Lecture 6 (chapter 3): Connectivity of ER graph

- ~~So far~~ last lecture: 어떤 조건에서 giant component 가 나오는가?
- (Question) 언제 ER graph가 connected(연결) 되나? (전체 node가 모두 연결)
  - ↳ random graph : 확률적인 argument
  - → 언제(어떤 조건에서) 굉장히 높은 확률로 "connected"
- (Intuition) $\lambda = np > 1$: λ가 1보다 큰 상수라면 giant component → connected : 훨씬 더 challenging
  - $p = \theta(\frac{1}{n})$, λ: 상수 $= 1$, < 1, > 1
  - λ: 상수 ↑ n에 증가 order가 필요
  - → $\lambda = \log\log n$? $\log n$? n^{α}? ($\alpha < 1$)
  - ↓ $p = \frac{\log\log n}{n}$, $p = \frac{\log n}{n}$, $p = \frac{n^{\alpha}}{n}$ — ?????

(접근방법) 수학적, 추상적 → "connected 되어 있다"

① $C(v)$ (임의의 노드) $P(|C(v)| = n) \longrightarrow 1$

⇕

isolation 된 node가 없다 : 훨씬 쉽게 잘 할 수 있다.

Let X be the R.V of # of isolated nodes

(관심) $P(X = \text{작은 수})$ $P(X \to 0)$??? → connected 된다.

Let R.V

$$I_v = \begin{cases} 1 & \text{if isolated} \\ 0 & \text{otherwise} \end{cases}$$

$X = \sum_v I_v$ → $\sum$ Bernoulli, $\text{Bin}(n, (1-p)^{n-1})$

$\{I_v\}$: identical, independent ??? (✗) dependent

$$\Pr(I_v = 1) = (1-p)^{n-1}$$

질문): $\sum$ Bernoulli, dependent 한 것을 다룰 수 있는 tool?

$B(n, p)$ → poisson

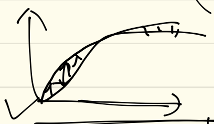
↳ Googling : "Bernoulli dependent approximation, poisson"

찾은 → "Stein-Chen method" : dependent 하는 Bernoulli RV의 합과 poisson 분포와의 거리

$$d\Big(\underbrace{\sum \text{Bernoulli}_i}_{\text{R.V}}^{\text{dependent}},\ \underbrace{\text{Poisson}(\lambda)}_{\text{R.V}}\Big) < \epsilon$$

⇒ 두개의 RV를 (두개의 distribution) 간의 거리를 개념화하는 수학이 필요할것같다

→ 우리 용도에 맞게



⇒ **Total variation** : 2개의 distribution 간의 distance metric

(수리순서) (접근 방법) ⇒ 거리순의 차이 ~을 정리 → 증명, ... 수학에 매몰(X) tool

X : 확률변수

$X = \sum_{v \in N} I_v$, (I_v is Bernoulli RV $((1-p)^{n-1})$, I_v들은 dependent하다) → stein-chen method

만약 independent 하다면 $Bin(n, (1-p)^{n-1}) \xrightarrow{n\to\infty} Poisson(\lambda)$, $\lambda = n\cdot(1-p)^{n-1}$

→ dependent

X ⟷ (가까울것같다, stein-chen method) $Poisson(\lambda)$, $\lambda = n\cdot(1-p)^{n-1}$ ⟷ (더 같게될것 있을것같다) $Poisson(■)$

p를 적당히 잘선택하면 $n\cdot(1-p)^{n-1} \xrightarrow{n\to\infty} ■$ → handling

가깝다

$X \geq 0$ → 일정하게 작게

(Next) $n(1-p)^{n-1} \xrightarrow{n\to\infty}$ 상수로 보내는 p의 선택은? $p = □$

$p = f(n)$ 이라고 하자. $n\cdot p = f(n)$

$n(1-f(n))^{n-1} \xrightarrow{n\to\infty} □$

$n\left(1 - \frac{f(n)}{n}\right)^{n-1} \xrightarrow{n\to} □$

(i) $f(n)$이 linear order $np = \Theta(n)$ → 큰 의미가 없다. $\left(1-\frac{1}{n}\right)^n \to \frac{1}{e}$

