TD4 : RANDOM GRAPHS

Exercice 1 Erdős-Rényi model with $n$ vertices and $m$ edges

Originally, random graphs have been defined as $G(n,m)$, which is the set of graphs with $n$ vertices and $m$ vertices exactly. Graphs in $G(n,m)$ are uniformly distributed.

1. Show that a graph in $G(n,m)$ can be constructed as follows: starting from a graph with $n$ vertices and no edge, choose one edge uniformly at random among the $N$ possible edges. Add a second edge chosen uniformly at random from the $N-1$ remaining edges and continue the same way until the graph has $m$ edges.

**Correction :** The final graph has $n$ vertices and $m$ edges. From the construction, every sequence of $m$ edges has the same probability, $(N-m)!/N!$, and for one graph, there are $m!$ possible orders. Then, the probability to get a graph is $1/\binom{N}{m}$, which is exactly the desired probability.

2. Show that conditionally on having $m$ edges, $G_{n,p}$ has the same distribution as in $G(n,m)$ (whatever $p \in (0,1)$).

**Correction :** Let $G$ be a graph with $n$ vertices and $m$ edges.

$$P(G_{n,p} = G \mid G_{n,p} \text{ has } m \text{ edges}) = \frac{P(G_{n,p} = G)}{P(G_{n,p} \text{ has } m \text{ edges})} = \frac{p^m(1-p)^{N-m}}{\binom{N}{m}p^m(1-p)^{N-m}} = \frac{1}{\binom{N}{m}}.$$

Exercice 2 Conditional expectation inequality

1. *(Jensen inequality)* Let $X$ be a real random variable, $I$ an interval and $\phi : I \to \mathbb{R}$ a convex function, such that $P(X \in I) = 1$. Show that if $X$ and $\phi(X)$ are integrable, then $E[\phi(X)] \geq \phi(E[X])$.

**Correction :** If $X$ is a constant random variable, then $X = E[X]$ and $\phi(X) = E(\phi[X]) = \phi(E[X])$. Otherwise, $E(X) \in \hat{\phi}$. But, for all $x_0 \in I$, there exists $\alpha$ such that for all $x$, $\phi(x) \geq \phi(x_0) + \alpha(x - x_0)$. So taking $x_0 = E[X]$, $\phi(X) \geq \phi(E[X]) + \alpha(X - E[X])$, and the expectation, by linearity, $E[\phi(X)] \geq E[\phi(E[X])] + \alpha(E[X] - E[E[X]])$. Hence $E[\phi(X)] \geq E[\phi(E[X])]$.

Set $X = \sum_{i=1}^{n} X_i$, where $X_i$ are random variables in $\{0,1\}$, we want to show that

$$P(X > 0) \geq \sum_{i=1}^{n} \frac{P(X_i = 1)}{E(X \mid X_i = 1)}.$$

Set $Y = 1/X$ if $X \neq 0$ and $Y = 0$ otherwise.

2. Show that $P(X > 0) = E(XY)$.

**Correction :** $X = 0 \Leftrightarrow XY = 0$ and $XY \in \{0,1\}$, so $P(X > 0) = P(XY > 0) = E(XY)$.

3. Show that $E(X_iY) \geq \frac{P(X_i = 1)}{E[X_i \mid X_i = 1]}$.
Correction :

\[ E(X,Y) = E(X,Y \mid X_i = 0)P(X_i = 0) + E(X,Y \mid X_i = 1)P(X_i = 1) \]
\[ \geq E(X,Y \mid X_i = 1)P(X_i = 1) \]
\[ = E(1/X \mid X_i = 1)P(X_i = 1) \geq \frac{P(X_i = 1)}{E(X \mid X_i = 1)}. \]


Correction : 

\[ P(X > 0) = E(XY) = \sum_{i=1}^{n} E(X,Y) = \sum_{i=1}^{n} \frac{P(X_i = 1)}{E(X \mid X_i = 1)}. \]

Exercice 3

Number of triangles in a graph

Consider a graph of \( G_{n,p} \) with \( p = 1/n \). Let \( X \) be its number of triangles.

