proof. We can associate the characteristic vector $\underline{1}_A \in \{0,1\}^n$ with a set A. Let

$$\mu(\underline{x}) = \prod_{i=1}^{n} p_i^{x_i} (1 - p_i)^{1 - x_i}.$$

Then $\mathbb{P}_p(A) = \mu(\underline{1}_A)$. Then

$$\mu(\underline{x})\mu(\underline{y}) = \mu(\underline{x} \vee \underline{y})\mu(\underline{x} \wedge \underline{y})$$

or equivalently $\mathbb{P}_p(A)\mathbb{P}_p(B) = \mathbb{P}_p(A \cup B)\mathbb{P}_p(A \cap B)$. Furthermore, let f be the characteristic functions of the family of sets $\mathcal{A}$, i. e., $f(1_A) = 1$ if $A \in \mathcal{A}$ and 0

otherwise. Similarly, let g be the characteristic functions of the family of sets $\mathcal{B}$. Then f and g are monotone increasing functions. The inequality

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geq \mathbb{P}_p(\mathcal{A}) \cdot \mathbb{P}_p(\mathcal{B})$$

is simply Theorem 7.1.5 applied to μ, f and g. The other parts of the theorem follow similarly.

□

Theorem 7.1.10. *Given a graph G on the vertex set $[n]$, and a parameter $\beta > 0$ with a vector $\underline{B} = (B_1, \ldots, B_n)$. For an $\underline{x} = (x_1, \ldots, x_n) \in \{-1, 1\}^n$ set*

$$\mathbb{P}_{\beta,\underline{B}}(\underline{x}) = \frac{1}{Z} \exp\left(\beta \sum_{(i,j)\in E(G)} x_i x_j + \sum_{i=1}^{n} B_i x_i\right),$$

where Z is some normalizing constant. For vectors $\underline{x}$ and $\underline{y}$ let $(\underline{x} \wedge \underline{y})_i = \min(x_i, y_i)$ and $(\underline{x} \vee \underline{y})_i = \max(x_i, y_i)$. Then we have

$$\mathbb{P}_{\beta,\underline{B}}(\underline{x}) \cdot \mathbb{P}_{\beta,\underline{B}}(\underline{y}) \leq \mathbb{P}_{\beta,\underline{B}}(\underline{x} \wedge \underline{y}) \cdot \mathbb{P}_{\beta,\underline{B}}(\underline{x} \vee \underline{y}).$$

Proof. Clearly, the statement is equivalent with

$$\left(\beta \sum_{(i,j)\in E(G)} x_i x_j + \sum_{i=1}^{n} B_i x_i\right) + \left(\beta \sum_{(i,j)\in E(G)} y_i y_j + \sum_{i=1}^{n} B_i y_i\right) \leq$$

$$\leq \left(\beta \sum_{(i,j)\in E(G)} \min(x_i, y_i)\min(x_j, y_j) + \sum_{i=1}^{n} B_i \min(x_i, y_i)\right)$$

$$+ \left(\beta \sum_{(i,j)\in E(G)} \max(x_i, y_i)\max(x_j, y_j) + \sum_{i=1}^{n} B_i \max(x_i, y_i)\right).$$

It is enough to prove the statement term by term. Note that $x_i + y_i = \min(x_i, y_i) + \max(x_i, y_i)$ so we only need to prove that

$$x_i x_j + y_i y_j \leq \min(x_i, y_i)\min(x_j, y_j) + \max(x_i, y_i)\max(x_j, y_j).$$

If $x_i \leq y_i$ and $x_j \leq y_j$ then this statement holds true with equality. Of course, the same is true if $x_i \geq y_i$ and $x_j \geq y_j$. If $x_i \leq y_i$ and $x_j \geq y_j$ then we need to prove that

$$x_i x_j + y_i y_j \leq x_i y_j + x_j y_i$$

which is equivalent with

$$(x_i - y_i)(y_j - x_j) \geq 0$$

which is true by assumption. Again the same is true if $x_i \geq y_i$ and $x_j \leq y_j$. □

Theorem 7.1.11. *Let $G = (A, B, E)$ be a bipartite graph, and let $\lambda \geq 0$. Let $\mathcal{I}(G)$ be the set of independent sets of G. Let $\boldsymbol{I}$ be a random independent set of G such that for an independent set I of G we have*

$$\mathbb{P}_\lambda(\boldsymbol{I} = I) = \frac{\lambda^{|I|}}{I(G, \lambda)}$$

where

$$I(G, \lambda) = \sum_I \lambda^{|I|}.$$

Then for all $u, v \in A$ we have

$$\mathbb{P}_\lambda[u, v \in \boldsymbol{I}] \geq \mathbb{P}_\lambda[u \in \boldsymbol{I}]\mathbb{P}_\lambda[v \in \boldsymbol{I}],$$

and for $u \in A$ and $v \in B$ we have

$$\mathbb{P}_\lambda[u, v \in \boldsymbol{I}] \leq \mathbb{P}_\lambda[u \in \boldsymbol{I}]\mathbb{P}_\lambda[v \in \boldsymbol{I}].$$

Proof. Consider the following function $\mu : \{0, 1\}^{A \cup B}$:

$$\mu(\underline{x}, \underline{y}) = \exp\left(\ln(\lambda)\left(\sum_{u \in A} x_u + \sum_{v \in B}(1 - y_v)\right)\right) \prod_{(u,v) \in E(G)} (1 - x_u(1 - y_v)).$$

First we show the connection between μ and $\mathbb{P}_\lambda$. For $(\underline{x}, \underline{y}) \in \{0, 1\}^{A \cup B}$ set

$$S = \{u \in A \mid x_u = 1\} \cup \{v \in B \mid y_v = 0\}.$$

Note that if S is not an independent set then there exists a $(u, v) \in E(G)$ such that $x_u = 1, y_v = 0$. Then $1 - x_u(1 - y_v) = 0$ and $\mu(\underline{x}, \underline{y}) = 0$. If S is an independent set for all $(u, v) \in E(G)$ we have $1 - x_u(1 - y_v) = 1$, and

$$\exp\left(\ln(\lambda)\left(\sum_{u \in A} x_u + \sum_{v \in B}(1 - y_v)\right)\right) = \lambda^{|S|}.$$

So up to the normalization constant $I(G, \lambda)$ the function μ and $\mathbb{P}_\lambda$ are the same. Next we show that $\mu(\underline{x}, \underline{y})$ is log-supermodular. It is clear that if $(\underline{x}, \underline{y})$ and $(\underline{x}', \underline{y}')$ are two vectors then

$$\left(\sum_{u \in A} x_u + \sum_{v \in B}(1 - y_v)\right) + \left(\sum_{u \in A} x'_u + \sum_{v \in B}(1 - y'_v)\right) =$$

$$= \left(\sum_{u\in A} \min(x_u, x'_u) + \sum_{v\in B}(1 - \min(y_v, y'_v))\right) + \left(\sum_{u\in A} \max(x_u, x'_u) + \sum_{v\in B}(1 - \max(y_v, y'_v))\right).$$

So we only need to prove that

$$(1-x_u(1-y_v))(1-x'_u(1-y'_v)) \leq (1-\min(x_u, x'_u)(1-\min(y_v, y'_v))(1-\max(x_u, x'_u)(1-\max(y_v, y'_v)).$$

One can do it by checking 16 cases, but it is possible to speed up the checking by some observations. The right hand side is non-negative so we only need to exclude the cases where the left hand side is 1 (and the right hand side is 0). If $x_u = x'_u = 1$ then $y_v = y'_v = 0$ and then the right hand side is 1. If $x_u = x'_u = 0$ then the right hand side is again 1. Similarly, if $y_v = y'_v = 0$ then $x_u = x'_u = 0$, or $y_v = y'_v = 1$, then right hand side is again 1. So we only need to check when one of x_u and x'_u is 1, the other 0, and one of y_v and y'_v is 1, the other 0. By symmetry we can assume that $x_u = 0, x'_u = 1$: if $y_v = 0, y'_v = 1$ then both sides is 1, and if $y_v = 1, y'_v = 0$ then the left hand side is 0, but the right hand side is still 1. Hence the inequality is indeed true.

Finally, fix a u and v as in the statement in the theorem. Now we can apply Theorem 7.1.5 in the first case to the functions

$$\mu(\underline{x}, \underline{y}), \quad f(\underline{x}, \underline{y}) = x_u, \quad g(\underline{x}, \underline{y}) = x_v.$$

Clearly,

$$\sum_{(\underline{x},\underline{y})} \mu(\underline{x}, \underline{y}) x_u x_v = I(G, \lambda)\mathbb{P}_\lambda[u, v \in \mathbf{I}].$$

In the second case we can apply Theorem 7.1.5 to the functions.

$$\mu(\underline{x}, \underline{y}), \quad f(\underline{x}, \underline{y}) = x_u, \quad g(\underline{x}, \underline{y}) = 1 - y_v.$$

In the first case, both f and g are monotone increasing, in the second case f is monotone increasing and g is monotone decreasing. After dividing by $I(G, \lambda)^2$ we get that for all $u, v \in A$ we have

$$\mathbb{P}_\lambda[u, v \in \mathbf{I}] \geq \mathbb{P}_\lambda[u \in \mathbf{I}]\mathbb{P}_\lambda[v \in \mathbf{I}],$$

and for a $u \in A$ and $v \in B$ we have

$$\mathbb{P}_\lambda[u, v \in \mathbf{I}] \leq \mathbb{P}_\lambda[u \in \mathbf{I}]\mathbb{P}_\lambda[v \in \mathbf{I}].$$

□

8. Poisson paradigm

In the chapter about the Lovász local lemma we have seen that we can give a lower bound for the probability $\mathbb{P}\left(\bigcap_{i\in I}\overline{B_i}\right)$ if the bad events B_i have sufficiently small probabilities, and all bad events are mutually independent of every other events, but a small number of exceptions. If the bad events were independent then we would expect that the number X of occuring bad events is close to a Poisson distribution. One might hope that this is true in more generality, that is, if we allow some small dependence. This is the content of this chapter. In the first section we study the so-called Janson's inequalities that –at least for a special, but important setup– provide such a result: $\mathbb{P}(X = 0)$ is close to $e^{-\mu}$, the probability of a Poisson distribution with parameter μ taking value 0. Then we study another tool to prove that certain distributions converge to a Poisson distribution.

