Asymptotic Degree Distribution of Erdős-Rényi Binomial Random Graphs[1]

Anirban Nath

Indian Statistical Institute, Kolkata

June 15, 2020

[1] Submitted as a part of Random Graphs course at ISI Kolkata

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.

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.

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.

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 where $p = \lambda/n$. Why this specific form of p shall be useful, we shall see in a bit.

Degree Distribution

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.

Degree Distribution

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

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

Degree Distribution

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

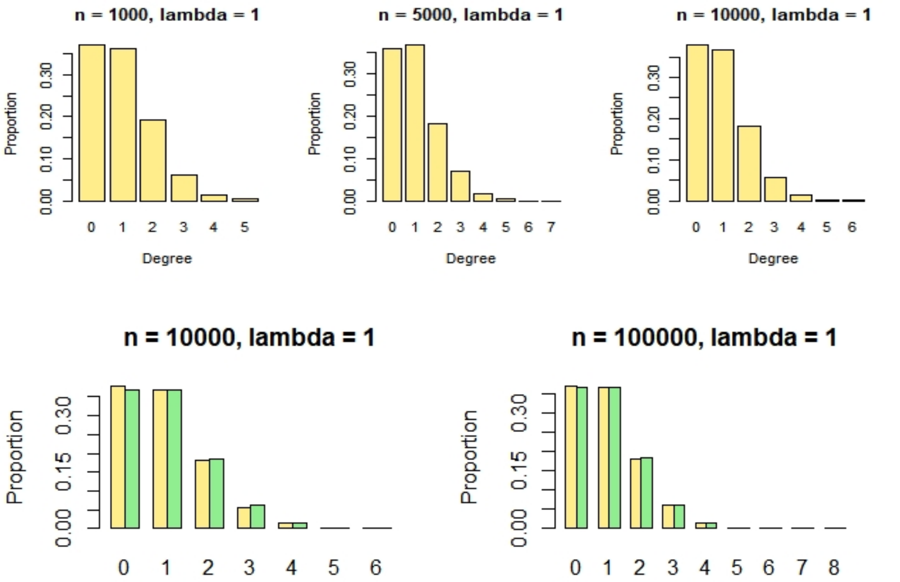
$$\mathbb{P}_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}}$$

Remark

Even if we fix a vertex and check its degree the distribution will be same, but the choice should be arbitrary i.e. if we intentionally choose an isolated vertex, the result will obviously not follow.

We also define the Poisson(λ) pmf as $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k \geq 0$

Simulation


n = 1000, lambda = 1
Proportion
0.00 0.10 0.20 0.30
0 1 2 3 4 5
Degree
n = 5000, lambda = 1
Proportion
0.00 0.10 0.20 0.30
0 1 2 3 4 5 6 7
Degree
n = 10000, lambda = 1
Proportion
0.00 0.10 0.20 0.30
0 1 2 3 4 5 6
Degree
n = 10000, lambda = 1
Proportion
0.00 0.15 0.30
0 1 2 3 4 5 6
n = 100000, lambda = 1
Proportion
0.00 0.15 0.30
0 1 2 3 4 5 6 7 8

Main Theorem

Theorem

Fix $\lambda > 0$. Then, for every ϵ_n such that $n\epsilon_n^2 \to \infty$,

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

Main Theorem

Theorem

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.

Main Theorem

Theorem

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,

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

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,

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

Main Theorem

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.

Main Theorem

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 $Y_n \sim Bin(n, \frac{\lambda}{n})$. Then, $Y_n = X_n + I_n$ where $I_n \sim Ber\left(\frac{\lambda}{n}\right)$ and X_n and I_n are independent.

$$\sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(Y_n = k)|$$

$$= \sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k, I_n = 0) - \mathbb{P}(X_n = k-1, I_n = 1)|$$

$$= \frac{\lambda}{n} \sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k-1)| \leq \frac{2\lambda}{n}$$

Main Theorem

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 $Y_n \sim Bin(n, \frac{\lambda}{n})$. Then, $Y_n = X_n + I_n$ where $I_n \sim Ber\left(\frac{\lambda}{n}\right)$ and X_n and I_n are independent.