1. Show that \( P(X \geq 1) \leq 1/6. \)

Correction : Let \( I \) be the set of the subsets of 3 vertices of the graph. For \( i \in I \), set \( X_i = 1 \) if \( i \) is a triangle of the graph and \( X_i = 0 \) otherwise, and \( X = \sum_{i \in I} X_i \). One have \( P(X_i = 1) = p^3 \). So

\[ E(X) = \sum_{i \in I} P(X_i = 1) = \binom{n}{3} \frac{n(n-2)(n-2)}{6n^3} \leq \frac{1}{6}. \]

From the Markov inequality, we get \( P(X \geq 1) \leq E(X) \leq 1/6. \)

2. Show that \( \lim_{n \to \infty} P(X \geq 1) \geq 1/7. \) Indication : Use the previous exercise

Correction : We apply the previous exercise, and for this compute \( E(X \mid X_i = 1). \)

\[ E(X \mid X_i = 1) = \begin{cases} 1 & \text{if } i \text{ is a triangle} \\ + \ 3(n-3)p^2 & \text{triangles with one common edge with } i \\ + \ 3\binom{n-2}{2}p^3 & \text{triangles with one common vertex with } i \\ + \ \binom{3}{n-3}p^3 & \text{triangles disjoint from } i. \end{cases} \]

So, \( E(X \mid X_i = 1) = \frac{7n^3+39n^2-52n-6}{6n^3}. \) Then,

\[ P(X \geq 1) \geq \frac{\binom{3}{n}}{n^3} \frac{6n^3}{7n^3+39n^2-52n-6} \to \frac{1}{7}. \]

Exercice 4

Isolated vertices and connectivity

Let us recall the inequalities :

- for all \( p \in [0,1], \ (1-p) \leq e^{-p}, \) and for all \( p \in [0,1/2], \ (1-p) \geq e^{-p-p^2}; \)
- \( \binom{n}{k} \leq \frac{n^k}{k!} \) and \( k! \geq k^k e^{-k}. \)

For a vertex \( x \), define the random variable

\[ I(x) = \begin{cases} 1 & \text{if } x \text{ is isolated} \\ 0 & \text{otherwise}. \end{cases} \]

Set
1. \( I = \sum_x I(x) \) the number of isolated vertices,
2. \( C = 1 \) if and only if \( G_{n,p} \) is connected.

We first deal with isolated vertices, but the threshold function is the same: when there is no isolated vertex, then with high probability, the graph will be connected.

We say that property \( A \) (depending on \( n \)) is satisfied with \textit{high probability} if \( \lim_{n \to \infty} P(A) = 1 \). The goal of this exercise is to prove the following theorem:

**Theorem 1.** If \( pn - \ln n \to \infty \), then \( G_{n,p} \) is connected with high probability and if \( pn - \ln n \to -\infty \), then the \( G_{n,p} \) is disconnected with high probability.

In this statement and in the following, \( p \) depends on \( n \), but we write \( p \) instead of \( p(n) \).

Let us first focus on the isolated vertices.

1. Show that if \( pn - \ln n \to +\infty \) then \( \lim_{n \to \infty} P(I \neq 0) = 0 \).

\textit{Correction:} For each vertex \( x \), \( P(I(x)) = (1 - p)^{n-1} \). So \( E[I] = n(1 - p)^{n-1} \leq ne^{-p(n-1)} \).

Then by the Markov inequality, \( P(I \neq 0) \leq e^p e^{-pn + \ln n} \). So if \( pn - \ln n \to +\infty \) when \( n \to \infty \), then \( P(I \neq 0) \to 0 \).

We now assume that \( pn - \ln n \to -\infty \).

2. Compute \( \text{Var}(I) \) and show that \( \text{Var}(I) \leq E[I] + E[I]^2 \frac{p}{1 - p} \).

\textit{Correction:}

\[
E[I^2] = E[\sum_x I^2(x)] = \sum_x E[I(x)] + \sum_{x \neq y} E[I(x)I(y)]
\]

\[
= E[I] + n(n - 1)P(I(x) = 1)P(I(y) = 1 | I(x) = 1)
\]

\[
= n(1 - p)^{n-1} + n(n - 1)(1 - p)^{2n-3}.
\]

Then,

\[
\text{Var}(I) = E[I] + n(n - 1)(1 - p)^{2n-3} - n^2(1 - p)^{2n-2} \leq E[I] + E[I]^2 \frac{p}{1 - p}.
\]

3. Show that \( \lim_{n \to \infty} P(I = 0) = 0 \).

\textit{Correction:} With the inequality of question 2, \( P(I = 0) \leq E[I]^{-1} + \frac{p}{1 - p} \). But \( E[I] = ne^{(n-1)\ln(1-p)} \geq ne^{(n-1)(-p-p^2)} \geq e^{np^2} e^{\ln n -pn} \). This inequality holds when \( p < 1/2 \), which is asymptotically true. The second terms tends to \( \infty \), as \( pn - \ln n \to -\infty \), and the first term tends to \( 1 \) : for \( n \) large enough, \( pn < \ln n \), so \( pn < \ln n/n \) and \( p^2 < (\ln n)^2/n^2 \), so \( p^2 n < (\ln n)^2/n \to 0 \). Then \( E[I] \to +\infty \). We also have \( \frac{p}{1 - p} \to 0 \), so \( \lim_{n \to \infty} P(I = 0) = 0 \).