8.1 Janson's inequalities

Setup. Let Ω be a fixed set and let R be a random subset of Ω by choosing $r \in R$ with probability p_r mutually independently of each other. Let $(A_i)_{i\in I}$ be subsets of Ω for some index set I. Let B_i be the event that $A_i \subseteq R$. Let X_i be the indicator random variable for the event B_i. Set

$$X = \sum_{i\in I} X_i.$$

It is, of course, the number of $A_i \subseteq R$. So the events $\cap_{i\in I}\overline{B_i}$ and $X = 0$ are identical. For $i, j \in I$ we say that $i \sim j$ if $A_i \cap A_j \neq \emptyset$, not that if $i \not\sim j$ then this is consistent with our previous notation that B_i and B_j are independent. Let

$$\Delta = \sum_{i\sim j} \mathbb{P}\left(B_i \cap B_j\right),$$

where the sum is over all ordered pairs, so $\Delta/2$ is the same sum for unordered pairs. This notation is again consistent with the notation introduced in the chapter on the

second moment method. Set

$$M = \prod_{i \in I} \mathbb{P}\left(\overline{B_i}\right).$$

This would be the probability of $\cap_{i \in I} \overline{B_i}$ if the events B_i were independent. Finally, set

$$\mu = \mathbb{E}X = \sum_{i \in I} \mathbb{P}(B_i).$$

Now we are ready to state Janson's inequalities.

Theorem 8.1.1 (Janson inequality [14])**.** *Let $(B_i)_{i \in I}, M, \Delta, \mu$ be as above, and assume that $\mathbb{P}(B_i) \leq \varepsilon$ for all $i \in I$. Then*

$$M \leq \mathbb{P}\left(\bigcap_{i \in I} \overline{B_i}\right) \leq M \exp\left(\frac{1}{1-\varepsilon} \cdot \frac{\Delta}{2}\right).$$

Furthermore,

$$\mathbb{P}\left(\bigcap_{i \in I} \overline{B_i}\right) \leq \exp\left(-\mu + \frac{\Delta}{2}\right).$$

Theorem 8.1.2 (Extended Janson inequality)**.** *Let $(B_i)_{i \in I}, M, \Delta, \mu$ be as above, and further assume that $\Delta \geq \mu$. Then*

$$\mathbb{P}\left(\bigcap_{i \in I} \overline{B_i}\right) \leq e^{-\mu^2/2\Delta}.$$

Remark 8.1.3. Note that

$$M = \prod_{i \in I} \mathbb{P}\left(\overline{B_i}\right) = \prod_{i \in I} (1 - \mathbb{P}(B_i)) \leq \prod_{i \in I} e^{-\mathbb{P}(B_i)} = e^{-\mu}.$$

This shows that two upper bounds in Theorem 8.1.1 are not really different in the important case when $\varepsilon \to 0$ as the hidden parameter $n \to \infty$. In fact, $M \sim e^{-\mu}$ in the case when $\varepsilon = o(1)$, $m\varepsilon^2 = o(1)$ and $\Delta = o(1)$. To see this let us introduce the function $\kappa(x)$ for which

$$1 - x = e^{-x + \kappa(x)}$$

for $x \in [0, 1)$. Then $\kappa(x) \leq 0$, monotone decreasing function, and for $x \leq 1/2$ we have $\kappa(x) \geq -x^2$. Hence

$$M = \prod_{i \in I} \mathbb{P}\left(\overline{B_i}\right) = \prod_{i \in I} (1 - \mathbb{P}(B_i)) = \prod_{i \in I} e^{-\mathbb{P}(B_i) + \kappa(\mathbb{P}(B_i))} = e^{-\mu} \exp\left(\sum_{i \in I} \kappa(\mathbb{P}(B_i))\right).$$

Here

$$\left|\sum_{i\in I}\kappa(\mathbb{P}(B_i))\right| \le m\kappa(\varepsilon) \le m\varepsilon^2$$

if $\varepsilon \le 1/2$. Since $\Delta = o(1)$ we get that both the upper and lower bound is $e^{-\mu}(1 + o(1))$, that is,

$$\mathbb{P}\left(\bigcap_{i\in I}\overline{B_i}\right) = e^{-\mu}(1 + o(1))$$

in this case.

Proof of Theorem 8.1.1. First let us prove that

$$\mathbb{P}\left(B_i \;\middle|\; \bigcap_{j\in J}\overline{B_j}\right) \le \mathbb{P}(B_i)$$

for $J \subset I$ with $i \notin J$. Using the definition of the conditional probability this is equivalent with

$$\mathbb{P}\left(B_i \cap \left(\bigcap_{j\in J}\overline{B_j}\right)\right) \le \mathbb{P}(B_i)\mathbb{P}\left(\bigcap_{j\in J}\overline{B_j}\right).$$

Observe that B_i is an increasing event and $\cap_{j\in J}\overline{B_j}$ is a decreasing event, so we can use the FKG correlation inequality to obtain the above inequality.

It is also true that

$$\mathbb{P}\left(B_i \;\middle|\; B_k \cap \left(\bigcap_{j\in J}\overline{B_j}\right)\right) \le \mathbb{P}(B_i|B_k)$$

for $J \subset I$ with $i, k \notin J$. This inequality follows from the first one since conditioning on B_k is simply equivalent with assuming that $p_r = \mathbb{P}(r \in R) = 1$ for all $r \in A_k$.

Next we prove the lower bound. Let $I = \{1, 2, \ldots, m\}$. Since

$$\mathbb{P}\left(B_i \;\middle|\; \bigcap_{1\le j<i}\overline{B_j}\right) \le \mathbb{P}(B_i)$$

we have

$$\mathbb{P}\left(\overline{B_i} \;\middle|\; \bigcap_{1\le j<i}\overline{B_j}\right) \ge \mathbb{P}(\overline{B_i}).$$

Hence

$$\mathbb{P}\left(\bigcap_{i\in I}\overline{B_i}\right) = \prod_{i=1}^{m}\mathbb{P}\left(\overline{B_i} \;\middle|\; \bigcap_{1\le j<i}\overline{B_j}\right) \ge \prod_{i=1}^{m}\mathbb{P}(\overline{B_i}).$$

We are done.

Next we prove the upper bound. First observe that for any event A, B, C we have

$$\mathbb{P}(A|B \cap C) \geq \mathbb{P}(A \cap B|C)$$

by using the definition of the conditional probability and that $\mathbb{P}(C) \geq P(B \cap C)$ trivially holds true. For a fixed $i \in I$ let us apply it as follows:

$$A = B_i \quad B = \bigcap_{\substack{j \sim i \\ j<i}} \overline{B_j} \quad C = \bigcap_{\substack{j \not\sim i \\ j<i}} \overline{B_j}.$$

Then

$$\mathbb{P}\left(B_i \;\middle|\; \bigcap_{1 \leq j<i} \overline{B_j}\right) = \mathbb{P}(A|B \cap C) \geq \mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|A \cap C) = \mathbb{P}(A)\mathbb{P}(B|A \cap C)$$

In the above computation, in the second step we can use the definition of the conditional probability, and in the third step we used that B_i is independent of C. Next we bound $\mathbb{P}(B|A \cap C)$. We have

$$\mathbb{P}(B|A \cap C) \geq 1 - \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j|A \cap C) = 1 - \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j|B_i \cap C) \geq 1 - \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j|B_i),$$

where in the last step we used the correlation inequality established at the beginning of the proof. Putting together the last two inequalities we get that

$$\mathbb{P}\left(B_i \;\middle|\; \bigcap_{1 \leq j<i} \overline{B_j}\right) \geq \mathbb{P}(B_i) - \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i).$$

Taking the complement event we get that

$$\mathbb{P}\left(\overline{B_i} \;\middle|\; \bigcap_{1 \leq j<i} \overline{B_j}\right) \leq \mathbb{P}(\overline{B_i}) + \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i).$$

Now we have only some algebraic manipulations left to do. Since $\mathbb{P}(\overline{B_i}) \geq 1 - \varepsilon$ we get that

$$\mathbb{P}(\overline{B_i}) + \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i) \leq \mathbb{P}(\overline{B_i}) \left(1 + \frac{1}{1-\varepsilon} \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i)\right).$$

Furthermore, since $1 + x \leq e^x$ we get that

$$\mathbb{P}(\overline{B_i}) \left(1 + \frac{1}{1-\varepsilon} \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i) \right) \leq \mathbb{P}(\overline{B_i}) \exp\left(\frac{1}{1-\varepsilon} \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i) \right).$$

Hence

$$\mathbb{P}\left(\overline{B_i} \,\middle|\, \bigcap_{1 \leq j < i} \overline{B_j} \right) \leq \mathbb{P}(\overline{B_i}) \exp\left(\frac{1}{1-\varepsilon} \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i) \right).$$

Multiplying together these inequalities for all $1 \leq i \leq m$ we get that

$$\mathbb{P}\left(\bigcap_{i \in I} \overline{B_i} \right) \leq M \exp\left(\frac{1}{1-\varepsilon} \cdot \frac{\Delta}{2} \right).$$

To prove the second upper bound we use the inequalities:

$$\mathbb{P}\left(\overline{B_i} \,\middle|\, \bigcap_{1 \leq j < i} \overline{B_j} \right) \leq 1 - \mathbb{P}(B_i) + \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i) \leq \exp\left(-\mathbb{P}(B_i) + \sum_{\substack{j \sim i \\ j<i}} \mathbb{P}(B_j \cap B_i) \right),$$

and by multiplying together these inequalities for all $1 \leq i \leq m$ we get that

$$\mathbb{P}\left(\bigcap_{i \in I} \overline{B_i} \right) \leq \exp\left(-\mu + \frac{\Delta}{2} \right).$$

□

Next we prove the extended Janson's inequality.

