The probabilistic method

We will introduce a general methodology which will provide us with tools to prove existence theorems. The main idea is based on the following observation: "if an event has probability >0 to occur, then this event exists".

The main philosophy will be to construct certain probability spaces and study them, deducing that certain event have probability >0 to happen.

A crash course on probability theory.

We introduce the basic tools needed to understand the method. We introduce a definition adapted to our framework

Def / A <u>probability space</u> is a triple $(\Omega, \mathcal{A}, \mathbb{P})$ where

i) Ω is finite or countable
ii) $\mathcal{A}$ is the set of subsets of Ω (set of <u>events</u>)
iii) $\mathbb{P}: \mathcal{A} \to [0,1]$, such that $\mathbb{P}(\emptyset)=0$, $\mathbb{P}(\Omega)=1$ and if $A_i \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ then $\mathbb{P}(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1} \mathbb{P}(A_i)$ (<u>probability function</u>) $\longrightarrow$ $\mathbb{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$ (<u>union bound</u>)

Another key definition is the one of random variable:

Def / Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A <u>random variable</u> X over $(\Omega, \mathcal{A}, \mathbb{P})$ is a function $X: \Omega \to \mathbb{R}$ (which needs some natural properties...)

Hence, if $a \in \mathbb{R}$, the set $\{\omega \in \Omega : X(\omega) = a\} := \{X = a\}$ is an event in $\mathcal{A}$ $\Rightarrow$ we can compute its probability.

Comment / i) We will usually deal with random variables whose image is in $\{0,1\}$ or in $\mathbb{N}$.
ii) A very important type of random variables are <u>indicator functions</u>: for $a \in \Omega$, we consider

$$\mathbb{1}_a : \Omega \to \{0,1\} \subseteq \mathbb{R}, \quad A \mapsto \begin{cases} 0 \text{ if } a \notin A \\ 1 \text{ if } a \in A \end{cases}$$

$\Rightarrow$ Then, if Ω is finite, $\mathbb{P}(\mathbb{1}_a = 1)$,

Some parameters will be of importance in the study of probability spaces.

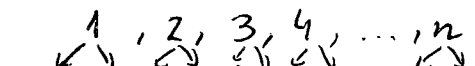
Def / Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and X a random variable with image in $\mathbb{N}$. Then, the <u>expectation</u> and the <u>variance</u> of X are:

$$\mathbb{E}[X] = \sum_{i \geq 0} i\, \mathbb{P}(X=i) \; ; \; \mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Def / Two events A_1 and A_2 in $\mathcal{A}$ are <u>independent</u> iff $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$.

Let us show some examples that will appear later in this part of the course:

Example / Random sets in $[n]$. Take $\Omega = 2^{[n]}$ and $\mathcal{A}$ subsets of Ω. Let $p \in [0,1]$. Then $A \in \mathcal{A}$, $\mathbb{P}(A) = p^{|A|}(1-p)^{n-|A|}$. This space arises from the following model



$\Rightarrow$ we choose each element independently with probability p $\Rightarrow$ <u>binomial model</u>.

Example / Random graph $G(n,p)$: similarly as in the previous example, now Ω = subgraphs of K_N, and A is the set of subsets of such graphs. Then, for $A \in \Omega$, $P(\{A\}) = p^{|E(A)|}(1-p)^{\binom{n}{2}-|E(A)|}$. This can be also be interpreted as a graph constructed randomly by choosing independently edges in K_n with probability p $\Rightarrow$ random graph model $G(n,p)$.

Theorem / (Erdős, 1947) $R(k,k) > 2^{k/2}$.

Proof / We translate the proof in the probabilistic language: take K_N and colour each edge RED / BLUE independently with probability $1/2$. Let Ω the set of all possible colourings.

For each subset S of k vertices of K_N, write A_S the event: "the k vertices of S define a monochromatic K_k". Then, $P(A_S) = 2 \cdot 2^{-\binom{k}{2}}$. Then, the event "$K_N$ has a monochromatic K_k" is exactly $\bigcup_S A_S$. Then, if such event exist, then $P(\bigcup_S A_S) > 0$:

$$P\left(\bigcup_S A_S\right) \leq \sum_S P(A_S) = \binom{N}{k} 2^{1-\binom{k}{2}} \leq \frac{N^k}{2^k} 2^{1-\binom{k}{2}}$$

Now, if $N \leq 2^{k/2}$, then the previous value is < 1. This means that $\bigcup A_S$ is not Ω, and consequently there is a 2-colouring without monochromatic copies of K_k.

