MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Lecture 6 Mar 1, SP17
Lecturer: Guy Bresler Scribe notes by Suhas S Kowshik

Coupling and giant component in random graphs

Content.

1. Main theorem
2. Coupling of random processes
3. Strassen's theorem
4. Stochastic domination
5. Idea of the proof of the main theorem

1 Main theorem

Theorem 1 (Erdős-Rényi). *Let $G(n, p = \frac{\lambda}{n})$ be an Erdős-Rényi random graph. For $v \in [n]$ let $C(v)$ denote the component containing v. Let $L_1 = \max_{v \in [n]} |C(v)|$ be the size of the largest component of G and let L_2 be the size of the second largest component. Then*

$$\lambda < 1 \Rightarrow L_1 = \Theta(\log(n)) \text{ (Sub critical)}$$
$$\lambda > 1 \Rightarrow L_1 \sim c_\lambda n,\ L_2 = \Theta(\log(n)) \text{ (Super critical)}$$

where $c_\lambda > 0$ satisfies $1 - c = e^{-c\lambda}$.

An idea of the proof of this theorem is given towards the end.

2 Coupling of stochastic processes

Definition 1 (Coupling). *Random variables $(\hat{X}_1, \hat{X}_2, ..., \hat{X}_n)$ defined on same probability space, is a coupling of the random variables $X_1, X_2, ..., X_n$ if $\hat{X}_i \stackrel{d}{=} X_i, i \in [n]$.*

Note that the original variables do not need to be defined on the same probability space.

2.1 Examples

1. Let $X \sim \mathrm{Ber}(p)$ and $Y \sim \mathrm{Ber}(q)$ with $q > p$. Let $U \sim \mathrm{Unif}[0,1]$. Define $\hat{X} = 1_{U \le p}$ and $\hat{Y} = 1_{U \le q}$. Clearly $\hat{X} \overset{d}{=} X$ and $\hat{Y} \overset{d}{=} Y$. Also

$$\mathbb{P}(\hat{X} \le \hat{Y}) = 1.$$

   This is called "Monotone coupling".

2. For $i \in \mathbb{N}$, let $X_i \sim \mathrm{Ber}(p)$, $Y_i \sim \mathrm{Ber}(q)$ be iid random variables such that (X_i, Y_i) are coupled as in the previous example with $q > p$. Let $S_n^{(p)} = \sum_{i=1}^n (2X_i - 1)$ and $S_n^{(q)} = \sum_{i=1}^n (2Y_i - 1)$. With $S_0^{(p)} = 0 = S_0^q$, the two random walks are such that

$$S_n^{(p)} \le S_n^{(q)} \ a.s \ \forall n$$

3 A special coupling and Strassen's theorem

Given two random variables X and Y, our goal is to maximize $\mathbb{P}[X = Y]$. Let $\mathbb{P}[X = x] = p_x$ and $\mathbb{P}[Y = x] = q_x$. Let $d_{TV}(X,Y) = \sup_{A \in \mathcal{F}}(\mathbb{P}[X \in A] - \mathbb{P}[Y \in A])$ be the total variation distance between X and Y. We have the following theorem.

Theorem 2 (Strassen). *$\mathbb{P}[X \neq Y] \ge d_{TV}(X,Y)$ for any coupling (X,Y). Further, there exists a coupling (X,Y) such that $\mathbb{P}[X \neq Y] = d_{TV}(X,Y)$.*

Proof. For any $A \in \mathcal{F}$, we have

$$\begin{aligned}
\mathbb{P}[X \in A] - \mathbb{P}[Y \in A] = & \\
& \mathbb{P}[X \in A, X = Y] + \mathbb{P}[X \in A, X \neq Y] \\
& - \mathbb{P}[Y \in A, X = Y] - \mathbb{P}[Y \in A, X \neq Y] \\
& \le \mathbb{P}[X \in A, X \neq Y] \\
& \le \mathbb{P}[X \neq Y].
\end{aligned}$$

Taking supremum over $A \in \mathcal{F}$ proves the first part of the theorem. We know that, for discrete random variables X and Y, $d_{TV}(X,Y) = \frac{1}{2}\sum_x |p_x - q_x|$. Let $A = \{x : p_x > q_x\}$. Then $d_{TV}(X,Y) = \mathbb{P}[X \in A] - \mathbb{P}[Y \in A]$. To prove the second part, it is enough to show that $\mathbb{P}[X \in A] - \mathbb{P}[Y \in A] = \mathbb{P}[X \neq Y]$. That is

$$\mathbb{P}[X \in A, X \neq Y] - \mathbb{P}[Y \in A, X \neq Y] = \mathbb{P}[X \neq Y].$$

So it suffices to find a coupling such that

$$\begin{aligned} X \notin A &\Rightarrow X = Y \\ Y \in A &\Rightarrow X = Y. \end{aligned}$$

One can verify that the following joint distribution on (X, Y) is a valid coupling and has the above required property

$$\mathbb{P}[X = x, Y = y] = \begin{cases} \min\{p_x, q_x\} & x = y \\ \frac{\max\{p_x - q_x, 0\} \cdot \max\{q_y - p_y, 0\}}{d_{TV}(X,Y)} & x \neq y \end{cases}$$

□

4 Poisson approximation

Lemma 3. *If* $X \sim \mathrm{Ber}(p)$, $Y \sim \mathrm{Poisson}(p)$ *then* $d_{TV}(X, Y) \leq p^2$

Proof. We use the optimal coupling to prove this lemma. Let

$$\begin{aligned} X = Y = 0 & \quad w.p. \quad p \\ X = Y = 1 & \quad w.p. \quad pe^{-p} \\ X \neq Y & \quad w.p. \quad (1 - p - pe^{-p}). \end{aligned}$$