Proof of Theorem 8.1.2. The second upper bound of Theorem 8.1.1 can be rewritten as follows:

$$-\ln\left(\mathbb{P}\left(\bigcap_{i \in I} \overline{B_i} \right) \right) \geq \sum_{i \in I} \mathbb{P}(B_i) - \frac{1}{2} \sum_{i \sim j} \mathbb{P}(B_i \cap B_j).$$

Of course, this is true for any $S \subseteq I$:

$$-\ln\left(\mathbb{P}\left(\bigcap_{i \in S} \overline{B_i} \right) \right) \geq \sum_{i \in S} \mathbb{P}(B_i) - \frac{1}{2} \sum_{\substack{i \sim j \\ i,j \in S}} \mathbb{P}(B_i \cap B_j).$$

Now let us apply it to a random subset S by choosing an $i \in S$ with probability p, where p is chosen later. Then

$$\mathbb{E}\left(-\ln\left(\mathbb{P}\left(\bigcap_{i \in S} \overline{B_i} \right) \right) \right) \geq \mathbb{E}\left(\sum_{i \in S} \mathbb{P}(B_i) \right) - \frac{1}{2} \mathbb{E}\left(\sum_{\substack{i \sim j \\ i,j \in S}} \mathbb{P}(B_i \cap B_j) \right) = p\mu - p^2 \frac{\Delta}{2}.$$

Let $p = \frac{\mu}{\Delta}$. This is at most 1 by the condition of the theorem. Hence

$$\mathbb{E}\left(-\ln\left(\mathbb{P}\left(\bigcap_{i\in S}\overline{B_i}\right)\right)\right) \geq p\mu - p^2\frac{\Delta}{2} = \frac{\mu^2}{2\Delta}.$$

So there must be a set S for which

$$-\ln\left(\mathbb{P}\left(\bigcap_{i\in S}\overline{B_i}\right)\right) \geq \frac{\mu^2}{2\Delta}.$$

Then

$$\mathbb{P}\left(\bigcap_{i\in I}\overline{B_i}\right) \leq \mathbb{P}\left(\bigcap_{i\in S}\overline{B_i}\right) \leq e^{-\mu^2/2\Delta}.$$

We are done! □

8.2 Brun's sieve

Setup. Let $B_1, \ldots, B_m$ be events with X_i indicator random variables. Set $X = X_1 + \cdots + X_m$. As usual there is a hidden parameter n: $B_i = B_i(n)$ and $m = m(n)$. Set

$$S^{(r)} = \sum_{i_1<i_2<\cdots<i_r} \mathbb{P}(B_{i_1} \cap \cdots \cap B_{i_r}).$$

Let

$$(X)_r = X(X-1)\ldots(X-r+1),$$

and

$$\binom{X}{r} = \frac{1}{r!}(X)_r.$$

The next theorem is the well-known inclusion-exclusion principle and the Bonferroni's inequalities.

Theorem 8.2.1. *(a) We have*

$$\mathbb{P}(X=0) = \mathbb{P}\left(\bigcap_{i=1}^{m}\overline{B_i}\right) = \sum_{r=0}^{m}(-1)^r S^{(r)}.$$

(b) For every s we have

$$\sum_{r=0}^{2s-1}(-1)^r S^{(r)} \leq \mathbb{P}\left(\bigcap_{i=1}^{m}\overline{B_i}\right) = \sum_{r=0}^{2s}(-1)^r S^{(r)}.$$

(c) In general, for every s and k we have

$$\sum_{r=0}^{2s-1}(-1)^r\binom{k+r}{k}S^{(k+r)} \le \mathbb{P}(X=k) \le \sum_{r=0}^{2s}(-1)^r\binom{k+r}{k}S^{(k+r)}.$$

Proof. Clearly, it is enough to prove part (c) since part (a) and (b) are special cases of part (c).

The sets $B_1,\dots,B_m$ partition the ground set into 2^m events called atoms. A set $B_{i_1}\cap\cdots\cap B_{i_r}$ can be decomposed into 2^{m-r} such atoms. When we decompose each $S^{(k+r)}$ into 2^{m-k-r} sums of probabilities of atoms, we get that

$$\sum_{r=0}^{2s-1}(-1)^r\binom{k+r}{k}S^{(k+r)} = \sum_E c(E)\mathbb{P}(E),$$

where the sum on the right hand side runs over atoms, and $c(E)$ is some constant. Suppose that the atom E is the intersection of t sets B_i and $m-t$ sets $\overline{B_j}$. Then $\mathbb{P}(E)$ will appear in $\binom{t}{k+r}$ expansion of terms $S^{(k+r)}$, whence

$$c(E) = \sum_{r=0}^{2s-1}(-1)^r\binom{k+r}{k}\binom{t}{k+r} = \binom{t}{k}\sum_{r=0}^{2s-1}(-1)^r\binom{t-k}{r} = -\binom{t}{k}\binom{t-k-1}{2s-1}$$

if $t>k$ and $c(E)=1$ if $t=k$. This shows that

$$\sum_{r=0}^{2s-1}(-1)^r\binom{k+r}{k}S^{(k+r)} \le \mathbb{P}(X=k)$$

since $\mathbb{P}(X=k)$ is the sum of the probabilities of those atoms that are the intersection of k sets B_i and $m-k$ sets $\overline{B_j}$. Very similarly,

$$\sum_{r=0}^{2s-1}(-1)^r\binom{k+r}{k}S^{(k+r)} = \sum_E c'(E)\mathbb{P}(E),$$

where

$$c'(E) = \sum_{r=0}^{2s}(-1)^r\binom{k+r}{k}\binom{t}{k+r} = \binom{t}{k}\sum_{r=0}^{2s-1}(-1)^r\binom{t-k}{r} = \binom{t}{k}\binom{t-k-1}{2s}$$

if $t>k$ and $c'(E)=1$ if $t=k$. □

Theorem 8.2.2. *Suppose that for some constant μ we have that for every fixed r*

$$\mathbb{E}\binom{X}{r} = S^{(r)} \to \frac{\mu^r}{r!}.$$

Then for every fixed k

$$\mathbb{P}(X = k) \to \frac{\mu^k}{k!} e^{-\mu}.$$

In particular, $\mathbb{P}(X = 0) \to e^{-\mu}$.

Proof. Let k be fixed. Note that

$$\sum_{r=0}^{\infty} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{(k+r)!} = \frac{\mu^k}{k!} e^{-\mu}.$$

For a fixed ε let s be chosen such that

$$\max\left(\left| \sum_{r=0}^{2s-1} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{(k+r)!} - \frac{\mu^k}{k!} e^{-\mu} \right|, \left| \sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{(k+r)!} - \frac{\mu^k}{k!} e^{-\mu} \right| \right) \leq \frac{\varepsilon}{2}.$$

Let n_0 be large enough such that for every $n \geq n_0$ and $0 \leq r \leq 2s$ we have

$$\binom{k+r}{k} \left| S^{(k+r)} - \frac{\mu^r}{r!} \right| \leq \frac{\varepsilon}{2(2s+1)}.$$

Then

$$\begin{aligned}
\mathbb{P}(X = k) - \frac{\mu^k}{k!} e^{-\mu} &= \left(\mathbb{P}(X = k) - \sum_{r=0}^{2s-1} (-1)^r \binom{k+r}{k} S^{(k+r)} \right) \\
&\quad + \left(\sum_{r=0}^{2s-1} (-1)^r \binom{k+r}{k} \left(S^{(k+r)} - \frac{\mu^{k+r}}{(k+r)!} \right) \right) \\
&\quad + \left(\sum_{r=0}^{2s-1} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{(k+r)!} - \frac{\mu^k}{k!} e^{-\mu} \right) \\
&\geq 0 - 2s \frac{\varepsilon}{2(2s+1)} - \frac{\varepsilon}{2} \geq -\varepsilon
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(X = k) - \frac{\mu^k}{k!} e^{-\mu} &= \left(\mathbb{P}(X = k) - \sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} S^{(k+r)} \right) \\
&\quad + \left(\sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} \left(S^{(k+r)} - \frac{\mu^{k+r}}{(k+r)!} \right) \right) \\
&\quad + \left(\sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{(k+r)!} - \frac{\mu^k}{k!} e^{-\mu} \right) \\
&\leq 0 + (2s+1) \frac{\varepsilon}{2(2s+1)} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

Hence $\left| \mathbb{P}(X = k) - \frac{\mu^k}{k!} e^{-\mu} \right| \leq \varepsilon$. Since ε was arbitrary we get that $P(X = k) \to \frac{\mu^k}{k!} e^{-\mu}$. □

Without proof let us mention the multivariate version of this claim.

Theorem 8.2.3. *Let $\lambda_1 = \lambda_1(n), \ldots, \lambda_m = \lambda_m(n)$ be non-negative bounded functions. For each n let $X_1(n), \ldots, X_m(n)$ be non-negative integer valued function in the same space. Suppose for all $r_1, \ldots, r_m \in \mathbb{Z}_+$ we have*

$$\lim_{n\to\infty} (\mathbb{E}((X_1)_{r_1}(X_2)_{r_2}\ldots(X_m)_{r_m}) - \lambda_1^{r_1}\ldots\lambda_m^{r_m}) = 0.$$

Then $X_1(n), \ldots, X_m(n)$ are asymptotically independent Poisson random variables with means $\lambda_1, \ldots, \lambda_m$, that is

$$\lim_{n\to\infty} \left(\mathbb{P}(X_1 = k_1, \ldots X_n = k_n) - \prod_{i=1}^{m} \left(e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!} \right) \right) = 0$$

for all $k_1, \ldots, k_m \in \mathbb{Z}_+$.

8.3 Vertices and triangles

In this section we give an application of the above theorems.

Let $G \sim G(n,p)$ and let EPIT represent the statement that every vertex lies in a triangle.

Theorem 8.3.1. *Let $c > 0$ be fixed and let $p = p(n)$, $\mu = \mu(n)$ satisfy that*

$$e^{-\mu} = \frac{c}{n} \quad \textit{and} \quad \binom{n-1}{2} p^3 = \mu.$$

Then

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \textit{ satisfies } EPIT) = e^{-c}.$$

Proof. For vertices x, y, z let B_{xyz} be the event that xyz form a triangle in G. Let $C_x = \bigcap_{yz} \overline{B_{xyz}}$ be the event that there is no triangle on x, and let X_x be the corresponding indicator random variable. Finally, set

$$X = \sum_{x\in V(G)} X_x.$$

We will show that X has an asymptotic Poisson distribution with parameter c. According to Theorem 8.2.2 we need to show that for every fixed r

$$\mathbb{E}\binom{X}{r} = S^{(r)} \to \frac{c^r}{r!}.$$

Consider one term in the sum

$$S^{(r)} = \sum_{1 \le i_1 < \cdots < i_r \le n} \mathbb{P}(C_{x_{i_1}} \cap \cdots \cap C_{x_{i_r}}),$$

that is, $\mathbb{P}(C_{x_{i_1}} \cap \cdots \cap C_{x_{i_r}})$. Clearly,

$$C_{x_{i_1}} \cap \cdots \cap C_{x_{i_r}} = \bigcap_{i=1}^{r} \bigcap_{y,z} \overline{B_{x_i yz}}.$$

To bound this probability we use Janson's inequalities. We have $\mathbb{P}(B_{xyz}) = p^3$ which can be assumed to be less than $\varepsilon = 1/8$ as it is $o(1)$. The number of terms is

$$m_r(n) = r\binom{n-1}{2} - \binom{r}{2}(n-2) + \binom{r}{3} = r\binom{n-1}{2} + O(n)$$

as r is fixed. Finally, $\Delta = O(n^3 p^5)$ since the events $B_{x_i y_i z_i}$ and $B_{x_j y_j z_j}$ are not independent if they share an edge which means that they have altogether 4 vertices from which at least one (as maybe $x_i = x_j$) is from the fixed r vertices, and we can choose at most 3 vertices freely. Hence by Janson's inequality

$$S^{(r)} = \binom{n}{r}(1-p^3)^{m_r(n)}\eta_{n,r},$$

where

$$1 \le \eta_{n,r} \le \exp\left(\frac{1}{1-\varepsilon} \cdot \frac{\Delta}{2}\right).$$