Theorem / $W(2,k) > 2^{k/2}$.

Proof / We take the set $[n]$, and we colour each element independently using colours RED / BLUE, with probability $p = 1/2$. For each k-AP S, let us call A_S the event "S is a monochromatic k-AP". Then $P(A_S) = 2 \cdot 2^{-k}$. There are at most $\binom{n}{2}$ such sets, hence, as before,

$$P\left(\bigcup_S A_S\right) \leq \sum_S P(A_S) \leq \binom{n}{2} 2^{1-k} < 1 \Rightarrow n^2 - n < 2^k$$

taking now $n \leq 2^{k/2}$, $n^2 - n \leq 2^k - 2^{k/2} < 2^k$, and $\bigcup_S A_S \neq \Omega$, hence we have 2-colourings without k-AP's.

Let us see another application.

Def / A <u>tournament</u> is the graph K_N (for some N) with an orientation of each edge.

Def / A tournament T_N has <u>the property S_k</u> iff for all set of k vertices, there is a vertex $v \in K_N$ where all edges start at it.

Theorem / (Erdős, 1963) If $\binom{N}{k}(1-2^{-k})^{N-k} < 1$, then there is a tournament T_N with the property S_k.

Proof / Consider random tournaments on K_N by orienting each edge independently with prob. $1/2$. For every fixed subset K of $V(K_N)$ of size k, let A_K be the event "there is not a vertex where all edges start at it". Then, $P(A_K) = (1-2^{-k})^{N-k}$: for every choice of $v \in K_N - K$, the probability that none edge from v to K does not start at v. Hence,

$$P\left(\bigcup_K A_K\right) \leq \sum_K P(A_K) = \binom{N}{k}(1-2^{-k})^{N-k} < 1$$

then there are tournaments without prop S_k.

In all the previous examples we have obtained existence proofs of certain objects, but in any case the proof does not give us how to construct such objects. This is, in general, something very difficult.

Study of the expectation: first moment method.

Next step is based on studying certain random variables and get information about them just from the expectation (also called first moment).

We start exploiting the following easy property from the expectation: both $\mathbb{P}(X \geq \mathbb{E}[X])$ and $\mathbb{P}(X \leq \mathbb{E}[X])$ are greater or equal than 0. A more quantitative version of this is the following theorem:

Theorem / (Markov's inequality) Let X be a non-negative random variable. Then, for every $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}\, \mathbb{E}[X]$$

Proof / Let $\mathbb{I}_\lambda$ be the random variable such that if $A \in \mathcal{A}$, $\mathbb{I}_\lambda(A) = \begin{cases} 1 \text{ if } X(A) \geq \lambda \\ 0 \text{ if } X(A) < \lambda \end{cases}$. Then, $X \geq \lambda\, \mathbb{I}_\lambda$ for all choice of $A \in \mathcal{A}$. Taking now expectation we obtain the result as claimed.

Roughly speaking, this tells us that $X = O(\mathbb{E}[X])$ with high probability: for instance, taking $\lambda = 10\,\mathbb{E}[X]$, then $\mathbb{P}(X \geq 10\,\mathbb{E}[X]) \leq 0.1$.

Another important ingredient is the linearity of the expectation: if $X = c_1 X_1 + c_2 X_2$, then $\mathbb{E}[X] = c_1\,\mathbb{E}[X_1] + c_2\,\mathbb{E}[X_2]$:

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i\, \mathbb{P}(X = i) = \sum_{\omega \in \Omega} X(\omega)\, \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} \{c_1 X_1(\omega) + c_2 X_2(\omega)\}\, \mathbb{P}(\{\omega\})$$

$$= c_1\, \mathbb{E}[X_1] + c_2\, \mathbb{E}[X_2].$$

So, if $X = \sum_{i=1}^{n} c_i X_i$, then $\mathbb{E}[X] = \sum_{i=1}^{n} c_i\, \mathbb{E}[X_i]$.

