Asymptotic Degree Distribution of Erdős-Rényi Binomial Random Graphs*

Student: Anirban Nath **Instructor:** Antar Bandopadhyay

May 31^{st} 2020

1 Introduction

Erdős-Rényi model for random graphs is one of the most popular models in graph theory. They are named after mathematicians Paul Erdős and Alfréd Rényi, who first introduced one of the models in 1959, while Edgar Gilbert introduced the other model contemporaneously and independently of Erdős and Rényi. There are two closely related variants of the Erdős–Rényi random graph model.

- In the $G(n, M)$ model, a graph is chosen uniformly at random from the collection of all graphs which have n nodes and M edges.
- In the $G(n, p)$ model, a graph is constructed by connecting nodes randomly. Each edge is included in the graph with probability p independent from every other edge.

However, for the rest of this article, we shall be considering the $G(n, p)$ model for our purpose.

2 Degree Sequence of the Erdős-Rényi Random Graph

In this section, we would like to investigate the nature of the degree of *a uniformly selected vertex* **given** an Erdős-Rényi random graph. Since the vertex to be selected is arbitrary, its distribution will asymptotically be same as the empirical distribution of all the vertices **given** the graph. Suppose the graph has n vertices and D_i denote the degree of vertex i. Then the empirical degree distribution will be

$$P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}}$$

Remark. Even if we fix a vertex and check its degree the distribution will be same, but the choice should be arbitrary i.e. there is a fair possibility of occurrence of

*Submitted as a part of Random Graphs course at ISI Kolkata an isolated vertex. If we intentionally choose that one, the result will obviously not follow.

Now, let's do some simulations and try to figure out how the distributions will look like.

2.1 Simulation of ER Random graphs with Different Parameters

For the simulation, we have used a R-package called "igraph". It allows us to generate the graph right away and also can calculate the degree distribution with built-in commands. Once the graph is generated, we have taken its degree distribution and compared it to the poisson pmf side by side using 'barplot'.For some given parameter, say 'n' and 'λ/n', we compare the degree distributopn with the pmf of Poisson(λ). The code will look like the following:

```
> install.packages("igraph")
> library(igraph)
> g = erdos.renyi.game(n,\lambda/n,type="gnp",directed=FALSE,
  loops=FALSE)
> l = length(degree.distribution(g))
> barplot(rbind(degree.distribution(g),dpois(c(0:l-1),\lambda))
  ,beside = TRUE, xlab = "Degree",ylab = "Proportion")
```

The followings are some of the simulations with proper parameters.

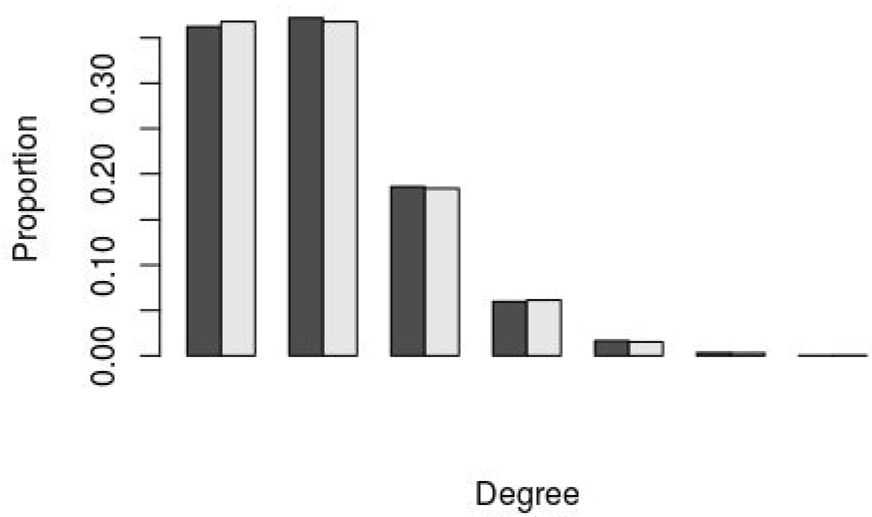

Figure 1: The degree distribution of $ER_n(\lambda/n)$ with $n = 10000$ and $\lambda = 1$ (in the left) and the probability mass function of a Poisson random variable with parameter 1