Note that

$$p = \left(\frac{2\ln(n/c)}{n(n-1)}\right)^{1/3}$$

which means that $p = o(n^{-3/5})$, consequently $\Delta = O(n^3 p^5) = o(1)$. This means that

$$\exp\left(\frac{1}{1-\varepsilon} \cdot \frac{\Delta}{2}\right) = 1 + o(1).$$

Then

$$\binom{n}{r}(1-p^3)^{m_r(n)}\eta_{n,r} \sim \binom{n}{r}e^{-p^3 m_r(n)} \sim \binom{n}{r}e^{-p^3 r\binom{n-1}{2}} = \binom{n}{r}e^{-r\mu} = \binom{n}{r}\left(\frac{c}{n}\right)^r \sim \frac{c^r}{r!}.$$

(Note that we could have used Remark 8.1.3 too.) Hence by Theorem 8.2.2 we have

$$\mathbb{P}(G(n,p) \text{ satisfies } EPIT) = P(X = 0) \to e^{-c}.$$

□

9. Martingales

In this chapter we study martingales. It turns out that martingales are especially amenable to provide tight concentration inequalities in combinatorial problems. Martingales are tied to various exposure processes that arise very naturally in these problems. Rather than fighting for independence, martingales seem to take an alternative route by grasping the phenomenon that if a random variable can change only a small amount by a small perturbation then it is strongly concentrated around its mean. This introduction will be more clear after some examples and applications.

9.1 Martingales

Definition 9.1.1. A sequence of random variables $(X_n)_n$ is called a martingale if for all n we have

$$\mathbb{E}(X_{n+1}|X_n, X_{n-1}, \ldots, X_0) = X_n.$$

Example 9.1.2. Edge exposure martingale. Let $G \sim G(n,p)$, and f any graph theoretic function. For a fixed ordering $e_1, \ldots, e_{\binom{n}{2}}$ of the pair of vertices define the random variables $X_1, \ldots, X_{\binom{n}{2}}$ as follows:

$$X_i(H) = \mathbb{E}(f(G) \mid \text{for } j \leq i,\ e_j \in E(G) \Leftrightarrow e_j \in E(H)).$$

This means that we expose the first i pairs (whether they are in G or not), and then we consider the conditional expectation of f based on this observation.

Example 9.1.3. Vertex exposure martingale. Again let $G \sim G(n,p)$, and f any graph theoretic function. For a fixed ordering of the vertices define the random variables $X_1, \ldots, X_n$ as follows:

$$X_i(H) = \mathbb{E}(f(G) \mid \text{for } x, y \leq i,\ (x,y) \in E(G) \Leftrightarrow (x,y) \in E(H)).$$

This means that we expose the first i vertices, and then we consider the conditional expectation of f based on this observation.

Theorem 9.1.4 (Azuma's inequality)**.** *Let $0 = X_0, \ldots, X_m$ be a martingale with $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i \leq m$. Let $\lambda > 0$ be arbitrary. Then*

$$\mathbb{P}(X_m > \lambda\sqrt{m}) < e^{-\lambda^2/2}.$$

Proof. Set $\alpha = \frac{\lambda}{\sqrt{m}}$, and define $Y_i = X_i - X_{i-1}$. Then $|Y_i| \leq 1$, and $\mathbb{E}(Y_i|X_{i-1}, \ldots, X_0) = 0$. Furthermore, for $x \in [-1, 1]$ let

$$h(x) = \frac{e^\alpha + e^{-\alpha}}{2} + \frac{e^\alpha - e^{-\alpha}}{2}x.$$

Then $e^{\alpha x} \leq h(x)$. Hence –by some abuse of notation– we have

$$\begin{aligned}
\mathbb{E}(e^{\alpha Y_i}|X_{i-1}, \ldots, X_0)(t_{i-1}, \ldots, t_0) &= \sum_{s_i} e^{\alpha s_i}\mathbb{P}(Y_i = s_i|X_0 = t_0, \ldots, X_{i-1} = t_{i-1}) \\
&\leq \sum_{s_i} h(s_i)\mathbb{P}(Y_i = s_i|X_0 = t_0, \ldots, X_{i-1} = t_{i-1}) \\
&= \sum_{s_i} \left(\frac{e^\alpha + e^{-\alpha}}{2} + \frac{e^\alpha - e^{-\alpha}}{2}s_i\right)\mathbb{P}(Y_i = s_i|X_0 = t_0, \ldots, X_{i-1} = t_{i-1}) \\
&= \frac{e^\alpha + e^{-\alpha}}{2} + \frac{e^\alpha - e^{-\alpha}}{2}\mathbb{E}(Y_i|X_{i-1}, \ldots, X_0)(t_{i-1}, \ldots, t_0) \\
&= \frac{e^\alpha + e^{-\alpha}}{2}.
\end{aligned}$$

It is easy to prove that

$$\frac{e^\alpha + e^{-\alpha}}{2} \leq e^{\alpha^2/2}.$$

Then

$$\begin{aligned}
\mathbb{E}(e^{\alpha X_m}) = \mathbb{E}\left(\prod_{i=1}^{m} e^{\alpha Y_i}\right) &= \mathbb{E}\left(\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right)\mathbb{E}(e^{\alpha Y_m}|X_{m-1}, \ldots, X_0)\right) \\
&\leq \mathbb{E}\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right)e^{\alpha^2/2} = \mathbb{E}(e^{\alpha X_{m-1}})e^{\alpha^2/2}
\end{aligned}$$

Then by induction we get that $\mathbb{E}(e^{\alpha X_m}) \leq e^{\alpha^2 m/2}$. Hence

$$\mathbb{P}(X_m > \lambda\sqrt{m}) = \mathbb{P}(e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}) < \mathbb{E}(e^{\alpha X_m})e^{-\alpha\lambda\sqrt{m}} \leq e^{\alpha^2 m/2}e^{-\alpha\lambda\sqrt{m}} = e^{-\lambda^2/2}.$$

We are done!

□

Theorem 9.1.5. *Let $c = X_0, \ldots, X_m$ be a martingale with $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i \leq m$. Let $\lambda > 0$ be arbitrary. Then*

$$\mathbb{P}(|X_m - c| > \lambda\sqrt{m}) < 2e^{-\lambda^2/2}.$$

Definition 9.1.6. A graph theoretic function f is said to satisfy the edge Lipshitz condition if whenever the graphs H and H' differ only in an edge, then $|f(H) - f(H')| \leq 1$. It satisfies the vertex Lipshitz condition if whenever the graphs H and H' differ only in a vertex, then $|f(H) - f(H')| \leq 1$.

Theorem 9.1.7. *(a) When f satisfies the edge Lipschitz condition the corresponding edge exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$.*

(b) When f satisfies the vertex Lipschitz condition the corresponding vertex exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$.

We will prove this intuitively clear statement a bit later. See the next section.

Theorem 9.1.8 (Shamir and Spencer [15])**.** *Let n, p be arbitrary and let $c = \mathbb{E}\chi(G)$, where $G \sim G(n,p)$. Then*

$$\mathbb{P}(|\chi(G) - c| > \lambda\sqrt{n-1}) < 2e^{-\lambda^2/2}.$$

Proof. Consider the vertex exposure martingale. Since $|\chi(H) - \chi(H')| \leq 1$ the vertex Lipshitz condition is satisfied, and by the previous theorem we have $|X_{i+1} - X_i| \leq 1$. Then the claim follows from Azuma's inequality. □

Theorem 9.1.9. *Let $p = n^{-\alpha}$, where $\alpha > \frac{5}{6}$ fixed. Let $G = G(n,p)$. Then there exists a $u = u(n,p)$ such that almost always $u \leq \chi(G) \leq u + 3$. In other words, $\chi(G)$ is concentrated on 4 values.*

Lemma 9.1.10. *Let α, c be fixed, $\alpha > \frac{5}{6}$. Let $p = n^{-\alpha}$. Then for almost every $G \sim G(n,p)$ every $c\sqrt{n}$ vertices of $G = G(n,p)$ can be three-colored.*

Proof. Let us bound the probability that for some graph G, where $G \sim G(n,p)$ there exists an $A \subseteq V(G)$ with $|A| = c\sqrt{n}$ the graph $G[A]$ cannot be three-colored. Then A has to contain some minimal subset T that $G[T]$ is not 3-colorable. Then each degree in $G[T]$ has to be at least 3, otherwise for a vertex x with degree at most 2

the graph $G[T-x]$ is 3-colorable by minimality, but then $G[T]$ is also 3-colorable. Hence

$$\sum_{\substack{G\\ \exists A:\chi(G[A])\geq 4}} \mathbb{P}(G) \leq \sum_{\substack{G\\ \exists T:e(G[T])\geq \frac{3}{2}|T|,|T|\leq c\sqrt{n}}} \mathbb{P}(G) \leq \sum_{\substack{T\\ e(G[T])\geq \frac{3}{2}|T|,|T|\leq c\sqrt{n}}} \mathbb{P}(T).$$

Furthermore,

$$\sum_{\substack{T\\ e(G[T])\geq \frac{3}{2}|T|,|T|\leq c\sqrt{n}}} \mathbb{P}(T) \leq \sum_{t=4}^{c\sqrt{n}} \binom{n}{t}\binom{\binom{t}{2}}{3t/2} p^{3t/2}.$$

Then we use the bounds

$$\binom{n}{t} \leq \left(\frac{ne}{t}\right)^t \quad \text{and} \quad \binom{\binom{t}{2}}{3t/2} \leq \left(\frac{te}{3}\right)^{3t/2}.$$

Then we can bound a term in the above sum as

$$\left(\frac{ne}{t}\frac{t^{3/2}e^{3/2}}{3^{3/2}}n^{-3\alpha/2}\right)^t \leq \left(c_1 n^{1-3\alpha/2}t^{1/2}\right)^t \leq \left(c_2 n^{1-3\alpha/2}n^{1/4}\right)^t = (c_2 n^{-\kappa})^t,$$

where $\kappa = \frac{3}{2}\alpha - \frac{5}{4} > 0$. And so the above sum is $o(1)$. □