Example / Let S_n be the set of permutations. Define the probability space $\Omega = S_n$, $\mathcal{A}$ = subsets of S_n, and $\mathbb{P}(\{\sigma\}) = 1/n!$ for all $\sigma \in S_n$ (uniform distribution).
Let $X : S_n \to \mathbb{R}$, $X(\sigma) = \#$ fixed points of σ. If we want to compute the expectation of X directly, this is a difficult question (try!). However, writing $X = \sum_{i=1} X_i$, where $X_i(\sigma) = 1$ if $\sigma(i) = i$, 0 otherwise, then:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n\, \mathbb{E}[X_1] = n\, \frac{(n-1)!}{n!} = \underline{1}$$

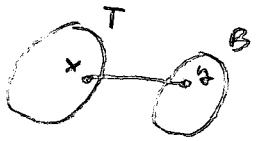
Let us see some applications of these ideas:

Theorem / Let $G = (V, E)$ be a graph with n vertices and e edges. Then G contains a bipartite subgraph with at least $e/2$ edges.

Proof / We define a convenient probabilistic space and a convenient random variable. Let Ω be the set of subsets of V. For $T \in \Omega$, take $\mathbb{P}(\{T\}) = 2^{-|T|}$ (in other words, $\mathbb{P}(x \in T) = 1/2$, where $x \in V$, and the choice is done independent.. For a random subset $T \subseteq V$, write $B = V - T$.

Consider now the random variable $X(T) = \#$ of edges between T and B. Let us study its expectation. For $\overline{xy} \in E$, we write $X_{\overline{xy}}$ for:

$$X_{\overline{xy}}(T) = \begin{cases} 1 \text{ if exactly one of the } x, y \in T \\ 0 \text{ in the other case} \end{cases}$$



Then, $X = \sum X_{\overline{xy}}$, and consequently $\mathbb{E}[X] = \sum \mathbb{E}[X_{\overline{xy}}] = e \cdot \mathbb{E}[X_{\overline{xy}}] = \frac{e}{2}$

As $\mathbb{E}[X] = e/2$, there is some $T \in \Omega$ with $X(T) \geq e/2$, as we wanted to prove.

Theorem / Let $v_1, \dots, v_n \in \mathbb{R}^n$, $|v_i| = 1$. Then, there exist $\varepsilon_1, \dots, \varepsilon_n = \pm 1$ so that $|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n| \leq \sqrt{n}$.

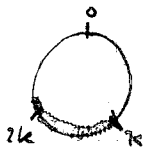
Proof / Let $\varepsilon_1, \dots, \varepsilon_n$ be selected uniformly and independently from $\{-1, +1\}$, and write $X = |\varepsilon_1 v_1 + \dots + \varepsilon_n v_n|^2$. Then,

$$X = \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j (v_i \cdot v_j) \Rightarrow \mathbb{E}[X] = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j \, \mathbb{E}[\varepsilon_i \cdot \varepsilon_j]$$

Let us compute $\mathbb{E}[\varepsilon_i \cdot \varepsilon_j]$. When $i = j$, then $\mathbb{E}[\varepsilon_i^2] = \mathbb{E}[1] = 1$. In the case $i \neq j$, $\mathbb{E}[\varepsilon_i \cdot \varepsilon_j] = 0$. So, $\mathbb{E}[X] = \sum v_i \cdot v_i = n \Rightarrow$ there is some choice of the ε's with value $\leq n$, and $\geq n$.

Theorem / (Erdős, 1965) A set A of nonzero integers is sum-free if there is no solution to the equation $a_1 + a_2 = a_3$ with $a_1, a_2, a_3 \in A$. Then, every set $B = \{b_1, \dots, b_n\}$ of n nonzero integers contain a sum-free subset A of size $|A| > n/3$.

Proof / Let p be a prime, $p \equiv 2 (3)$, which satisfies that $p > 2 \max\{|b_i|\}$. Write $p = 3k+2$. Consider the set $C = \{k+1, \dots, 2k+1\}$. Then $|C| = k+1$, and is sum-free in $\mathbb{Z}/p\mathbb{Z}$.