2.2 Main Theorems and Results

Now, we can have an intuition that the distributions follow power law which is the following:

Definition 2.1 (Power Law)**.** A Power Law is a function $f(x)$ where the value y is proportional to some power of the input x i.e. $f(x) = y = x^{-\alpha}$ If the function describes the probability of being greater than x, it is called a power law distribution (or cumulative distribution function - CDF) and is denoted $P(> x) = x^{-\alpha}$

In this section we will mainly focus on proving the following theorem. In order to be able to state the result, we first introduce some notation. We write

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}$$

for the Poisson probability mass function with parameter λ.

Theorem 2.1 (Degree sequence of the Erdős-Rényi random graph)**.** *Fix* $\lambda > 0$*. Then, for every* ϵ_n *such that* $n\epsilon_n^2 \to \infty$*,*

$$\mathbb{P}_\lambda\left(\max_{k\geq 0}|P_k^{(n)} - p_k| \geq \epsilon_n\right) \to 0$$

Remark. The proof involves some ideas and results of coupling which we shall discuss in the next section in details.

Proof. First, note that,

$$\mathbb{E}_\lambda\left[P_k^{(n)}\right] = \mathbb{E}_\lambda\left[\frac{1}{n}\sum_{i\in[n]}\mathbb{1}_{\{D_i=k\}}\right] = \mathbb{P}_\lambda\left(D_1 = k\right) = \binom{n-1}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-1}$$

because $D_1 \sim Bin(n-1, \frac{\lambda}{n})$ i.e. node 1 has $n-1$ edges to connect each with probability $\frac{\lambda}{n}$. Now,

$$\sum_{k\geq 0}\left|p_k - \mathbb{E}_\lambda[P_k^{(n)}]\right| = \sum_{k\geq 0}\left|p_k - \mathbb{P}_\lambda\left(D_1 = k\right)\right| = \sum_{k\geq 0}\left|\mathbb{P}_\lambda\left(X^* = k\right) - \mathbb{P}_\lambda\left(X_n = k\right)\right|$$

where $X^* \sim Poi(\lambda)$ and $X_n \sim Bin\left(n-1, \frac{\lambda}{n}\right)$. We shall bound the difference by a coupling argument as follows. Let Y_n be a random variable following binomial distribution with parameter n and $p = \frac{\lambda}{n}$. Then, Y_n can be written as $Y_n = X_n + I_n$ where $I_n \sim Ber\left(\frac{\lambda}{n}\right)$ and X_n and I_n are independent. We can couple Y_n and X_n such that the probability that they are different is bounded by,

$$\begin{aligned}
&\sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(Y_n = k)| \\
&\quad = \sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k, I_n = 0) - \mathbb{P}(X_n = k-1, I_n = 1)| \\
&\quad = \sum_{k\geq 0} \left|\mathbb{P}(X_n = k) - \left(1 - \frac{\lambda}{n}\right)\mathbb{P}(X_n = k) - \frac{\lambda}{n}\mathbb{P}(X_n = k-1)\right| \\
&\quad = \frac{\lambda}{n}\sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k-1)| \\
&\quad \leq \frac{2\lambda}{n}
\end{aligned}$$

Therefore, for all $k \geq 0$, we have,

$$\begin{aligned}
&\sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X^* = k)| \\
&\quad = \sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(Y_n = k)| + \sum_{k\geq 0} |\mathbb{P}(X_* = k) - \mathbb{P}(Y_n = k)| \\
&\quad \leq \frac{2\lambda}{n} + \sum_{k\geq 0} |\mathbb{P}(X_* = k) - \mathbb{P}(Y_n = k)| \leq \frac{2\lambda + \lambda^2}{n}
\end{aligned}$$

The last part of the equation seems strange as we have used a result due to coupling there which we shall discuss in the following section.