Proof of Theorem 9.1.9. Let $\varepsilon > 0$ be fixed. Set λ such that $\varepsilon = e^{-\lambda^2/2}$. Let us say that a graph G is good if every subset of size at most $2\lambda\sqrt{n-1}$ is 3-colorable. By the previous lemma if n is large enough and $G \sim G(n,p)$, then G is good with probability at least $1-\varepsilon$. Define $u = u(n,p,\varepsilon)$ as the least value for which $\mathbb{P}(\chi(G) \leq u) > \varepsilon$. By definition of u we have $\mathbb{P}(\chi(G) \leq u-1) \leq \varepsilon$. We will show that

$$\mathbb{P}(u \leq \chi(G) \leq u+3) \geq 1-3\varepsilon.$$

Let $Y = Y(G)$ be the following random variable:

$$Y(G) = \min_{\chi(G-S)\leq u} |S|.$$

Then

$$\mathbb{P}(Y=0) = \mathbb{P}(\chi(G) \leq u) > \varepsilon.$$

Set $\mathbb{E}Y = \mu$. Note that Y satisfies the vertex Lipshitz-condition since if G and G' differ in only one vertex then the size of the minimal sets S and S' can differ at most 1. Hence by Azuma's inequality

$$\mathbb{P}(Y \leq \mu - \lambda\sqrt{n-1}) < e^{-\lambda^2/2} = \varepsilon,$$

and

$$\mathbb{P}(Y \geq \mu + \lambda\sqrt{n-1}) < e^{-\lambda^2/2} = \varepsilon.$$

Note that $\mu \leq \lambda\sqrt{n-1}$, otherwise

$$\varepsilon \leq \mathbb{P}(Y = 0) \leq \mathbb{P}(Y \leq \mu - \lambda\sqrt{n-1}) < \varepsilon$$

would lead to a contradiction. Then

$$\mathbb{P}(Y \geq 2\lambda\sqrt{n-1}) \leq \mathbb{P}(Y \geq \mu + \lambda\sqrt{n-1}) \leq \varepsilon.$$

With probability at least $1 - 3\varepsilon$ the following things are satisfied: (1) $\chi(G) \geq u$, (2) G is good, (3) $Y(G) \leq 2\lambda\sqrt{n-1}$. Then there exists a set S of size at most $2\lambda\sqrt{n-1}$ such that $G - S$ is colorable with at most u colors, and we can color S with extra 3 colors since G is good. Hence $\chi(G) \leq u + 3$. Hence

$$\mathbb{P}(u \leq \chi(G) \leq u + 3) \geq 1 - 3\varepsilon.$$

□

9.2 Lipschitz condition

In this section we prove Theorem 9.1.7. It will be convenient to prove a slightly more general statement.

Setup. Let $\Omega = A^B$ be the set of functions $g : B \to A$. Consider the measure on Ω by setting

$$\mathbb{P}(g(b) = a) = p_{ab},$$

where the values are assumed to be mutually independent. Fix a gradation

$$\emptyset = B_0 \subset B_1 \subset B_2 \subset \cdots \subset B_m = B.$$

Let $L : \Omega \to \mathbb{R}$ be a functional, and define the martingale $X_0, \ldots, X_m$ by setting

$$X_i(h) = \mathbb{E}(L(g) \mid g(b) = h(b) \text{ for all } b \in B_i).$$

Hence $X_0 = \mathbb{E}L(g)$, and $X_m = L$. We say that the values $X_0, X_1, \ldots, X_m$ satisfies the Lipschitz condition with respect to the gradation if for all $0 \leq i \leq m$ the following holds true: if h, h' differ only on $B_{i+1} \setminus B_i$ then $|L(h') - L(h)| \leq 1$.

Example 9.2.1. When $B = \binom{[n]}{2}$ and $A = \{0, 1\}$, and $p_{1b} = p,\ p_{0b} = 1 - p$ we get back the $G(n, p)$ model. If B_i contains the first i pairs, then we get back the edge exposure martingale. If B_i contains all pairs induced by the first i vertices then we get back the vertex exposure martingale.

Theorem 9.2.2. *Let L satisfy the Lipschitz-condition. Then the corresponding martingale satisfies $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i \leq m - 1$ and $h \in \Omega$.*

Proof. Let H be the family of h' that agree with h on B_{i+1}. Then

$$X_{i+1}(h) = \sum_{h' \in H} L(h')\mathbb{P}(g = h' | g = h \text{ on } B_{i+1}).$$

For each h' in H let $H[h']$ be family of h^* that agree with h' on all points except on $B_{i+1} \setminus B_i$. Then the sets $H[h']$ is a partition of h^* agreeing on h with B_i. Hence

$$X_i(h) = \sum_{h' \in H} \sum_{h^* \in H[h']} L(h^*)\mathbb{P}(g = h^* | g = h^* \text{ on } B_{i+1})\mathbb{P}(g = h^* \text{ on } B_{i+1} | g = h^* \text{ on } B_i).$$

Note that

$$\mathbb{P}(g = h^* | g = h^* \text{ on } B_{i+1}) = \mathbb{P}(g = h' | g = h' \text{ on } B_{i+1})$$

for $h^* \in H[h']$, namely they are both equal to $\prod_{b \in B \setminus B_{i+1}} p_{h'(b),b} = \prod_{b \in B \setminus B_{i+1}} p_{h^*(b),b}$. Since $h' = h$ on B_{i+1} we can further write it as:

$$\mathbb{P}(g = h^* | g = h^* \text{ on } B_{i+1}) = \mathbb{P}(g = h' | g = h \text{ on } B_{i+1}).$$

For the ease of notation, set

$$w_{h'} = \mathbb{P}(g = h' | g = h \text{ on } B_{i+1}) \quad \text{and} \quad q_{h^*} = \mathbb{P}(g = h^* \text{ on } B_{i+1} | g = h \text{ on } B_i).$$

Then

$$\begin{aligned}
|X_{i+1}(h) - X_i(h)| &= \sum_{h' \in H[h]} w_{h'} \left(L(h') - \sum_{h^* \in H[h']} L(h^*) q_{h^*} \right) \\
&\leq \sum_{h' \in H} w_{h'} \sum_{h^* \in H(h')} q_{h^*} |L(h') - L(h^*)| \\
&\leq \sum_{h' \in H} w_{h'} \sum_{h^* \in H(h')} q_{h^*} \\
&= 1.
\end{aligned}$$

We are done.

□

9.3 More applications

9.3.1 Image of a random function

Let g be a random function from $[n] \to [n]$, and let $L(g)$ be the size of the image. Then

$$\mathbb{E}L(g) = n - n\left(1 - \frac{1}{n}\right)^n,$$

since the probability that g does not take a fixed value y is $\left(1 - \frac{1}{n}\right)^n$. Note that

$$n - \frac{n}{e} \leq n - n\left(1 - \frac{1}{n}\right)^n \leq n - \frac{n-1}{e}.$$

Clearly, L satisfies the Lipschitz condition. Hence

$$\mathbb{P}\left(\left|L(g) - n\left(1 - \frac{1}{e}\right)\right| > \lambda\sqrt{n} + 1\right) < 2e^{-\lambda^2/2}.$$

9.3.2 Expansions of sets in cubes

Let d_H be the Hamming distance on $\{0,1\}^n$, i. e., $d_H(\underline{x}, \underline{y}) = |\{i \mid x_i \neq y_i\}|$. For $A \subseteq \{0,1\}^n$ and some $s \in \mathbb{R}$ let

$$B(A, s) = \{\underline{y} \mid \exists \underline{x} \in A : \ d_H(\underline{x}, \underline{y}) \leq s\}.$$

Then $A \subseteq B(A, s)$ for every $s \geq 0$.

Theorem 9.3.1. *Let ε, λ satisfy $\varepsilon = e^{-\lambda^2/2}$. Then the following holds true: if $|A| \geq \varepsilon 2^n$ then $|B(A, 2\lambda\sqrt{n})| \geq (1-\varepsilon)2^n$.*

Proof. We can consider $\{0,1\}^n$ as a probability space with the uniform distribution on it. Let X be the following random variable:

$$X(\underline{y}) = \min_{\underline{x} \in A} d_H(\underline{x}, \underline{y}).$$

We can consider the martingale $X_0, X_1, \ldots, X_n$ with respect to the gradation, where at X_i we expose the first i coordinates. Note that X satisfies the Lipschitz condition as $|X(\underline{y}) - X(\underline{y}')| \leq 1$ if $\underline{y}$ and $\underline{y}'$ differ in at most 1 coordinate. Let $\mathbb{E}X = \mu$. Then

$$\mathbb{P}(X < \mu - \lambda\sqrt{n}) < e^{-\lambda^2/2} = \varepsilon \quad \text{and} \quad \mathbb{P}(X > \mu + \lambda\sqrt{n}) < e^{-\lambda^2/2}.$$

Note that $\mu \leq \lambda\sqrt{n}$ since if $\mu > \lambda\sqrt{n}$ then

$$\varepsilon \leq \frac{|A|}{2^n} = \mathbb{P}(X = 0) \leq \mathbb{P}(X < \mu - \lambda\sqrt{n}) < e^{-\lambda^2/2} = \varepsilon$$

would lead to a contradiction. Therefore $\mu \leq \lambda\sqrt{n}$, then

$$1 - \frac{|B(A, 2\lambda\sqrt{n})|}{2^n} = \mathbb{P}(X > 2\lambda\sqrt{n}) \leq \mathbb{P}(X > \mu + \lambda\sqrt{n}) < e^{-\lambda^2/2} = \varepsilon.$$

Then $|B(A, 2\lambda\sqrt{n})| \geq (1 - \varepsilon)2^n$. □

10. Entropy

10.1 Information and counting

The entropy of a probability distribution of $\underline{p} = (p_1, \ldots, p_n)$ is

$$H(\underline{p}) = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i}.$$

The intuition behind entropy that it encodes certain information contained in the probability distribution. This intuition can be formalized by various inequalities, see Proposition 10.2.1. For instance,

$$H(\underline{p}) \leq \ln n,$$

and equality holds true if and only if $\underline{p}$ is the uniform distribution, i. e., $\underline{p} = (\frac{1}{n}, \ldots, \frac{1}{n})$. Based on this inequality one can prove lower bounds in various counting problem. Suppose that we would like to give a lower bound to the cardinality of some set S. If we can give a probability distribution $\underline{p}$ on S and compute $H(\underline{p})$, then we know that $|S| \geq \exp(H(\underline{p}))$. Another idea provides upper bound on $|S|$. Here we start from the uniform distribution on the set S, and uses entropy inequalities such as Shearer's inequality (Theorem 10.2.6) to give an upper bound on the entropy of this uniform distribution that is $\ln |S|$. Such strategy will be carried out in the case of matchings, see Brégman's theorem (Theorem 10.3.1), and in the case of homomorphisms, see the theorem of Galvin and Tetali (Theorem 10.4.1).