Take now a random integer x, $1 \leq x < p$, according to the uniform distribution, and define d_i by $d_i \equiv b_i x (p)$. Then, when x is chosen uniformly in $\{1, \dots, p-1\}$, d_i varies uniformly in $\{1, \dots, p-1\} \subseteq \mathbb{Z}/p\mathbb{Z}$. Hence $\mathbb{P}(d_i \in C) = |C|/(p-1) = (k+1)/(3k+1) > 1/3$.

Let X be now the number of elements b_i such that $d_i \in C$. Then $X = \sum X_i$, where $X_i = 1$ iff $d_i \in C$. Then, by linearity of the expectation, $\mathbb{E}[X] > n/3$.

Consequently, there is an x, $1 \leq x < p$ and a subset $A \subseteq B$, such that $|A| > \frac{n}{3}$ and for all $a_1, a_2, a_3 \in A$, $a_1 + a_2 \neq a_3$: in the opposite situation ($\exists a_1, a_2, a_3 \in A : a_1 + a_2 = a_3$) we would have that $a_1 + a_2 \equiv a_3 (p) \Rightarrow x a_1 + x a_2 \equiv x a_3 (p)$, but $x \cdot a_1$, $x a_2$ and $x a_3 \in C$!

The alteration method

In many situations we will be able to construct a probabilistic space and study a certain random variable, BUT without getting positive results (namely, the desired structure). In some cases (and this is what is known as the alteration method) we will be able to "alter" the structure a bit, getting the desired structure.

Let us see some examples to show this:

Theorem / Let $G = (V, E)$ be a graph with n vertices and $nd/2$ edges, $d \geq 1$. Then $\alpha(G) \geq \frac{n}{2d}$.

Proof / Let $S \subseteq V$ be a random subset, built by taking $v \in S$ independently with probability p (we will determine p later). Let $X = \#$ number of element in S, and $Y = \#$ number of edges in the subgraph induced by S. Let us study $\mathbb{E}[X]$ and $\mathbb{E}[Y]$:

i) $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$, where $X_i(S) = 1$ iff $v_i \in S$.

ii) Write $Y = \sum_{e \in E} Y_e$, where $Y_e(S) = 1$ iff the endvertices of $e \in S$. Then

$$\mathbb{E}[Y] = \sum_{e \in E} \mathbb{E}[Y_e] = \frac{nd}{2} \cdot p^2$$

Then, $\mathbb{E}[X - Y] = np - \frac{nd}{2} p^2$. In order to maximize this value, choose $p = 1/d$, getting $n/2d \Rightarrow$ Take a set S with $X - Y \geq n/2d$. Now, select one vertex for each edge, and delete it. This creates a new set S^*, which is independent with at least $n/2d$ vertices!

We show now a major application of the first moment method + alteration method.

Theorem 1 (Erdős, 1959) For all $k, \ell > 0$, there exists a graph with girth $> \ell$ and $\chi(G) > k$.

Proof/ Let us start considering the random graph $G(n,p)$, and let X be the random variable which counts the number of cycles of length at most ℓ in $G(n,p)$. In particular, $X = \sum_{i=3}^{\ell} X_i$, where X_i count the number of cycles of length exactly i. Then:

$$\mathbb{E}[X] = \sum_{i=3}^{\ell} \mathbb{E}[X_i] = \sum_{i=3}^{\ell} \frac{n(n-1)\dots(n-i+1)}{2i} p^i \equiv \sum_{i=3}^{\ell} \frac{(n)_i}{2i} p^i$$

On the other side, we will use that $\chi(G)\alpha(G) \geq |V(G)| \Rightarrow \chi(G) \geq \frac{|V(G)|}{\alpha(G)} = \frac{n}{\alpha(G)}$.

Then,

$$\mathbb{P}(\alpha(G) \geq r) \leq \binom{n}{r} (1-p)^{\binom{r}{2}} \qquad \text{(Union bound)}$$

Let us see how to choose p and r. Fix $\theta < 1/\ell$, and $p = n^{\theta-1}$, and $r = \lceil \frac{3}{p} \log n \rceil$. Then,

i) $\mathbb{E}[X] = \sum_{i=3}^{\ell} \frac{(n)_i}{2i} p^i = \sum_{i=3}^{\ell} \frac{(n)_i}{2i} n^{i\theta - i} \leq \sum_{i=3}^{\ell} \frac{n^{\theta i}}{2i} \leq \frac{\ell}{2} n^{\theta \ell} = o(n)$ $(\theta \cdot \ell < 1)$

ii) $\mathbb{P}(\alpha(G) \geq r) \leq \binom{n}{r}(1-p)^{\binom{r}{2}} \leq n^r (1-p)^{\binom{r}{2}} = (n(1-p)^{r-1/2})^r$ $\leftarrow 1-p \leq e^{-p}$

$\leq (n e^{-p(r-1)/2})^r = o(1)$

Markov's ineq!