(ii) sublinear 하면 $f(n)$ 어떻게 되나 (?) $\left(1+\frac{1}{n}\right)^n \to e$

$$n\left(1-\frac{f(n)}{n}\right)^n \approx e^{\log n}\left(\left(1-\frac{f(n)}{n}\right)^{\frac{n}{f(n)}}\right)^{f(n)} \approx e^{\log n}\left(e^{-1}\right)^{f(n)} = e^{\log n - f(n)}$$

$f(n) = \log n$ $\quad \lambda = n\cdot(1-p)^{n-1} \to 1$ $\quad$ Poisson(1)

$f(n) = \log n + c$ $\quad \lambda = n\cdot(1-p)^{n-1} \to e^{-c}$, $X \to$ Poisson(e^{-c})

c는 충분히 크게 잡으면 Pr(isolation node의 개수 $= x$) ~ Poisson(e^{-c}) → 충분히 작아진다.

(Thm): If $\lambda = np = \log n + c$ for a constant c,

$X \sim$ Poisson(e^{-c}) →

다음 lecture: 지금까지의 intuition → 수학으로 정리

Lecture (Part I)   Convergence of R.Vs

- 4 definitions (a.s. in prob. m.s in distribution)

- A random variable X is a function (random) on Ω (→ sample space) for some probability space $(\Omega, \mathcal{F}, P)$

  - Ω ⇓ sample space
  - $\mathcal{F}$ ⇓ σ-algebra "set of events"
  - P → probability measure

event = set $\subset \Omega$

(Ex) Fair Dice $\Omega = \{1, 2, \ldots, 6\}$ (each $\frac{1}{6}$)

$X(\omega) = 1$ if ω is even
$\quad\;\; = 0$ otherwise

$X(1) = 0$, $X(2) = 1$, …, $X(6) = 1$ (집합이 존재)

$\emptyset, \Omega, \{1,2\}, \{1,3,\omega\}$

"event that $X=1$" $= \{\omega \mid X(\omega) = 1\} = \{2,4,6\}$

$$\Pr(\text{event}) = \Pr(X=1) = \Pr(\{\omega \mid X(\omega)=1\}) = \frac{1}{2}$$

- Interest: Given a seq. of random variable $\{X_n\}_{n \ge 0}^{\infty}$, talk about "Convergence"

$X_n(\omega) \longrightarrow X(\omega)$ …… "어떤 sense" (mode)

(Deterministic sequence) $a_n \xrightarrow{\infty} a \iff \forall \varepsilon > 0, \exists N$ s.t. whenever $n \ge N$, $|a_n - a| < \varepsilon$.

$\frac{1}{n} \xrightarrow{n \to \infty} 0$

$a_1, a_2, \ldots \ldots a_n\, a_{n+1} \ldots$

Def 1 (Almost sure convergence)

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1 \iff \Pr\left(X_n \xrightarrow{n\to\infty} X\right) = 1$$

$X_n \xrightarrow{n\to\infty} X$ a.s. (almost sure) 거의 확실히

$$\Pr\left(\left\{\omega \in \Omega \;\middle|\; \lim_{n\to\infty} X_n(\omega) \longrightarrow X(\omega)\right\}\right) = 1$$

(pointwise convergence가 만족되는 놈들이 전체 Ω 일 때)

Almost sure conv.를 test 하는 방법

X_n X,

① w를 하나 가져온다. $X_n(w)$ $X(w)$

deterministic

② check whether $X_n(w) \xrightarrow{n\to\infty} X(w)$ 수렴하는지 본다.