Now, since λ fixed, for sufficiently large n, we shall get $\frac{2\lambda+\lambda^2}{n} \leq \frac{\epsilon_n}{2}$ and thus we have shown that for sufficiently large n, the sum $\sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X^* = k)|$ is bounded by $\frac{\epsilon_n}{2}$. Therefore we have gathered the results needed to prove the actual theorem.

From triangle inequality, it follows that,

$$\begin{aligned}
\left|P_k^{(n)} - p_k\right| &\leq \left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| + \left|\mathbb{E}_\lambda\left[P_k^{(n)}\right] - p_k\right| \\
\implies \left|P_k^{(n)} - p_k\right| &\leq \left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| + \sum_{k\geq 0}\left|\mathbb{E}_\lambda\left[P_k^{(n)}\right] - p_k\right| \\
\implies \left|P_k^{(n)} - p_k\right| &\leq \left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| + \frac{\epsilon_n}{2} \\
\implies \max_{k\geq 0}\left|P_k^{(n)} - p_k\right| &\leq \max_{k\geq 0}\left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| + \frac{\epsilon_n}{2} \\
\implies \mathbb{P}_\lambda\left(\max_{k\geq 0}|P_k^{(n)} - p_k| \geq \epsilon_n\right) &\leq \mathbb{P}_\lambda\left(\max_{k\geq 0}\left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| \geq \frac{\epsilon_n}{2}\right)
\end{aligned}$$

Thus, it is enough to show that,

$$\mathbb{P}_\lambda\left(\max_{k\geq 0}\left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| \geq \frac{\epsilon_n}{2}\right) = o(1)$$

Now, we shall use Boole's Identity that gives us,

$$\mathbb{P}_\lambda\left(\max_{k\geq 0}\left|P_k^{(n)}-\mathbb{E}_\lambda\left[P_k^{(n)}\right]\right|\geq\frac{\epsilon_n}{2}\right)\leq\sum_{k\geq 0}\mathbb{P}_\lambda\left(\left|P_k^{(n)}-\mathbb{E}_\lambda\left[P_k^{(n)}\right]\right|\geq\frac{\epsilon_n}{2}\right)$$

Then, for a fixed $k\geq 0$, by Chebychev's Inquality, we get,

$$\mathbb{P}_\lambda\left(\left|P_k^{(n)}-\mathbb{E}_\lambda\left[P_k^{(n)}\right]\right|\geq\frac{\epsilon_n}{2}\right)\leq\frac{4\mathrm{Var}_\lambda(P_k^{(n)})}{\epsilon_n^2}$$

Thus, all that is left to show is that the right hand side of the above inequality goes to 0. So, we start calculating the variance. Now, we have to keep in mind that D_i's are identically distributed, but not independent. Hence,

$$\begin{aligned}\mathrm{Var}_\lambda(P_k^{(n)})&=\mathrm{Var}_\lambda\left[\frac{1}{n}\sum_{i\in[n]}\mathbb{1}_{\{D_i=k\}}\right]\\&=\frac{1}{n^2}\left[\sum_{i\in[n]}\mathrm{Var}_\lambda\left(\mathbb{1}_{\{D_i=k\}}\right)+\sum_{i\neq j}\mathrm{Cov}_\lambda\left(\mathbb{1}_{\{D_i=k\}},\mathbb{1}_{\{D_j=k\}}\right)\right]\\&=\frac{1}{n}\left[\mathbb{P}_\lambda(D_1=k)-\mathbb{P}_\lambda(D_1=k)^2\right]+\frac{n-1}{n}\left[\mathbb{P}_\lambda(D_1=D_2=k)-\mathbb{P}_\lambda(D_1=k)^2\right]\end{aligned}$$

Here, we again use a coupling argument. Let, $X_1,X_2\overset{i.i.d}{\sim}Bin(n-2,\lambda/n)$, and I_1, I_2 are two independent Bernoulli random variables with success probability λ/n. Then, the distribution of (D_1,D_2) is the same as the one of (X_1+I_1,X_2+I_1) while (X_1+I_1,X_2+I_2) are two independent copies of D_1. Thus,

$$\begin{aligned}\mathbb{P}_\lambda(D_1=D_2=k)&=\mathbb{P}_\lambda((X_1+I_1,X_2+I_1)=(k,k))\\\mathbb{P}_\lambda(D_1=k)^2&=\mathbb{P}_\lambda((X_1+I_1,X_2+I_2)=(k,k))\end{aligned}$$