10.2 Basic properties of entropy

In this section we give a brief account into the theory of entropy. For a thorough treatment, see for example [7].

Let X be a discrete random variable taking its values in a finite set. The range of X will be denoted by $R(X)$. The entropy of X is defined as

$$H(X) = \sum_{x \in R(X)} \mathbb{P}(X = x) \ln \frac{1}{\mathbb{P}(X = x)}.$$

In case of an event Q we write

$$H(X|Q) = \sum_{x \in R(X)} \mathbb{P}(X = x|Q) \ln \frac{1}{\mathbb{P}(X = x|Q)}.$$

If X and Y are discrete random variables then the conditional entropy is defined as

$$\begin{aligned} H(X|Y) &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) H(X|\{Y = y\}) \\ &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) \sum_{x \in R(X)} \mathbb{P}(X = x|Y = y) \ln \frac{1}{\mathbb{P}(X = x|Y = y)}. \end{aligned}$$

Next we collect some basic facts about entropy.

Proposition 10.2.1. *We have*
(a) $0 \leq H(X) \leq \ln |R(X)|$. *Furthermore, if* X *has the uniform distribution on* $R(X)$ *then* $H(X) = \ln |R(X)|$.
(b) $H(X|Y) = H(X, Y) - H(Y)$.
(c) $H(X, Y) \leq H(X) + H(Y)$.
(d) $H(X, Z|Y) \leq H(X|Y) + H(Z|Y)$.
(e) $H(X) \leq H(X, Y)$.
(f) $H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z)$.
(g) $H(X|Y, Z) \leq H(X|Y)$.
(h) $H(X|Y) \leq H(X|f(Y))$.
(j) $H(f(X)|X) = 0$, *alternatively,* $H(f(X), X) = H(X)$.
(k) $H(f(X)|Y) \leq H(X|Y)$.

Remark 10.2.2. In the following proof we will often need Jensen's inequality. Recall that if f is a convex function on the interval $[a, b]$, and $a_1, \ldots, a_n \in [a, b]$ and $p_1, p_2, \ldots, p_n$ non-negative numbers with sum 1, then

$$\sum_{i=1} p_i f(a_i) \geq f\left(\sum_{i=1}^{n} p_i a_i\right).$$

If f is concave, then the inequality is the opposite:

$$\sum_{i=1} p_i f(a_i) \leq f\left(\sum_{i=1}^{n} p_i a_i\right).$$

Proof. (a) The lower bound is clear from the definition since each term is non-negative. The upper bound follows from Jensen's inequality since $f(x) = \ln(x)$ is a concave function (indeed, $f''(x) = \frac{-1}{x^2} < 0$). So let $p_i = \mathbb{P}(X = x)$ and $a_i = \frac{1}{\mathbb{P}(X=x)}$ then

$$H(X) = \sum_{x \in R(X)} \mathbb{P}(X = x) \ln \frac{1}{\mathbb{P}(X = x)} \leq \ln \left(\sum_{x \in R(X)} \mathbb{P}(X = x) \cdot \frac{1}{\mathbb{P}(X = x)} \right) = \ln |R(X)|.$$

Clearly, we have equality if and only if X has the uniform distribution on $R(X)$, and then $H(X) = \ln |R(X)|$.
(b)

$$\begin{aligned} H(X|Y) &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) H(X|\{Y = y\}) \\ &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) \sum_{x \in R(X)} \mathbb{P}(X = x|Y = y) \ln \frac{1}{\mathbb{P}(X = x|Y = y)} \\ &= \sum_{y \in R(Y)} \mathbb{P}(Y = y) \sum_{x \in R(X)} \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \ln \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \\ &= \sum_{x \in R(X), y \in R(Y)} \mathbb{P}(X = x, Y = y) \ln \frac{\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \\ &= H(X, Y) - H(Y). \end{aligned}$$

(c) Note that

$$H(X) + H(Y) - H(X, Y) = \sum_{x \in R(X), y \in R(Y)} \mathbb{P}(X = x, Y = y) \ln \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)}.$$

Now let us apply Jensen's inequality to the function $f(x) = x \ln x$ with

$$p_i = \mathbb{P}(X = x)\mathbb{P}(Y = y) \text{ and } a_i = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)}.$$

Here $f(x)$ is convex as $f''(x) = \frac{1}{x} > 0$, and naturally the sum of p_i's is 1. Let us introduce the notation $I(X, Y) = H(X) + H(Y) - H(X, Y)$. This is called the mutual information.

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) \\ &= \sum_{x \in R(X), y \in R(Y)} \mathbb{P}(X = x, Y = y) \ln \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)} \end{aligned}$$

$$= \sum_{x\in R(X),y\in R(Y)} \mathbb{P}(X=x)\mathbb{P}(Y=y)\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)} \ln \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)}$$

$$= \sum_i p_i f(a_i) \geq f\left(\sum_i p_i a_i\right)$$

$$= f\left(\sum_{x\in R(X),y\in R(Y)} \mathbb{P}(X=x)\mathbb{P}(Y=y)\cdot \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)}\right)$$

$$= f(1) = 0$$

(d) By the previous part we have

$$H(X|Y=y) + H(Z|Y=y) \geq H(X,Z|Y=y)$$

for all $y \in R(Y)$. Hence

$$\begin{aligned} H(X|Y) + H(Z|Y) &= \sum_{y\in R(Y)} \mathbb{P}(Y=y)(H(X|Y=y) + H(Z|Y=y)) \\ &\geq \sum_{y\in R(Y)} \mathbb{P}(Y=y)H(X,Z|Y=y) \\ &= H(X,Z|Y) \end{aligned}$$

(e) We have

$$H(X,Y) - H(X) = H(Y|X) = \sum_{x\in X} \mathbb{P}(X=x)H(Y|X=x) \geq 0$$

termwise.

(f) This is a direct consequence of $H(X|Y) + H(Z|Y) \geq H(X,Z|Y)$ using that $H(X|Y) = H(X,Y) - H(Y)$, $H(Z|Y) = H(Z,Y) - H(Y)$, and $H(X,Z|Y) = H(X,Z,Y) - H(Y)$.

(g) This is again a direct consequence of $H(X|Y) + H(Z|Y) \geq H(X,Z|Y)$ or equivalently $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$ using that $H(X|Y,Z) = H(X,Y,Z) - H(Y,Z)$ and $H(X|Y) = H(X,Y) - H(Y)$.

(h) We have seen that

$$H(X|Y) = \sum_{x\in R(X),y\in R(Y)} \mathbb{P}(X=x,Y=y)\ln\frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x,Y=y)}.$$

Similarly,

$$H(X|f(Y)) = \sum_{x\in R(X),z\in R(f(Y))} \mathbb{P}(X=x,f(Y)=z)\ln\frac{\mathbb{P}(f(Y)=z)}{\mathbb{P}(X=x,f(Y)=z)}.$$

Now fix an $x \in R(X)$ and a $z \in R(f(Y))$, and observe that

$$\begin{aligned} T_{x,z} &:= \sum_{y:f(y)=z} \mathbb{P}(X=x, Y=y) \ln \frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x, Y=y)} \\ &= \mathbb{P}(X=x, f(Y)=z) \sum_{y:f(y)=z} \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(X=x, f(Y)=z)} \ln \frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x, Y=y)} \\ &\leq \mathbb{P}(X=x, f(Y)=z) \ln \left(\sum_{y:f(y)=z} \frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x, f(Y)=z)} \right) \\ &\leq \mathbb{P}(X=x, f(Y)=z) \ln \frac{P(f(Y)=z)}{\mathbb{P}(X=x, f(Y)=z)}. \end{aligned}$$

In the second step we used Jensen's inequality to the function $f(x) = \ln x$ with

$$p_i = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(X=x, f(Y)=z)} \quad \text{and} \quad a_i = \frac{\mathbb{P}(Y=y)}{\mathbb{P}(X=x, Y=y)}.$$

Now by summing these inequalities for all $x \in R(X)$ and $z \in R(f(Y))$ we get that $H(X|Y) \leq H(X|f(Y))$.
(j) $H(f(X)|X) = \sum_{x \in R(X)} \mathbb{P}(X=x) H(f(X)|\{X=x\}) = 0$ since the inner sum $H(f(X)|\{X=x\}) = 0$ for each $x \in R(X)$. By part (b) we have $0 = H(f(X)|X) = H(f(X), X) - H(X)$. □

An immediate corollary of part (c) is the following.

Theorem 10.2.3. *We have*

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i).$$

Let us see some applications of it.

Theorem 10.2.4. *Let $\mathcal{A}$ be a family of subsets of $\{1, 2, \ldots, n\}$ and suppose that the fraction of sets $A_k \in \mathcal{A}$ containing the element i is p_i. Then*

$$|\mathcal{A}| \leq \exp\left(\sum_{i=1}^{n} H(p_i) \right),$$

where $H(x) = -x \ln x - (1-x) \ln(1-x)$.

Proof. Pick an element of $\mathcal{A}$ uniformly at random, and let $(X_1, \ldots, X_n)$ be its characteristic vector. Then $\ln|\mathcal{A}| = H(X_1, \ldots, X_n)$ by part (a) of Proposition 10.2.1. Then

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i) = \sum_{i=1}^{n} H(p_i)$$

implies the claim. □

Theorem 10.2.5. *Let $p \leq 1/2$ then*

$$\sum_{k \leq np} \binom{n}{k} \leq \exp\left(nH(p)\right),$$

where $H(x) = -x \ln x - (1-x)\ln(1-x)$.

Proof. Let us consider the family $\mathcal{A}$ of subsets of $\{1, 2, \ldots, n\}$ of size at most np. Then

$$|\mathcal{A}| = \sum_{k \leq np} \binom{n}{k}.$$

Let p_i be the fraction of sets containing the element i. Then by symmetry $p_1 = \cdots = p_n$. We also have $np_1 = \sum_{i=1}^{n} p_i = \mathbb{E}|A| \leq np$, where the expected value refers to picking a set $A \in \mathcal{A}$ uniformly at random. Hence $p_1 \leq p \leq 1/2$, and so $H(p_1) \leq H(p)$ since $H(x)$ is monotone increasing on the interval $[0, 1/2]$. Hence

$$\sum_{k \leq np} \binom{n}{k} = |\mathcal{A}| \leq \exp\left(\sum_{i=1}^{n} H(p_i)\right) = \exp\left(nH(p_1)\right) \leq \exp\left(nH(p)\right).$$

□

For a random vector $X = (X_1, \ldots, X_m)$ and an $A \subseteq [m]$ let $X_A = (X_i \mid i \in A)$.