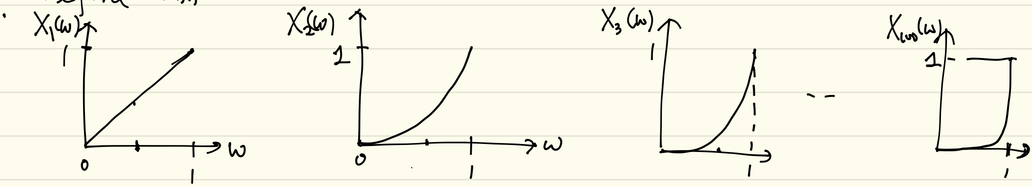
③ If yes, 그 "w"를 챙겨둔다.

④

A.S convergence

(Example) X_n on $\Omega=[0,1]$ $X=0$

Define $X_n(w)=w^n$



(Question) $X_n \longrightarrow X$

$\longrightarrow 0$ a.s. (?)

$Pr([0,1))$ = $Pr([0,1])$

w를 선택 $w=\frac{1}{2}$ $X_n(\frac{1}{2}) \xrightarrow{n\to\infty} X(\frac{1}{2})$ (Yes)
$= 0$

$w=\frac{1}{3}$

$w=\frac{3}{4}$

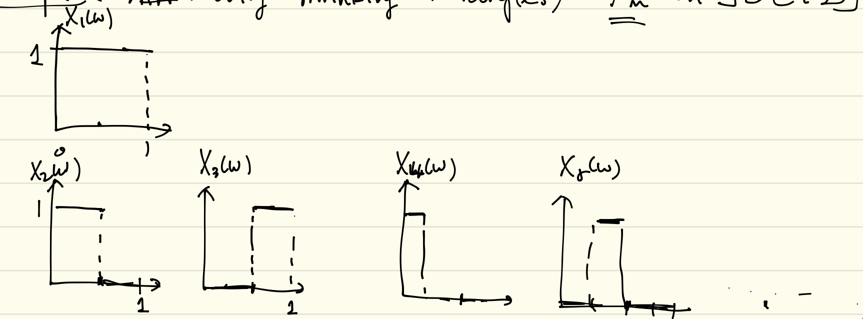
$w=1$ $X_n(1) \longrightarrow 1$

$Pr(\{w \mid X_n(w) \to 0\})$

$= Pr([0,1)) = 1$

$\therefore X_n \to X$ a.s.

Example (Moving shrinking rectangles) X_n on $\Omega=[0,1]$



$X_n \longrightarrow 0$ a.s (Yes / NO)

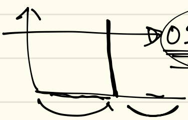
$\omega=\frac{1}{2}$ $X_n(\frac{1}{2}) \xrightarrow{n\to\infty} 0$

$\omega=\frac{1}{3}$ (X)

↳ 0과 1이 번갈아 나오므로: converge 할 수 없다

for large n, $P\{X_n=0\}$

$n=100$만
$n=10$억



0으로 가야함. ⇒ 다른 convergence concept

weaker 한 concept

(Def 2) (Convergence in Probability) for any $\varepsilon>0$

$$\lim_{n\to\infty} P_r\{|X_n-X|\geq\varepsilon\} = 0$$

$$\Leftrightarrow \lim_{n\to\infty} P_r\{|X_n-X|\leq\varepsilon\} \to 1$$

$\Leftrightarrow$ $|X_n-X|$ is small with high probability

pointwise로 converge 하는 애들이 "전부"

almost sure

$$P_r(|X_n-X|\leq\varepsilon) = \left(1-\frac{1}{n}\right) \longrightarrow 1$$

두개의 차이점 $n\to\infty$

Convergence in Probability

① n들 ___ ② $|X_n-X|\geq\varepsilon$, $|X_n-X|\leq\varepsilon$ 교차 ③ $P_r(|X_n-X|\geq\varepsilon)$ ④ $n\to\infty$

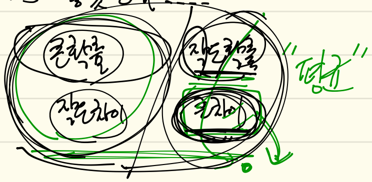
A.S $\xrightarrow{\ O\ }$ in Probability (증명)
A.S $\xleftarrow{\ X\ }$ in Probability

But $|X_n - X|$ 가 큰경우도 존재할수있다. (매우 작은 확률) ⇒ 문제가 안된다

conv. in probability

$X_n \xrightarrow{in\ P_r} X$

↳ "어쩌" 나는 $|X_n - X|$가 큰경우가 (너무) 안나왔으면 좋겠는데....