So, we have,

$$\begin{aligned}&\mathbb{P}_\lambda(D_1=D_2=k)-\mathbb{P}_\lambda(D_1=k)^2\\&=\mathbb{P}_\lambda((X_1+I_1,X_2+I_1)=(k,k))-\mathbb{P}_\lambda((X_1+I_1,X_2+I_2)=(k,k))\\&\leq\mathbb{P}_\lambda\big((X_1+I_1,X_2+I_1)=(k,k)),(X_1+I_1,X_2+I_2)\neq(k,k)\big)\end{aligned}$$

It can happen only when $I_1\neq I_2$. If, $I_1=0$, then $I_2=1$ and $X_1=k$ and when, $I_1=1$. then $I_2=0$ and $X_2=k-1$. Thus,

$$\mathrm{Var}_\lambda(P_k^{(n)})\leq\frac{1}{n}\mathbb{P}_\lambda(D_1=k)+\frac{\lambda}{n}\left[\mathbb{P}_\lambda(X_1=k)+\mathbb{P}_\lambda(X_2=k-1)\right]$$

Thus, from the previous inequality we get,

$$\begin{aligned}
&\mathbb{P}_\lambda\left(\max_{k\geq 0}\left|P_k^{(n)} - \mathbb{E}_\lambda\left[P_k^{(n)}\right]\right| \geq \frac{\epsilon_n}{2}\right)\\
&\leq \frac{4}{\epsilon_n^2}\sum_{k\geq 0}\left[\frac{1}{n}\mathbb{P}_\lambda(D_1 = k) + \frac{\lambda}{n}\mathbb{P}_\lambda(X_1 = k) + \mathbb{P}_\lambda(X_2 = k-1)\right]\\
&= \frac{4(2\lambda+1)}{\epsilon_n^2 n} \to 0
\end{aligned}$$

since $n\epsilon_n^2 \to \infty$. □

3 Coupling

Coupling is a very useful tool for proving results in statistics and probability theory. In this section, we shall discuss what a coupling is, and prove some results that we have used previously in the previous theorem. Two random variables, let say X and Y are said to be coupled, if they are defined on the same probability space, but have specific marginal distribution. The random variables are defined on same probability space means that, there exist a joint distribution of X and Y say, $\mathbb{P}$ such that, $\mathbb{P}(X \in \mathcal{E}, Y \in \mathcal{F})$ for all events $\mathcal{E}$ and $\mathcal{F}$.

Definition 3.1 (Coupling of Random Variables). The random variables $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$ are a coupling of the random variables $X_1, X_2, \ldots, X_n$, when $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$ are defined on the same probability space and are such that the marginal distribution of $\hat{X}_i$ is same as that of X_i for all $i = 1, 2, \ldots, n$ that is for all measurable set $\mathcal{E} \in \mathbb{R}$,

$$\mathbb{P}(\hat{X}_i \in \mathcal{E}) = \mathbb{P}(X_i \in \mathcal{E})$$

Remark. While the random variables $X_1, X_2, \ldots, X_n$ may not be defined on one probability space, the coupled random variables $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$ are defined on the same probability space. The coupled random variables $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$ are related to the original random variables $X_1, X_2, \ldots, X_n$ by the fact that the marginal distributions of $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$ are equal to those of the random variables $X_1, X_2, \ldots, X_n$.

Couplings are useful to prove that random variables are related. We now describe a general coupling between two random variables that makes them equal with high probability. Let, X and Y be two discrete random variables with the following probability mass functions

$$\mathbb{P}(X = x) = p_x \qquad \mathbb{P}(Y = y) = q_y \qquad x \in \mathcal{X}, y \in \mathcal{Y}$$

Now, a convenient distance between discrete probability distributions $(p_x)_{x\in\mathcal{X}}$ and $(q_y)_{y\in\mathcal{Y}}$ is the *total variation distance* between the discrete probability mass functions $(p_x)_{x\in\mathcal{X}}$ and $(q_y)_{y\in\mathcal{Y}}$.