From i), in particular, $\mathbb{P}(X \geq n/2) = o(1)$. Choose now n big enough so that both events $(X \geq n/2,\ \alpha(G) \geq r)$ occur with probability less than $1/2$. Then, there is a graph G with less than $n/2$ cycles of length at most ℓ, and with $\alpha(G) < 3n^{1-\theta}\log n$.

Now we apply the alteration argument: we remove from G a vertex from each cycle of length at most ℓ. This gives a graph G^* with at least $n/2$ vertices, whose girth is $> \ell$. Additionally, $\alpha(G^*) \leq \alpha(G)$, hence

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

To complete the proof, we need to take n large enough such that $\frac{n^\theta}{6\log n} > k$.

The second moment method

In the previous argument, we have just needed the information concerning the expectation of the random variable under study. However, we will show that when studying the variance we could obtain concentration type arguments that could be very useful.

The first result we need is the following application of Markov's inequality:

Theorem (Chebyshev's inequality) $\mathbb{P}(|X - \mathbb{E}[X]| \geq \lambda \mathbb{V}\mathrm{ar}[X]^{1/2}) \leq \frac{1}{\lambda^2}$

Proof/ We use the Markov's inequality: consider the random variable $Y = (X - \mathbb{E}[X])^2$. Then $\mathbb{E}[Y] = \mathbb{V}\mathrm{ar}[X]$, and by Markov:

$$\mathbb{P}(|Y|^{1/2} \geq \lambda) = \mathbb{P}(Y \geq \lambda^2) \leq \frac{1}{\lambda^2}\mathbb{E}[Y] \Rightarrow \mathbb{P}(|Y|^{1/2} \geq \lambda\, \mathbb{E}[Y]^{1/2}) \leq \frac{1}{\lambda^2}.$$

This relation tells us that the number of events such that $|X-\mathbb{E}[X]|$ is big is very small. In particular, we can deduce the following result:

Prop/ Let X be a nonnegative integral random variable. Then $\mathbb{P}(X=0) \leq \frac{\mathbb{V}ar[X]}{\mathbb{E}[X]^2}$.

Proof/ We write $\lambda = \mathbb{E}[X]/\mathbb{V}ar[X]^{1/2}$ on Chebyshev inequality:

$$\mathbb{P}(X=0) \leq \mathbb{P}\left(|X-\mathbb{E}[X]| \geq \mathbb{E}[X]\right) = \mathbb{P}\left(|X-\mathbb{E}[X]| \geq \lambda \mathbb{V}ar[X]^{1/2}\right)$$

$$\leq \frac{1}{\lambda^2} = \frac{\mathbb{V}ar[X]}{\mathbb{E}[X]^2}.$$

$\mathbb{P}(|X-\mathbb{E}[X]| \geq \varepsilon \mathbb{E}[X]) \leq \frac{1}{\varepsilon^2}\frac{\mathbb{V}ar[X]}{\mathbb{E}[X]^2}$

Let us see an application of all of this in the context of additive combinatorics:

Def/ A set $\{x_1, \dots, x_k\}$ of positive integers is said to have distinct sums iff all sums $\sum_{i \in S} x_i$, $S \subseteq [k]$ are different.

So the question is to get the minimal size of a subset of $[n]$ which has distinct sums. Of course, the set $2, 2^2, 2^3, \dots, 2^r$, $r \leq \log_2 n$ has distinct sums, so the value is at least $1 + \lfloor \log_2(n) \rfloor$. We show the following:

Theorem/ The size of a set with distinct sums is $< \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$.