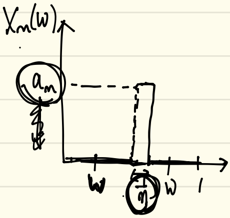
Def 3 (mean square convergence)

$E(X_n^2) < \infty$ for all n,

$$\lim_{n\to\infty} E\left((X_n - X)^2\right) = 0 \qquad X_n \xrightarrow{ms} X$$

AS →	in prob
MS →	in prob

Example (Another shrinking Rectangles) X_n on $\Omega = [0,1]$



(i) $X_n \xrightarrow{a.s} 0$ if $a_n \xrightarrow{n\to\infty} 0$ (a_n의 조건)   ω 선택 Ω $X_n(\omega) \xrightarrow{n\to\infty} 0$

$a_n = \text{constant}$

(ii) $a_n \not\xrightarrow{n\to\infty} 0$ $\exists k$ $k>0$, s.t $a_n \geq k$ ⇒ $\lim X_n(\omega)$ does not exist for any $\omega \in \Omega$

almost sure(X)

$X_n \xrightarrow{ms} 0$ $X_n \to 0$ in probability $\forall \varepsilon > 0$ (n 증가하면) $P_r(X_n \leq \varepsilon) \to 1$

$\leq 1 - \frac{1}{n} \to 1$

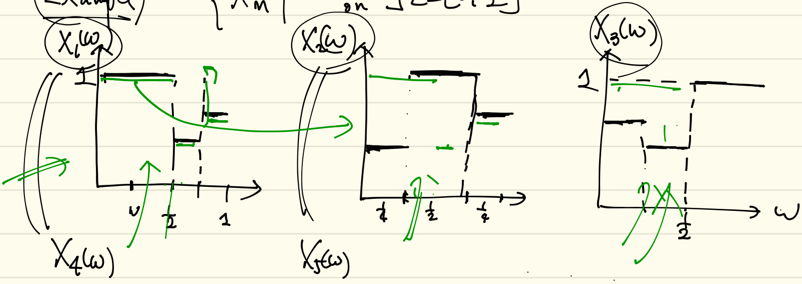
(iii) $E[|X_n - 0|^2] \to 0$ (?) $\frac{a_n^2}{n} \to 0$

큰차이는 작은 확률

(ex2) $a_n = \log n$ Yes

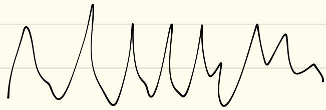
$a_n = \sqrt{n}$ No

(Example) $\{X_n\}$ on $\Omega = [0, 1]$



$X_{n+3} = X_n$ for all $n \geq 1$

① a.s convergence to some random variable (X)



② in prob. convergence to some random variable ③ m.s convergence

(작은 차이를 나타내는 확률이 high probability)

However, Consider some random variable X with the following distribution

$$\Pr(X=1) = \frac{1}{2}, \quad \Pr\left(X=\frac{3}{4}\right) = \frac{1}{4} \quad \Pr\left(X=\frac{1}{2}\right) = \frac{1}{4}$$

check?) X_n의 분포 $=\!=$ X의 분포

$X_n \xrightarrow{n\to\infty} X$ (분포) (weak)

(Def 4) (convergence in distribution). $\forall x$ $F_{X_n}(x) \to F_X(x)$ (CDF)

for all continuity point x (skip here)

$\Updownarrow$ $X_n \xrightarrow{\text{in distribution}} X$

$$E[f(X_n)] \xrightarrow{n\to\infty} E[f(X)] \text{ for all bounded continuous function } f.$$