Theorem 10.2.6 (Shearer [6])**.** *Let $X = (X_1, \ldots, X_m)$ be a random vector, and $\mathcal{A}$ be a collection of subsets of $[m]$ possibly with repeats such that each element of $[m]$ is contained in at least t members of $\mathcal{A}$. Then*

$$H(X) \leq \frac{1}{t} \sum_{A \in \mathcal{A}} H(X_A).$$

Proof. By part (f) of Proposition 10.2.1 we know that

$$H(X_A) + H(X_B) \geq H(X_{A \cap B}) + H(X_{A \cup B}).$$

So if we have two sets $A, B \in \mathcal{A}$ such that neither $A \subseteq B$ nor $B \subseteq A$, then we can replace them by $A \cap B$ and $A \cup B$. We can do this step even if A and B are disjoint. This way we cannot increase $\sum H(X_A)$ and every element will be contained in exactly the same number of sets. Moreover we cannot do this step infinitely many times because

$$|A|^2 + |B|^2 \leq |A \cap B|^2 + |A \cap B|^2$$

with strict inequality if $A \setminus B$ and $B \setminus A$ are non-empty. So in each step the sum $\sum |A|^2$ will increase by at least 1 and it is at most $|\mathcal{A}|m^2$. This means that the process will halt. For the final set system $\mathcal{B}$ it will be true that for any $A, B \in \mathcal{B}$ we have $A \subseteq B$ or $B \subseteq A$. This means that we get a set system $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n$ where $n = |\mathcal{A}| = |\mathcal{B}|$. Since every element is still contained in at least t sets we get that the last t sets must be the whole set $[m]$. Hence

$$H(X) \leq \frac{1}{t} \sum_{A \in \mathcal{B}} H(X_A) \leq \frac{1}{t} \sum_{A \in \mathcal{A}} H(X_A).$$

□

Let us see some applications of Shearer's inequality.

Theorem 10.2.7. *Let $\mathcal{F}$ be a family of vectors in $S_1 \times \cdots \times S_n$. Let $\mathcal{G} = \{G_1, \ldots, G_m\}$ be a collection of subsets of $N = \{1, 2, \ldots, n\}$, and suppose that each element $i \in N$ belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$ let $\mathcal{F}_i$ be the set of all projections of the members of $\mathcal{F}$ on G_i. Then*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

Proof. Pick an element of $\mathcal{F}$ uniformly at random, and let the corresponding random variable be $X = (X_1, \ldots, X_n)$. Then

$$k \ln |\mathcal{F}| = kH(X) \leq \sum_{i=1}^{m} H(X_{G_i}) \leq \sum_{i=1}^{m} \ln |\mathcal{F}_i|$$

by Theorem 10.2.6 and part (a) of Proposition 10.2.1. Hence

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

□

Theorem 10.2.8. *Let B be a measurable body in the n-dimensional Euclidean space, and let $\mathrm{Vol}(B)$ denote its n-dimensional volume, and let $\mathrm{Vol}(B_i)$ denote its $(n-1)$-dimensional volume of the projection of B on the hyperplane spanned by the coordinates besides the i-th one. Then*

$$\mathrm{Vol}(B)^{n-1} \leq \prod_{i=1}^{n} \mathrm{Vol}(B_i).$$

Proof. Take finer and finer grids, and let $\mathcal{F}$ be the set of lattice vectors contained in B. Apply the previous theorem to $G_i = N \setminus \{i\}$, and in the limit we get that

$$\mathrm{Vol}(B)^{n-1} \leq \prod_{i=1}^{n} \mathrm{Vol}(B_i).$$

□

10.3 Matchings: Brégman's theorem

Theorem 10.3.1 (Brégman [5])**.** *Let $G = (A, B, E)$ be a bipartite graph with $|A| = |B| = n$. Assume that the degrees of the vertices of A are $d_1, \ldots, d_n$. Let $\mathrm{pm}(G)$ denote the number of perfect matchings of G. Then*

$$\mathrm{pm}(G) \leq \prod_{i=1}^{n} (d_i!)^{1/d_i}.$$

The following theorem on regular bipartite graphs is an immediate corollary of Brégman's theorem. One can prove that the condition on bipartiteness can be dropped.

Theorem 10.3.2. *Let $\mathrm{pm}(G)$ denote the number of perfect matchings. Then for a d–regular bipartite graph G we have*

$$\mathrm{pm}(G)^{1/v(G)} \leq \mathrm{pm}(K_{d,d})^{1/v(K_{d,d})}$$

Proof of Theorem 10.3.1. We will consider a perfect matching as an $f : [n] \to [n]$, $f(i) = j$ if (a_i, b_j) is an edge of the perfect matching. Let X be the random vector $(f(1), \ldots, f(n))$, where we choose a perfect matching f uniformly among all perfect matchings. Then the entropy of X is $H(X) = \ln \mathrm{pm}(G)$. Next we will give an upper bound on $H(X)$. In general, we have

$$H(X) = H(X_1) + \sum_{i=2}^{n} H(X_i | X_{i-1}, \ldots, X_1).$$

We can think of this process as follows: we reveal one by one the neighbors of the vertices in A in the random matching f. When we arrive to some fixed vertex $a \in A$ it might occur that some of its neighbors are already covered by the perfect matching so we can be sure that the conditional entropy is definitely not $\ln d_a$, but something smaller. Unfortunately, it is not clear how much smaller it is since we have no control on how many neighbors of a are already occupied. We overcome this difficulty with a little trick: choose a random permutation $\pi \in S_n$ and apply the chain rule for this random order.

$$H(X) = H(X_{\pi(1)}) + \sum_{i=2}^{n} H(X_{\pi(i)}|X_{\pi(i-1)}, \dots, X_{\pi(1)}).$$

It will be more convenient to rewrite it as follows:

$$H(X) = \sum_{v \in A} H(X_v|X_{\{v':\pi(v')<\pi(v)\}}).$$

Let us average it over all $n!$ permutations of S_n:

$$H(X) = \sum_{v \in A} \frac{1}{n!} \sum_{\pi \in S_n} H(X_v|X_{\{v':\pi(v')<\pi(v)\}}).$$

For a fixed vertex $v \in A$ let us study the quantity

$$\frac{1}{n!} \sum_{\pi \in S_n} H(X_v|X_{\{v':\pi(v')<\pi(v)\}}).$$

For a moment let us stop to examine a general conditional entropy:

$$\begin{aligned} H(X|Y) &= \sum_{y \in R(Y)} \mathbb{P}(y) \sum_{x \in R(X)} \mathbb{P}(X = x|Y = y) \ln \frac{1}{\mathbb{P}(X = x|Y = y)} \\ &\leq \sum_{y \in R(Y)} \mathbb{P}(y) \ln |R(X|Y = y)|. \end{aligned}$$

The point is that the range of X conditioned on $Y = y$ might be smaller than the range of X. In particular, this happens if some neighbor of the vertex v is already occupied. So let $N_v(\pi, f)$ be the number of choices remaining for v if we already know $f(v')$ for all v' for which $\pi(v') < \pi(v)$. Then

$$\frac{1}{n!} \sum_{\pi \in S_n} H(X_v|X_{\{v':\pi(v')<\pi(v)\}}) \leq \sum_{j=1}^{d_v} \mathbb{P}(N_v(\pi, f) = j) \ln j$$

$$= \sum_{j=1}^{d_v} \ln j \frac{|\{(\pi, f) \mid N_v(\pi, f) = j\}|}{n! \cdot \mathrm{pm}(G)}.$$

Now the crucial observation is that

$$\frac{|\{(\pi, f) \mid N_v(\pi, f) = j\}|}{n! \cdot \mathrm{pm}(G)} = \frac{1}{d_v}$$

independently of j. In fact, it is independent of f: once we have fixed f the probability that $N_v(\pi, f) = j$ is $\frac{1}{d_v}$. The reason is simple: let us consider the d_v vertices in A whose f-neighbors are exactly the neighbors of v. If we keep only the ordering of these vertices from π, then with probability $\frac{1}{d_v}$ the vertex v will be the first, with probability $\frac{1}{d_v}$ the vertex v will be the second, etc. Hence

$$\frac{1}{n!} \sum_{\pi \in S_n} H(X_v | X_{\{v' : \pi(v') < \pi(v)\}}) \leq \sum_{j=1}^{d_v} \frac{\ln j}{d_v} = \frac{\ln(d_v!)}{d_v}.$$

Hence

$$H(X) \leq \sum_{v \in V} \frac{\ln(d_v!)}{d_v}.$$

Since $H(X) = \ln \mathrm{pm}(G)$ we get that

$$\mathrm{pm}(G) \leq \prod_{v \in A} (d_v!)^{1/d_v}.$$

□

10.4 Homomorphisms

Let G and H be two graphs. A map $\varphi : V(G) \to V(H)$ is a homomorphism if $(\varphi(u), \varphi(v)) \in E(H)$ whenever $(u, v) \in E(G)$. The number of homomorphisms from G to H is denoted by $\hom(G, H)$. Counting the number of independent sets of a graph G or the number of proper q-colorings of G are both special instances of $\hom(G, H)$. In the latter case $H = K_q$. In the former case H is the graph on 2 vertices that are adjacent, and one of them also has a self-loop.

Theorem 10.4.1 (Galvin and Tetali [13])**.** *Let G be a d–regular bipartite graph, and H be a fixed graph. Then*

$$\hom(G, H)^{1/v(G)} \leq \hom(K_{d,d}, H)^{1/v(K_{d,d})}.$$

Proof. Let $G = (A, B, E)$ and $K_{d,d} = (A_d, B_d, E_d)$. First, we suppose that H is a bipartite graph, and $V(H) = U \cup L$ is the partition. (Later we remove this condition on H.) Let

$$\mathrm{Hom}^{L,U}(G, H) = \{f \in \mathrm{Hom}(G, H) : f(A) \subseteq L, f(B) \subseteq U\}.$$

First we consider $|\mathrm{Hom}^{L,U}(K_{d,d}, H)|$. For any set $S \subseteq L$ let

$$\mathcal{H}(S) = \{f \in \mathrm{Hom}^{L,U}(K_{d,d}, H) \mid f(A_d) = S\},$$

$$T(S) = \{g : [d] \to S : g \text{ surjective}\},$$

and

$$C^U(S) = \{j \in U : (j, i) \in E(H)\ \forall i \in S\}.$$

Then

$$|\mathrm{Hom}^{L,U}(K_{d,d}, H)| = \sum_{S \subseteq L} |T(S)||C^U(S)|^d.$$

Next we show that

$$|\mathrm{Hom}^{L,U}(G, H)| \leq |\mathrm{Hom}^{L,U}(K_{d,d}, H)|^{v(G)/(2d)}.$$