Proof/ Take $\{x_1, \dots, x_k\} \subseteq [n]$ with distinct sums, and let $\varepsilon_1, \dots, \varepsilon_k$ be independent r.v. with $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = 0) = 1/2$. Write $X = \varepsilon_1 x_1 + \dots + \varepsilon_k x_k$. Then:

a) $\mathbb{E}[X] = \frac{1}{2}(x_1 + \dots + x_k)$

b) $\mathbb{V}ar[X] = \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2\sum_{\substack{i,j \\ i<j}} \mathbb{E}[\varepsilon_i \varepsilon_j] x_i x_j + \sum_{i=1}^{k} \mathbb{E}[\varepsilon_i^2] x_i^2$

$$- \frac{1}{4}(x_1^2 + \dots + x_k^2) - \frac{1}{2}\sum_{\substack{i,j \\ i<j}} x_i x_j = \frac{1}{4}(x_1^2 + \dots + x_k^2) \leq \frac{n^2 k}{4}.$$

Hence, $\sigma^2 = \mathbb{V}ar[X] \leq \frac{n^2 k}{4}$. So, by Chebyshev's inequality we have that

$$\mathbb{P}\left(|X-\mathbb{E}[X]| \geq \lambda\sigma\right) \leq \lambda^{-2} \Rightarrow \mathbb{P}(|X-\mathbb{E}[X]| < \lambda\sigma) > 1 - 1/\lambda^2$$

$\mathbb{P}(|X-\mathbb{E}[X]| < \lambda\sigma) <$
$\mathbb{P}(|X-\mathbb{E}[X]| < \lambda\frac{n\sqrt{k}}{2})$ So

But now observe that $\mathbb{P}(X=a) \overset{\circ}{\leq} 2^{-k}$, because the set have distinct sums.

$$1 - \frac{1}{\lambda^2} < \mathbb{P}(|X-\mathbb{E}[X]| < \lambda n\sqrt{k}/2) \leq 2^{-k}(\lambda n\sqrt{k}+1) \Rightarrow 1 - \frac{1}{\lambda^2} < 2^{-k}(\lambda n\sqrt{k}+1)$$

$$\Rightarrow n > (2^k(1-\lambda^{-2})-1)/\sqrt{k}\lambda \simeq C(\lambda)\frac{2^k}{\sqrt{k}} \rightarrow \log_2 n > k - \frac{1}{2}\log_2 k - \log_2 C(\lambda)$$

$$\Rightarrow k < \log_2 n + \frac{1}{2}\log_2 k + \log_2 C'(\lambda) < \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$$

$\mathbb{P}[|X-\mathbb{E}[X]| \geq \varepsilon \mathbb{E}[X]] \leq \mathbb{V}ar[X]/\varepsilon^2\mathbb{E}[X]^2$

When studying random models, the study of the variance will give concentration results. Assume that we have a probability space $(\Omega_n, \mathcal{A}, \mathbb{P})$, and a random variable X_n, such that $\mathbb{V}ar[X_n] = o(\mathbb{E}[X_n]^2)$. Then it is clear that $\mathbb{P}(X_n = 0) \to 0$. In this situation, $\mathbb{P}(X_n > 0) \to 1$, and we say that $X_n > 0$ asymptotically almost surely (a.a.s.). In fact, the proof of this fact gives that $X = \mathbb{E}[X](1+o(1))$.

The analysis can be made more accurate if $X = X_n$ is the sum of indicator random variables. Write $X = X_1 + \dots + X_m$, where X_i is the indicator random variable of the event A_i. Write $i \sim j$ if the events A_i, A_j are not independent, and

$$\Delta = \sum_{i \sim j} \mathbb{P}(A_i \cap A_j)$$

Let us study $\mathbb{V}ar[X]$ in terms of Δ:

$$\mathbb{V}ar[X] = \mathbb{E}\left[\left(\sum_{i=1}^{m} X_i - \mathbb{E}[X_i]\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{m}(X_i - \mathbb{E}[X_i])^2\right] + \sum_{i\neq j}\mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right]$$

$$= \sum_{i=1}^{m}\mathbb{V}ar[X_i] + \sum_{i\neq j}\Big(\underbrace{\mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]}_{Cov[X_i,X_j]}\Big)$$

In particular, if X_i, X_j are independent (the event A_i, A_j are independent), $Cov[X_i,X_j]=0$.