So

$$P[X \neq Y] = 1 - (1 - p) - pe^{-p} = p(1 - e^{-p}) \leq p^2$$

since $1 - e^{-p} \leq p$. □

Theorem 4. *Let* $X_i \sim \mathrm{Ber}(p_i)$, $S_n = \sum_{i=1}^n X_i$. *Let* $Z_n \sim \mathrm{Poisson}(\sum_{i=1}^n p_i)$. *Then*

$$d_{TV}(S_n, Z_n) \leq \sum_{i=1}^n p_i^2.$$

Hence

$$d_{TV}(\mathrm{Binom}(n, \frac{\lambda}{n}), \mathrm{Poisson}(\lambda)) \leq \frac{\lambda^2}{n}.$$

Proof. Let $Y_i = \mathrm{Poisson(p_i)}$. Then $Z_n \stackrel{d}{=} \sum_{i=1}^{n} Y_i$. Couple X_i and Y_i optimally as in lemma 3 for every i. Therefore

$$\begin{aligned} d_{TV}(S_n, Z_n) &\leq \mathbb{P}[S_n \neq Z_n] \\ &\leq \sum_{i=1}^{n} \mathbb{P}[X_i \neq Y_i] \leq \sum_{i=1}^{n} p_i^2. \end{aligned}$$

□

5 Stochastic domination

Definition 2 (Stochastic domination)**.** *Given two random variables X and Y, X is said to have stochastic dominance over Y, denoted $X \succeq Y$, if $\mathbb{P}[X \geq x] \geq \mathbb{P}[Y \geq x], \forall x$*

Example : For $\mu \geq \lambda$, we have $\mathrm{Poisson}(\mu) \succeq \mathrm{Poisson}(\lambda)$. To see this, let $X \sim \mathrm{Poisson}(\lambda)$ and $Z \sim \mathrm{Poisson}(\mu - \lambda)$. Let $Y = X + Z$. Then $Y \sim \mathrm{Poisson}(\mu)$. Hence

$$\mathbb{P}[Y \geq x] = \mathbb{P}[X + Z \geq x] \geq \mathbb{P}[X \geq x].$$

Theorem 5. *$X \succeq Y$ iff there exists a coupling (X, Y) such that $P(X \geq Y) = 1$.*

Proof. If there exists such a coupling, then

$$\mathbb{P}[Y \geq x] = \mathbb{P}[X \geq Y \geq x] \leq \mathbb{P}[X \geq x].$$

For the converse the idea is similar to monotone coupling using the generalized inverse CDF on $\mathrm{Unif}[0, 1]$ random variable. □

Example : Let $m \geq n$ and $p \geq q$. Then $\mathrm{Binom}(m, p) \succeq \mathrm{Binom}(n, q)$. To prove this one think of binomial random variable as a sum of independent Bernoulli random variables. Use monotone coupling to couple n $\mathrm{Ber}(p)$ variables (corresponding to $\mathrm{Binom}(m, p)$) to n $\mathrm{Ber}(q)$ variables. Then we are just adding more non-negative stuff in $\mathrm{Binom}(m, p)$ compared to $\mathrm{Binom}(n, q)$.

6 Proof idea of the main theorem

There are four key steps to proving the main theorem:

1. Define the graph exploration process
2. Relate to a branching process
3. Analyze the Poisson branching process
4. Apply to the random graph in the theorem

6.1 Graph exploration process

In a given graph G, we can define a procedure to find the component $C(v)$ containing the vertex v. Similar to the random walk representation of branching processes, we classify the vertices as active, explored or neutral. Start from v. It is active and all other vertices are neutral at $t = 0$. $Y_0 = 1$. At each time t, take an active vertex and make all its neutral neighbours active. Set that vertex as explored. Let Y_t denote the new number of active vertices (just after time t). The process terminates when there are no more active vertices. Hence $C(v)$ is the set of explored vertices. If Z_t denotes the number of neutral neighbours of an active vertex that is chosen to be explored, then

$$Y_t = Y_{t-1} + Z_t - 1$$

with $Y_0 = 1$.

In $G = G(n, p)$, each neutral w has independent probability p of becoming active. Let $N_t = n - t - Y_t$ denote the number of neutral vertices at time t. Then it can be seen that

$$Z_t \sim \mathrm{Binom}(N_{t-1}, p).$$

If $T = \inf\{t : Y_t = 0\}$, then $T = |C(v)|$. Since $N_t = N_{t-1} - Z_t$ it follows that $N_t \sim \mathrm{Binom(N_{t-1}, 1-p)}$. By induction, we have

$$N_t \sim \mathrm{Binom}(n - 1, (1 - p)^t).$$

The random quantities Y_t, Z_t and N_t can be compared to the corresponding quantities of the Poisson branching process.

The next step is to prove the following inequalities:

$$\mathbb{P}[T^{gr}_{n,p} \geq t] \leq \mathbb{P}[T^{\mathrm{Binom}}_{n,p} \geq t] \quad \text{(Stochastic domination)}$$
$$\mathbb{P}[T^{gr}_{n,p} \geq t] \geq \mathbb{P}[T^{\mathrm{Binom}}_{n-1,p} \geq t]$$

where $T^{gr}_{n,p}$ is the component size in the graph exploration process and $T^{\mathrm{Binom}}_{n,p}$ is the total progeny is the branching process with offspring distribution $\mathrm{Binom}(n, p)$. These binomial quantities are compared to Poisson to yield various estimates. Further details can be found in [Alon-Spencer].