(Page 37) Def 3.3 A seq of probability (measures / law) distribution $\{\mu_n\}_{n\geq 0}$ is said to "weakly converge" to μ_∞ for all bounded continuous function

$(\Omega, \mathcal{F}, P)$

$$\lim_{n\to\infty} \int_\Omega f(\omega)\, \mu_n(d\omega) = \int_\Omega f(\omega)\, \mu_\infty(d\omega)$$

"convergence in distribution" = "weak convergence in probability measure"

convergence : 주어진 distance metric

$F_{X_n}(x) \longrightarrow F_X(x)$

$a_n \to a \qquad |a_n - a| < \varepsilon$

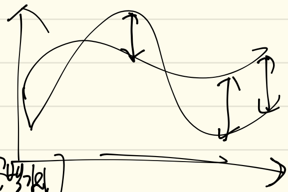
distribution들을 대상으로 하는 convergence

"convergence in distribution" ↳ "어떤 distance metric" 에서 convergence

ㅋㅋㅋ?

$d(F_{X_n}, F_X) \xrightarrow{n\to\infty} 0$

거리 — 두 함수간의 거리



distance metric

Total variation (Def 3.1)

$(\Omega, \mathcal{F}, P)$

$d_{var}(\mu_1, \mu_2) = 2 \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)|$

모든 event의 집합

Prop 3.4 total variation을 metric으로 사용하여 converge를 하면

$\Rightarrow$ convergence in distribution.

$$\left(\begin{array}{l} A.S \\ M.S \end{array} \right) \Rightarrow \text{in Prob} \Longrightarrow \text{in distribution}$$

under what condition (O)

상황에 맞게 증명

$\Rightarrow$ 오늘 강의 내용은 pdf upload

Lecture 6 (part 3) Stein-Chen method (Stein method)

→ Bernoulli RV의 sum ⟶ Poisson (Bernoulli: independent; convergence in distribution)

$Bin(n,p) \xrightarrow{n\to\infty} Poisson(\lambda)$, $np=\lambda$

Question) $|Bin(n,p) - Poisson(\lambda)| \le f(n)$

↳ 확장: not necessarily independent

$$\left|\sum_{i=1}^{n} Bernoulli - Poisson(\lambda)\right| \le f(n)$$

↳ 얼마나 dependent 한가?

Let $\{I_i\}_{1\le i\le n}$ be a sequence of Bernoulli random variables (not necessarily independent), with $Pr(I_i=1)=p_i$

$\lambda = \sum_{i=1}^{n} p_i$ $(= E(W))$, and $A \subset N$. $W=\sum_{i=1}^{n} I_i$

Our interest: $\left| Pr(W\in A) - Po_\lambda(A) = \sum_{k\in A} \frac{e^{-\lambda}\lambda^k}{k!} \right|$

Thm

$$\left| Pr(W\in A) - Po_\lambda(A) \right| \le 2\left(\frac{1-e^{-\lambda}}{\lambda}\right)\left(\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n}\sum_{j=1, j\ne i}^{n} Cov(I_i, I_j)\right)$$

(the double sum $\le$ ⬭; $Cov(I_i,I_j)$ → dependent)

easily check that $\frac{1-e^{-\lambda}}{\lambda} \le \min(1, \lambda)$

$Cov(X,Y)$ → X와 Y가 얼마나 correlated 되어있는가?