4. Show that in this case \( \lim_{n \to \infty} P(C = 1) = 0 \).

\textit{Correction:} If \( \lim_{n \to \infty} P(I = 0) = 0 \), obviously \( \lim_{n \to \infty} P(C = 1) = 0 \), as if there is an isolated vertex, the graph is disconnected.

Now, let us deal with the connectivity above the threshold and compute the probability that there is no isolated vertex, but the graph is disconnected: \( P(C = 0, I = 0) \). Let \( X_k \) be the number of spanning tree of size \( k \) in the components of size \( k \).
5. Show that $P(C = 0, I = 0) \leq \sum_{k=2}^{n/2} E[X_k]$. 

**Correction:** First notice that if there is no isolated vertex, but the graph is not connected, there exists a connected component of size between 2 and $n/2$. Then

$$P(C = 0, I = 0) \leq \sum_{k=2}^{n/2} P(X_k \geq 1) \leq \sum_{k=2}^{n/2} E[X_k].$$

6. Show that $E[X_k] \leq \binom{n}{k}p^{k-1}k^{k-2}(1-p)^{k(n-k)} \leq n(epm)^k \frac{1}{k!}e^{-pk(n-k)}$. **Indication:** One can admit that the number of spanning trees of a connected graph with $k$ vertices is at most $k^{k-2}$.

**Correction:** for each set of $k$ vertices, there can be a spanning tree if they are in the same connected component, which can happen if there are at least $k - 1$ edges among them. Moreover, among those vertices, there are at most $k^{k-2}$ spanning trees and those vertices are not connected to any other vertex (otherwise it would not be a spanning tree):

$$E[X_k] \leq \binom{n}{k}p^{k-1}k^{k-2}(1-p)^{k(n-k)}.$$

Now, as $(1-u) \leq e^{-u}$, $\binom{n}{k} \leq n^k$ and $k! \geq k^ke^{-k}$,

$$E[X_k] \leq n^k p^{k-1}k^{k-2}e^{-pk(n-k)}k^ke^k = n(epm)^k \frac{1}{k!}e^{-pk(n-k)}$$

We now investigate the case where $p = a \ln n/n$ with $a > 1/2$. This case will be sufficient to study the case where $pn - \ln n \to +\infty$.

7. Show that when $k$ is fixed, $E[X_k] = o(1)$.

**Correction:** For $k \geq 2$, $E[X_k] \leq \frac{np^k}{k!}n(epm)^ke^{-pm} = \frac{np^k}{k!}n(ea(\ln n)n^{-a})^k \to 0$ when $n \to \infty$.

8. Using that $k(n-k) \geq kn/2$, with $x \mapsto xe^{-x/2}$ is decreasing for $x > 2$, show that $E[X_k] \leq n^{1-k/4}$ for $n$ large enough.

**Correction:** $E[X_k] \leq n(epm)^k \frac{1}{k!}e^{-pk(n-k)} \leq n(epm)^ke^{-pk(n-k)/2} = n(ea\ln nn^{-a/2})^k$. As $a > 1/2$, there exists $N$ such that for all $n \geq N$, $ea\ln nn^{-a/2} \leq n^{-1/4}$. Then, $E[X_k] \leq n^{1-k/4}$.

9. Conclude by showing that $P(C = 0, I = 0) \xrightarrow{n \to \infty} 0$ when $pn - \ln n \to +\infty$.

**Correction:** If $p = a \ln n/n$, we now compute $\sum_{k=2}^{n/2} E[X_k] = o(1) + \sum_{k=2}^{n/2} E[X_k] \leq o(1) + \sum_{k=3}^{n/2} n^{1-k/4} = o(1) + \sum_{k=1}^{n/2} n^{-k/4} = o(1) + \frac{n^{-1/4} - n^{-n/8}}{1 - n^{-1/4}} = o(1)$. Now, if $p > a \ln n/n$, $E[X_k]$ decreases, so $P(C = 0, I = 0) \xrightarrow{n \to \infty} 0$. 

4