Definition 3.2 (total variation distance)**.** For two probability measures μ and ν, the total variation distance between them is defined as

$$d_{TV}(\mu,\nu) = \sup_{A\subseteq\mathbb{R}} |\mu(A) - \nu(A)|$$

where, $\mu(A)$ and $\nu(A)$ are the probabilities of the event A under the measures and the supremum is taken over all the possible Borel Subsets A of $\mathbb{R}$.

For discrete probability mass functions,according to the above example we have,

$$\mu(A) = \sum_{a\in A} p_a \qquad \nu(A) = \sum_{a\in A} y_a \qquad A = \mathcal{X}\cup\mathcal{Y}$$

and the total variation distance between them will be

$$d_{TV}(p,q) = \sup_{A\in\mathbb{R}} \left|\sum_{a\in A}(p_a - y_a)\right| = \max\left\{\sum_x (p_x - (p_x\wedge q_x)), \sum_x (q_x - (p_x\wedge q_x))\right\}$$
$$= \frac{1}{2}\sum_x |p_x - q_x|$$

For continuous random variables, the expressions are analogical. If F and G are two distribution functions corresponding to two continuous densities $f = (f(x))_{x\in\mathcal{X}}$ and $g = (g(y))_{y\in\mathcal{Y}}$ respectively, then the total variation distance will be,

$$d_{TV}(f,g) = \frac{1}{2}\int_{-\infty}^{\infty} |f(x) - g(x)|dx$$

Theorem 3.1 (Maximal coupling)**.** *For any two discrete random variables X and Y , there exists a coupling $(\hat{X},\hat{Y})$ of X and Y such that,*

$$\mathbb{P}(\hat{X}\neq\hat{Y}) = d_{TV}(p,q)$$

while, for any other coupling, we have,

$$\mathbb{P}(\hat{X}\neq\hat{Y}) \geq d_{TV}(p,q)$$

Proof. We start by defining the coupling that achieves the equality. For this, we define the random vector $(\hat{X},\hat{Y})$ by,

$$\mathbb{P}(\hat{X} = \hat{Y} = x) = p_x\wedge q_x$$
$$\mathbb{P}(\hat{X} = x, \hat{Y} = y) = \frac{(p_x - (p_x\wedge q_x))(q_y - (p_y\wedge q_y))}{\frac{1}{2}\sum_z |p_z - q_z|}, x\neq y$$

First of all, observe that,

$$\sum_x (p_x - (p_x\wedge q_x)) = 1 - \sum_x (p_x\wedge q_x) = \sum_x (q_x - (p_x\wedge q_x))$$
$$= 1 - \frac{1}{2}\sum_x (|p_x + q_x| - |p_x - q_x|) = \frac{1}{2}\sum_x |p_x - q_x|$$

Thus, the joint distribution defined above is indeed a probability distribution. Also note that,

$$\mathbb{P}(\hat{X} = x) = p_x \qquad\qquad \mathbb{P}(\hat{Y} = y) = q_y$$

Then,

$$\mathbb{P}(\hat{X} \neq \hat{Y}) = 1 - \mathbb{P}(\hat{X} = \hat{Y}) = 1 - \sum_x (p_x \wedge q_x) = \frac{1}{2} \sum_x |p_x - q_x|$$