$Cov(X,Y) = E(XY) - E(X)\cdot E(Y)$

If we can find another sequence of random variables J_{ji} defined on the same probability space as I_i whose distribution given that $I_j=1$ is identical to I_i, i.e.,

$$\begin{pmatrix} Pr(J_{ji}=1) \triangleq P(I_i=1 \mid I_j=1) \\ Pr(J_{ji}=0) \triangleq P(I_i=1 \mid I_j=1) \end{pmatrix}$$

($I_j=1$일 때 given I_i)

"coupling"

This coupling

$J_{ji} \ge I_i$

의미: I_j가 일어났을 때 I_i 의 분포와 같다

$$\mathrm{Cov}(I_i, I_j) = E[I_i I_j] - E(I_i)\cdot E(I_j) = E[I_i I_j] - p_i p_j$$

$$= E(I_i \mid I_j = 1)\, p_j - E(I_i)\cdot p_j = p_j\big(E(I_i \mid I_j = 1) - E(I_i)\big)$$

$$= p_j\big(E[J_{ji} - I_i]\big) \geq 0$$

Stein-Chen bound $\leq 2\cdot\min(1, \lambda^{-1})\cdot\left(\sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} p_j\big(E[J_{ji} - I_i]\big)\right)$

(책에나온것) (증명) (생각)

Connectivity of ER graph

Thm 3.8 Assume $np = \log n + c$ for some $c > 0$. $\left(p \approx \frac{\log n}{n}\right)$

Then, the distribution of number of isolated nodes in $G(n,p)$ converges in distribution to $\mathrm{Poisson}(e^{-c})$.

denote by X

$$X \longrightarrow \mathrm{Poisson}(\lambda) \longrightarrow \mathrm{Poisson}(e^{-c})$$

Let $I_v = 1$ if v is isolated, 0 otherwise.

$X = \sum_{v=1}^{n} I_v$, $\lambda = E(X) = \sum_{v=1}^{n} E(I_v) = \sum_{v=1}^{n} P_r(I_v = 1) = n\cdot(1-p)^{n-1}$

(Bernoulli의 sum)

From triangle inequality of dvar,

$$d_{var}(X, P_{e^{-c}}) \leq d_{var}(P_{e^{-c}}, P_\lambda) + d_{var}(X, P_\lambda)$$

(ii) Lemma 3.5   (i) Stein-Chen method

(i) $d_{var}(X, P_\lambda) \to 0$

$d_{var}(X, P_\lambda) \leq 2\cdot \min(1, \lambda^{-1}) \left(\quad \right)$

$\to I_v = \prod_{w \neq v} (1 - \xi_{vw})$, where $\xi_{uw} = 1$ if the edge (u,w) is connected, $= 0$

↳ v가 isolation

$Pr(\xi_{u,w}) = p$

$- J_{vw} \overset{d}{=} I_w \mid I_v = 1$

$= \prod_{u \neq v,w} (1 - \xi_{uw})$

coupling $J_{vw} \geq I_w$

$2 \cdot \min(1, \lambda^{-1}) \cdot \left(\sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} p_j \left(E[J_{ji} - I_j] \right) \right) \leq 2 \left(p_i + \lambda \frac{p \to 0}{1-p \to 0} \right) \xrightarrow{n\to\infty} 0$

$Pr(I_i = 1) \to e^{-c}$, $p_i \to 0$

Homework ① ⟸ 생략

$p = \frac{\log n + c}{n}$

$\lambda = n \cdot (1-p)^{n-1} \to e^{-c}$

$= 2 \left((1-p)^{n-1} + \lambda \cdot \frac{\frac{\log n}{n}}{1 - \frac{\log n}{n}} \right)$

(ii) $d_{var}(P_{e^{-c}}, P_\lambda) \leq$ ~~$2(1 \wedge \lambda^{-1})$~~ $2 |e^{-c} - \lambda| \longrightarrow 0$

Homework ②

$Pr(G(n,p) = \text{connected}) \to 1$ 로 만들기 위한 조건 $np = \log n + c$

So far, X: # of isolated nodes → X의 분포 $\overset{n\to\infty}{\sim}$ Poisson(e^{-c})

$X = 0 \Leftrightarrow$ 至 size 2 component,

size 3 〃 ,

size n 〃

외톨이가 없다.