Let f be chosen uniformly at random from $\mathrm{Hom}^{L,U}(G, H)$. We think of f as a vector $(f(v))_{v \in V}$, and f_S denotes the random vector $(f(v))_{v \in S}$. Let $M_v = \{f(w) \mid w \in N(v)\}$. Note that M_v is a set, while $f_{N(v)}$ is a vector. Clearly, $f_{N(v)}$ carries more information than M_v, or in other words, M_v is a function of $f_{N(v)}$. For $v \in B$ and $S \subseteq L$ let $m_v(S)$ denote the probability $\mathbb{P}(M_v = S)$. Clearly, $\sum_S m_v(S) = 1$. Then

$$\begin{aligned}
\ln |\mathrm{Hom}^{L,U}(G, H)| &\overset{(a)}{=} H(f_V) \\
&\overset{(b)}{=} H(f_A) + H(f_B | f_A) \\
&\overset{(d)}{\leq} H(f_A) + \sum_{v \in B} H(f(v) | f_A) \\
&\overset{(g)}{\leq} H(f_A) + \sum_{v \in B} H(f(v) | f_{N(v)}) \\
&\overset{(S)}{\leq} \frac{1}{d} \sum_{v \in B} H(f_{N(v)}) + \sum_{v \in B} H(f(v) | f_{N(v)}) \\
&\overset{(j)}{=} \frac{1}{d} \sum_{v \in B} H(f_{N(v)}, M_v) + \sum_{v \in B} H(f(v) | f_{N(v)})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{=} \frac{1}{d}\sum_{v\in B}(H(M_v)+H(f_{N(v)}|M_v))+\sum_{v\in B}H(f(v)|f_{N(v)})\\
&=\frac{1}{d}\sum_{v\in B}(H(M_v)+H(f_{N(v)}|M_v)+dH(f(v)|f_{N(v)}))\\
&\overset{(h)}{\leq} \frac{1}{d}\sum_{v\in B}(H(M_v)+H(f_{N(v)}|M_v)+dH(f(v)|M_v))\\
&\overset{\text{def}}{=} \frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}\left(m_v(S)\ln\frac{1}{m_v(S)}+m_v(S)H(f_{N(v)}|\{M_v=S\})\right)+\\
&+\frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}(dm_v(S)H(f(v)|\{M_v=S\}))\\
&\overset{(a)}{\leq} \frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}\left(m_v(S)\ln\frac{1}{m_v(S)}+m_v(S)\ln|T(S)|+dm_v(S)\ln|C^U(S)|\right)\\
&=\frac{1}{d}\sum_{v\in B}\sum_{S\subseteq L}m_v(S)\ln\frac{|T(S)||C^U(S)|^d}{m_v(S)}\\
&\overset{(J)}{\leq} \frac{1}{d}\sum_{v\in B}\ln\left(\sum_{S\subseteq L}|T(S)||C^U(S)|^d\right)\\
&=\frac{v(G)}{2d}\ln|\mathrm{Hom}^{L,U}(K_{d,d},H)|.
\end{aligned}
$$

On the top of the signs $=$ or $\leq$ one can see which part of Proposition 10.2.1 we have used. The sign S refers to Shearer's inequality, Theorem 10.2.6. The sign J refers to Jensen's inequality applied to $\ln x$. The def simply means that we use the definition of the (conditional) entropy. Finally, we use our previously found formula for $|\mathrm{Hom}^{L,U}(K_{d,d},H)|$.

Now to finish the proof of the theorem we remove the condition that H is bipartite. Let $H'=H\times K_2$, so H' is a bipartite graph with vertex set $V(H')=V(H)\times\{0,1\}$ and $((v,0),(w,1))\in E(H')$ if $(v,w)\in E(H)$. Let $U=\{(v,0)\mid v\in V(H)\}$ and $L=\{(v,1)\mid v\in V(H)\}$. Then

$$|\mathrm{Hom}^{L,U}(G,H')|=|\mathrm{Hom}(G,H)|.$$

Then we are done. □

10.5 Frankl's union-closed set conjecture

A set system $\mathcal{F} \subseteq 2^{[n]}$ is called union-closed if $A, B \in \mathcal{F}$ implies that $A \cup B \in \mathcal{F}$. The following conjecture due to Péter Frankl is one of the best known conjectures in combinatorics.

Conjecture 10.5.1. Let $\mathcal{F} \subseteq 2^{[n]}$ be a union-closed set system that contains at least one non-empty set. Then there is an element $i \in [n]$ that is contained in at least $1/2$ of the sets of $\mathcal{F}$.

Though this conjecture is not proved there was a significant breakthrough due to Gilmour and the authors of the follow-up papers: Alweiss, Huang, Sellke; Sahin; Chase and Lovett.

Theorem 10.5.2. *Let $\mathcal{F} \subseteq 2^{[n]}$ be a union-closed set system that contains at least one non-empty set. Then there is an element $i \in [n]$ that is contained in at least $\frac{3-\sqrt{5}}{2}$ of the sets of $\mathcal{F}$.*

From now on let $\psi = \frac{3-\sqrt{5}}{2} \approx 0.381$ and $\varphi = 1 - \psi = \frac{\sqrt{5}-1}{2} \approx 0.618$.

In the proof of Theorem 10.5.2 we will use the following lemma whose proof we omit.

Lemma 10.5.3. *Let $h(x) = -x\ln(x) - (1-x)\ln(1-x)$. Then*
(a) The minimum of $\frac{h(x^2)}{xh(x)}$ for $x \in [0,1]$ is obtained at $x = \varphi$, where its value is $\frac{1}{\varphi}$.
(b) The minimum of the function

$$f(x,y) := \frac{h(xy)}{xh(y) + yh(x)}$$

for $x, y \in [0,1]$ is attained at $(x,y) = (\varphi, \varphi)$ where its value is $\frac{1}{2\varphi}$.

Note that $\varphi^2 = 1 - \varphi$, and so $h(\varphi^2) = h(1-\varphi) = h(\varphi)$.

The following lemma is the key lemma, the proof of Theorem 10.5.2 will immediately follows from it.

Lemma 10.5.4. *Let A, B be two independent random variables taking values in $\{0,1\}^n$. Assume for all $i \in [n]$ we have $\mathbb{P}(A_i = 0) \geq p$ and $\mathbb{P}(B_i = 0) \geq p$. Then*

$$H(A \cup B) \geq \frac{p}{2\varphi}(H(A) + H(B)).$$

Proof. Let $A_{<i} = (A_1, \ldots, A_{i-1})$. Then

$$H(A \cup B) = \sum_{i=1}^{n} H((A \cup B)_i | (A \cup B)_{<i}) \geq \sum_{i=1}^{n} H((A \cup B)_i | A_{<i}, B_{<i})$$

since $(A \cup B)_{<i}$ can be determined given by $A_{<i}, B_{<i}$. Let $p(x) = \mathbb{P}(A_i = 0 | A_{<i} = x)$ and $q(y) = \mathbb{P}(B_i = 0 | B_{<i} = y)$. Then

$$H((A \cup B)_i | A_{<i} = x, B_{<i} = y) = h(p(x)q(y)) \geq \frac{1}{2\varphi}(p(x)h(q(y)) + q(y)h(p(y)).$$

Averaging for all x, y we get that

$$\begin{aligned} H((A \cup B)_i | A_{<i}, B_{<i}) &\geq \frac{1}{2\varphi} \left(\mathbb{E}_{A_{<i}} p(A_{<i}) \cdot \mathbb{E}_{B_{<i}} h(q(B_{<i})) + \mathbb{E}_{B_{<i}} q(B_{<i}) \cdot \mathbb{E}_{A_{<i}} h(p((A_{<i}))\right. \\ &= \frac{1}{2\varphi} \left(\mathbb{P}(A_i = 0) H(B_i | B_{<i}) + \mathbb{P}(B_i = 0) H(A_i | A_{<i})\right) \\ &\geq \frac{p}{2\varphi} \left(H(A_i | A_{<i}) + H(B_i | B_{<i})\right) \end{aligned}$$

By summing it for $i \in [n]$ we get the claim. □

Now we are ready to prove Theorem 10.5.2.

Proof of Theorem 10.5.2. Suppose for contradiction that for some union-closed set family $\mathcal{F}$ every $i \in [n]$ appears in at most ψ fraction of the sets. Let A, B be two independent copies of the uniform distribution on $\mathcal{F}$. Then $\mathbb{P}(A_i = 0) > 1 - \Psi = \varphi$ and $\mathbb{P}(A_i = 0) > \varphi$ since ψ is irrational. So

$$H(A \cup B) > \frac{1}{2}(H(A) + H(B)) = \ln |\mathcal{F}|.$$

Since $\mathcal{F}$ is union-closed we get that $H(A \cup B) \leq \ln |\mathcal{F}|$, contradiction.

□

Bibliography

[1] Rudolf Ahlswede and David E Daykin. An inequality for the weights of two families of sets, their unions and intersections. *Probability Theory and Related Fields*, 43(3):183–185, 1978.

[2] Miklós Ajtai, Vašek Chvátal, Monroe M Newborn, and Endre Szemerédi. Crossing-free subgraphs. *North-Holland Mathematics Studies*, 60(C):9–12, 1982.

[3] Noga Alon and Nathan Linial. Cycles of length 0 modulo k in directed graphs. *Journal of Combinatorial Theory, Series B*, 47(1):114–119, 1989.

[4] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[5] Lev M Brégman. Some properties of nonnegative matrices and their permanents. In *Soviet Math. Dokl*, volume 14, pages 945–949, 1973.

[6] Fan R. K. Chung, Péter Frankl, Ronald L. Graham, and James B. Shearer. Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, Series A*, 43(1):23–37, 1986.

[7] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[8] György Elekes. On the number of sums and products. *Acta Arithmetica*, 81(4):365–367, 1997.

[9] Paul Erdős. Graph theory and probability. *Canad. J. Math*, 11(11):34–38, 1959.

[10] Paul Erdős. On a problem of graph theory. *Math. Gaz.*, 47:220–223, 1963.

[11] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(17-61):43, 1960.

[12] Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[13] David Galvin and Prasad Tetali. On weighted graph homomorphisms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 63:97–104, 2004.

[14] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *An exponential bound for the probability of nonexistence of a specified subgraph in a random graph.* Institute for Mathematics and its Applications (USA), 1988.

[15] Eli Shamir and Joel Spencer. Sharp concentration of the chromatic number on random graphsg n, p. *Combinatorica*, 7(1):121–129, 1987.

[16] József Solymosi. Bounding multiplicative energy by the sumset. *Advances in mathematics*, 222(2):402–408, 2009.

[17] Endre Szemerédi and William T. Trotter. Extremal problems in discrete geometry. *Combinatorica*, 3(3-4):381–392, 1983.