This proves the first part of the theorem.
For the latter part, we proceed as follows. For all x, and any coupling $(\hat{X}, \hat{Y})$ of X and Y, we must have,

$$\mathbb{P}(\hat{X} = \hat{Y} = x) \leq \mathbb{P}(\hat{X} = x) = \mathbb{P}(X = x) = p_x$$

and also,

$$\mathbb{P}(\hat{X} = \hat{Y} = x) \leq \mathbb{P}(\hat{Y} = x) = \mathbb{P}(Y = x) = q_x$$

which implies that,

$$\begin{aligned} \mathbb{P}(\hat{X} = \hat{Y} = x) &\leq (p_x \wedge q_x) \\ \implies \mathbb{P}(\hat{X} = \hat{Y}) = \sum_x \mathbb{P}(\hat{X} = \hat{Y} = x) &\leq \sum_x (p_x \wedge q_x) \\ \implies \mathbb{P}(\hat{X} \neq \hat{Y}) = 1 - \mathbb{P}(\hat{X} = \hat{Y}) &\geq 1 - \sum_x (p_x \wedge q_x) = \frac{1}{2} \sum_x |p_x - q_x| \end{aligned}$$

The coupling above attains this equality, which makes it the best coupling possible, in the sense that it maximizes $\mathbb{P}(\hat{X} = \hat{Y})$. □

The following theorem is going to explain the result we used earlier in the preceding section.

Theorem 3.2 (Poisson limit for binomial random variables). *Let $(I_i)_{i=1}^n$ be independent with $I_i \sim Ber(p_i)$, and let $\lambda = \sum_{i=1}^n p_i$. Let $X = \sum_{i=1}^n I_i$ and Y be a Poisson random variable with parameter λ. Then there exists a coupling $(\hat{X}, \hat{Y})$ of random variables X and Y such that*

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \sum_{i=1}^n p_i^2$$

Consequently, for any $\lambda \geq 0$ and $n \in \mathbf{N}$, there exists a coupling $(\hat{X}, \hat{Y})$ of random variables X and Y where $X \sim Bin(n, \lambda/n)$ and $Y \sim Poi(\lambda)$ such that

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \frac{\lambda^2}{n}$$

Proof. Let us define random variables $J_i \sim Poi(p_i)$ for all $i = 1, 2, \ldots, n$ and they are independent. Moreover we write their p.m.fs as

$$p_{i,x} = \mathbb{P}(I_i = x) = p_i \mathbb{1}_{\{x=1\}} + (1 - p_i) \mathbb{1}_{\{x=0\}}$$

$$q_{i,x} = \mathbb{P}(J_i = x) = e^{-p_i}\frac{p_i^x}{x!}$$

Let, $(\hat{I}_i, \hat{J}_i)$ be a coupling of I_i and J_i where $(\hat{I}_i, \hat{J}_i)$ are independent for different i. Now, for each pair I_i,J_i, the maximal coupling $(\hat{I}_i, \hat{J}_i)$ described above satisfies

$$\mathbb{P}(\hat{I}_i = \hat{J}_i = x) = p_{i,x} \wedge q_{i,x} = \begin{cases} 1 - p_i & x = 0 \\ p_i e^{-p_i} & x = 1 \\ 0 & x \geq 2 \end{cases}$$

Thus, we obtain,

$$\mathbb{P}(\hat{I}_i = \hat{J}_i) = 1 - \mathbb{P}(\hat{I}_i \neq \hat{J}_i) = 1 - (1 - p_i) - (p_i e^{-p_i}) = p_i(1 - e^{-p_i}) \leq p_i^2$$

Now, let $\hat{X} = \sum_{i=1}^n \hat{I}_i$ and $\hat{Y} = \sum_{i=1}^n \hat{J}_i$. Then $\hat{X}$ has the same distribution as $X = \sum_{i=1}^n I_i$ and $\hat{Y}$ has the same distribution as $Y = \sum_{i=1}^n J_i \sim Poi(p_1 + p_2 + \cdots + p_n)$. Finally, by Boole's Inequality, we obtain

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \mathbb{P}\left(\bigcup_{i=1}^n \{\hat{I}_i \neq \hat{J}_i\}\right) \leq \sum_{i=1}^n \mathbb{P}\left(\hat{I}_i \neq \hat{J}_i\right) \leq \sum_{i=1}^n p_i^2$$

For the later part, we choose $p_i = \lambda/n$ and the result follows. □

4 Reference

For the above article, we have sought help from the following book(s):

[1] Random graphs and complex networks (Vol-1) by Remco van der Hofstad.