Then,

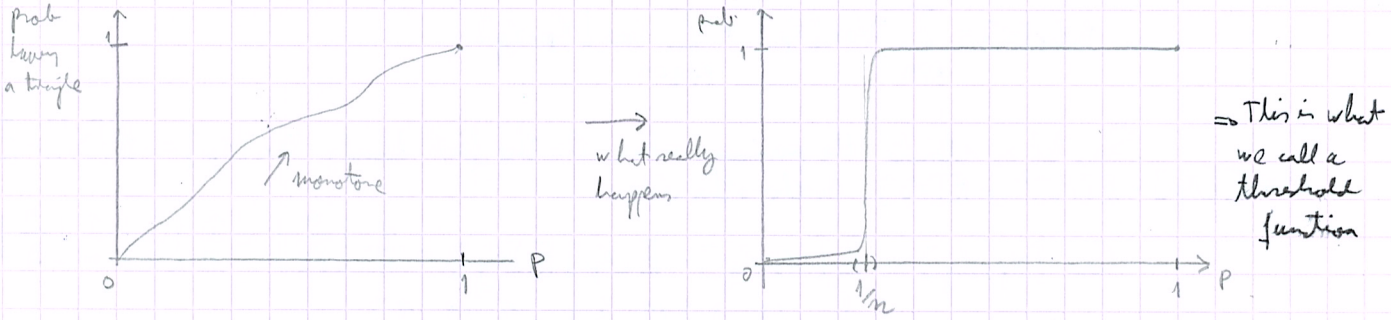
$$\mathbb{V}ar[X_i] = \mathbb{P}(A_i)(1-\mathbb{P}(A_i)) \leq \mathbb{P}(A_i) \overset{\mathbb{E}[X_i]}{=} \Rightarrow \sum_{i=1}^{m}\mathbb{V}ar[X_i] \leq \sum_{i=1}^{m}\mathbb{E}[X_i] = \mathbb{E}[X].$$

$$i\neq j \rightarrow Cov[X_i,X_j] = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_iX_j] = \mathbb{P}(A_i\cap A_j)$$

So, $\mathbb{V}ar[X] \leq \mathbb{E}[X] + \Delta$. Then it is obvious the following:

Theorem / If $\mathbb{E}[X]\to\infty$ and $\Delta = o(\mathbb{E}[X]^2)$, then $X>0$ a.a.s. Furthermore, $X = \mathbb{E}[X](1+o(1))$ a.a.s.

Let us use this result to study the existence of substructures in the model $G(n,p)$. Let us study the number of triangles that $G(n,p)$ has once moving p from 0 to 1. Or even better, we would like to see which is the probability of having a triangle:



Def / Let P be a graph property. A function $r(n)$ is a threshold function for P if

i) $p(n) \ll r(n) \Rightarrow \mathbb{P}(G(n,p) \text{ satisfies } P) \to 0$

ii) $r(n) \ll p(n) \Rightarrow \mathbb{P}(G(n,p) \text{ satisfies } P) \to 1$

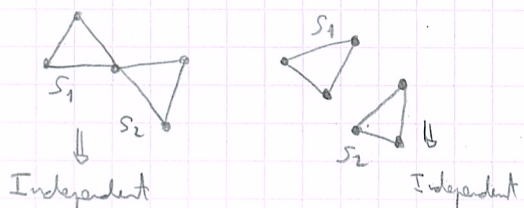
If $r(n)$ is a threshold, $Cr(n)$ it is also a threshold.

Prop / The property $G(n,p)$ has a triangle has threshold function $r(n) = n^{-1}$.

Proof / Let $X_n := X$ be the random variable which counts the number of triangles in $G(n,p)$. Then, $X = \sum_{|S|=3} X_S$, where X_S is the indicator random variable which counts if S forms a triangle or not:

$$\mathbb{E}[X] = \sum_{|S|=3}\mathbb{E}[X_S] = \binom{n}{3}p^3 = \frac{n^3}{6}p^3(1+o(1))$$

Now, if $p \ll n^{-1}$, $\mathbb{E}[X] = o(1)$ and necessarily $X=0$ a.a.s. ($\mathbb{P}(X>0) \leq \mathbb{E}[X] = o(1)$). Let us study the second situation, namely $n^{-1} \ll p(n)$. Then $\mathbb{E}[X]\to\infty$. From this we cannot say that $\mathbb{P}(X>0)=1$ a.a.s., because it could be possible that $\mathbb{P}(X=0)=c$ not depending on n. Let us study Δ:



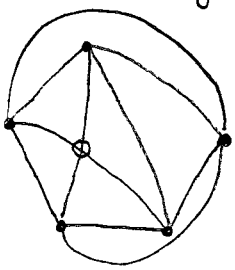
$$\Rightarrow \Delta = \binom{4}{2}\binom{n}{4}\cdot p^5 = \frac{n^4}{2}p^5(1+o(1))$$

$$\frac{\Delta}{\mathbb{E}[X]^2} = \frac{\frac{n^4}{2}p^5}{\frac{n^6}{36}p^6}(1+o(1)) = 18\,n^{-2}p^{-1}(1+o(1)) \to 0$$

Note added: first moment Method

Let us see an extra application of the first moment method in the context of embedded graphs:

Def/ The crossing number of a graph G ($\text{cross}(G)$) is the minimum number of edge crossings of G when drawing it in the plane.



$\text{cross}(K_5) = 1$

In particular, the crossing number of a planar graph is equal to 0.

Additionally, if G has many edges, one would expect that $\text{cross}(G)$ is big. This is the case, as it is stated in the following theorem of Ajtai, Chvátal, Newborn and Szemerédi:

Theorem/ Let G be a graph with $|E(G)| \geq 4|V|$. Then $\text{cross}(G) \geq \dfrac{|E(G)|^3}{64\,|V(G)|^2}$.

Proof/ Recall that a planar graph with n vertices has at most $3n-6$ edges. Then, observe also that any graph G can be made planar by deleting at most $\text{cross}(G)$ edges (namely, one for each crossing). Hence,

$$|E(G)| - \text{cross}(G) \leq 3|V(G)| - 6 \leq 3|V(G)| \Rightarrow \text{cross}(G) \geq |E(G)| - 3|V(G)|.$$
($|E|$ and $|V|$ above)

This is not what we want. Hence, we will amplify it using the first moment method. Take the graph G, $|E(G)| \geq 4|V(G)|$, and take $0 < p < 1$ that will be determined later. We choose now V' a random subset of $V(G)$, where each element is chosen independently with probability p. Write $G' = (V', E')$ for the subgraph of G induced by V'. Then, we have that $\text{cross}(G')$, $|E(G')|$ and $|V(G')|$ become random variables:

- $\mathbb{E}[|V(G')|]$: $|V(G')| = \sum_{i=1}^{|V|} X_i$, where X_i is $\{0,1\}$ depending on wether or not $v_i \in V$ is in V'. Hence, this expectation is $p|V(G)|$.
- $\mathbb{E}[|E(G')|]$: similarly, we have that this value is equal to $p^2|E(G)|$.

Hence, we have that $\mathbb{E}[\text{cross}(G')] \geq \mathbb{E}[|E(G')|] - 3\,\mathbb{E}[|V(G')|] = p^2|E| - p|V|$.

On the other side, consider a drawing of G with $\text{cross}(G)$ crossings. Observe that a crossing remains in G' if the 4 vertices defining a crossing are in G'. Thus, we have, for each crossing, a probability of p^4 to survive. Hence,

$\mathbb{E}[\text{cross}(G')] \leq p^4\,\text{cross}(G)$ (is an upper bound because crossings can dissappear)

Hence, $p^4\,\text{cross}(G) \geq p^2|E| - p|V| \Rightarrow \text{cross}(G) \geq p^{-2}|E| - 3p^{-3}|V|$. Optimizing now p:

$$\frac{\partial}{\partial p}\left(p^{-2}|E| - 3p^{-3}|V|\right) = 0 \Rightarrow -2p^{-3}|E| + 9p^{-4}|V| = 0 \Rightarrow$$

$$\Rightarrow -2p|E| + 9|V| = 0 \Rightarrow p = \frac{9}{2}\frac{|V|}{|E|} \to 4\frac{|V|}{|E|} \leq 1$$

Finally, $\text{cross}(G) \geq \dfrac{1}{64}\dfrac{|E|^3}{|V|^2}$.