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

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

$$\sum_{k\geq 0} |\mathbb{P}(X_n = k) - \mathbb{P}(X^* = k)| \leq \frac{2\lambda + \lambda^2}{n}$$

Main Theorem

Thus, it is enough to show that,

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

Main Theorem

Thus, it is enough to show that,

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

Boole's Identity gives us,

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

Main Theorem

Thus, it is enough to show that,

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

Boole's Identity gives us,

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

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

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

$$\text{Var}_\lambda(P_k^{(n)}) = \frac{1}{n} \left[\mathbb{P}_\lambda(D_1 = k) - \mathbb{P}_\lambda(D_1 = k)^2 \right. \\ \left. + \frac{n-1}{n} \left[\mathbb{P}_\lambda(D_1 = D_2 = k) - \mathbb{P}_\lambda(D_1 = k)^2 \right] \right.$$

Main Theorem

Let, $X_1, X_2 \overset{i.i.d}{\sim} Bin(n-2, \lambda/n)$, and $I_1, I_2 \overset{i.i.d}{\sim} Ber(\lambda/n)$. Then, $(D_1, D_2) \overset{d}{=} (X_1 + I_1, X_2 + I_1)$ while $(X_1 + I_1, X_2 + I_2)$ are two independent copies of D_1.Thus,

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

Using the above coupling, we get,

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

Main Theorem

Let, $X_1, X_2 \overset{i.i.d}{\sim} Bin(n-2, \lambda/n)$, and $I_1, I_2 \overset{i.i.d}{\sim} Ber(\lambda/n)$. Then, $(D_1, D_2) \overset{d}{=} (X_1 + I_1, X_2 + I_1)$ while $(X_1 + I_1, X_2 + I_2)$ are two independent copies of D_1.Thus,

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

Using the above coupling, we get,

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

$$\mathbb{P}_\lambda\left(\max_{k\geq 0}\left|P_k^{(n)} - \mathbb{E}_\lambda\left[\mathbb{P}_k^{(n)}\right]\right| \geq \frac{\epsilon_n}{2}\right)$$
$$\leq \frac{4}{\epsilon_n^2}\sum_{k\geq 0}\left[\frac{1}{n}\mathbb{P}_\lambda(D_1 = k) + \frac{\lambda}{n}\mathbb{P}_\lambda(X_1 = k) + \mathbb{P}_\lambda(X_2 = k-1)\right]$$
$$= \frac{4(2\lambda + 1)}{\epsilon_n^2 n} \to 0 \qquad \text{since} \quad n\epsilon_n^2 \to \infty$$

Coupling

Coupling is nothing but a joint distribution of random variables that may not be individually defined on the same probability space having the same marginal distribution.

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 $\mathcal{E} \in \mathbb{R}$,

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

Coupling

Coupling is nothing but a joint distribution of random variables that may not be individually defined on the same probability space having the same marginal distribution.

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 $\mathcal{E} \in \mathbb{R}$,

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

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

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

Total Variation Distance

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

Total variation distance

For two probability measures μ and ν, the total variation distance between them is defined as $d_{TV}(\mu,\nu) = \sup_{A\subseteq\mathbb{R}} |\mu(A) - \nu(A)|$

Total Variation Distance

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

Total variation distance

For two probability measures μ and ν, the total variation distance between them is defined as $d_{TV}(\mu, \nu) = \sup_{A \subseteq \mathbb{R}} |\mu(A) - \nu(A)|$

For discrete probability mass functions,the total variation distance between them will be

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

Total Variation Distance

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

Total variation distance

For two probability measures μ and ν, the total variation distance between them is defined as $d_{TV}(\mu, \nu) = \sup_{A \subseteq \mathbb{R}} |\mu(A) - \nu(A)|$

For discrete probability mass functions,the total variation distance between them will be

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

For continuous random variables, the total variation distance will be,

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

Useful Theorems

These are the theorems we have used previously in our proof earlier.

Useful Theorems

These are the theorems we have used previously in our proof earlier.

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

Useful Theorems

These are the theorems we have used previously in our proof earlier.

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

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

Reference

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

Extra Slides

Extra Slides for Convenience

Why Enough?

From triangle inequality, it follows that,

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

Thus, it is enough to show that,

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

The bound on variance

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

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,

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

Proof of maximal coupling theorem

We start by defining the coupling that achieves the equality.

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

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

First of all, observe that,

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

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,

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

Proof of maximal coupling theorem

and also,

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

which implies that,

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

$$\implies \mathbb{P}(\hat{X} = \hat{Y}) = \sum_x \mathbb{P}(\hat{X} = \hat{Y} = x) \leq \sum_x (p_x \wedge q_x)$$

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

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

Proof of poisson limit for binomial RVs

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

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

Proof of poisson limit for binomial RVs

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.