반대 → size 2 ... n, $\frac{n+1}{2}$ (connected) component가 존재 없고

$Pr \to 0$

$X = 0 \Leftrightarrow$ 그래프가 연결

$\lambda = np > 1 \rightarrow$ giant component 등장 ($\Theta(n)$, ~~연결~~ $O(\log n)$)

상수 $\lambda = np = \log n + c$ 연결성↑ 없어질 것 같다

$$\Pr(G(n,p) \text{ is connected}) = \Pr(X=0) = e^{-e^{-c}}$$

$X \sim \text{Poisson}(e^{-c})$ Poisson(λ) $e^{-\lambda}$

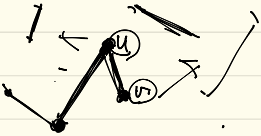
Need to show $\Pr(\text{size } 2, \dots \frac{n}{2} \text{ component}) \xrightarrow{n \to \infty} 0$

$p(A \cup B) \le p(A) + p(B)$

(Pf) (i) size = 2

$\Pr(\exists$ connected component of size 2)

$\le \binom{n}{2} \times \Pr($ ~~u와 v가~~ (u,v) size 2 component)



$\sum$(any two pair 연결 size 2 component)

$$= \frac{n(n-1)}{2} \cdot p\,(1-p)^{n-2} \cdot (1-p)^{n-2} = \frac{n(n-1)}{2} \cdot p \cdot (1-p)^{2(n-2)} \longrightarrow 0$$

$p = \frac{\log n + c}{n}$, $p \to 0$

$$\le \frac{\frac{n(n-1)}{2} \cdot \left(\frac{\log n + c}{n}\right) (1-p)^{2n}}{(1-p)^4} = \frac{\frac{n(n-1)}{2}\, p\, e^{-2np}}{(1-p)^4} \le \frac{n^2 \left(\frac{\log n + c}{n}\right) e^{-2n\frac{(\log n + c)}{n}}}{(1-p)^4}$$

$1-p \le e^{-p}$, $x \ge 0$

$np = \log n + c$

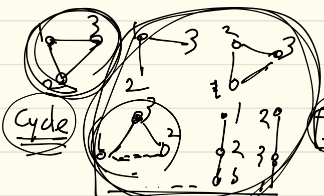
$e^{-2n\frac{(\log n+c)}{n}} = e^{-2\log n} \cdot e^{-2c}$, $e^{-2\log n} = \frac{1}{n^2}$

$$= \frac{n^2 \left(\frac{\log n + c}{n}\right) \frac{1}{n^2} e^{-2c}}{(1-p)^4} \longrightarrow 0$$

(ii) $\Pr(\text{size} = 3 \sim \frac{n}{2} \text{ exist})$

size r인 component

연결되는 pattern 다양



Cycle

Tree 형태로 연결 경우

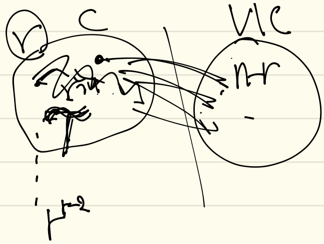
$p(A \cap B \cap C) \le p(A \cap B)$

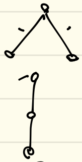
(Question) r개의 label이 달린 node 주어지고 만들 수 있는 tree의 수 몇 개?

(Thm) Cayley's Theorem: r^{r-2}. (증명 생략)

For any $r \in \{3, \dots, \frac{n}{2}\}$, and an arbitrary set C of r nodes,

$$Pr(C \text{ is connected}) \leq \sum_{T_C} Pr(\text{edges in } T_C \text{ present and no edge between } C \text{ and } V \setminus C)$$





$$\leq r^{r-2} p^{r-1} \left((1-p)^{n-r}\right)^r$$

$$Pr(\exists \text{ a connected component of size} \in \{3, 4, \dots, \frac{n}{2}\}) \qquad np = \lg n + c$$

$$\leq \sum_{r=3}^{\frac{n}{2}} \binom{n}{r} \cdot r^{r-2} p^{r-1} \left((1-p)^{n-r}\right)^r \xrightarrow{n\to\infty} 0$$

$$\binom{n}{r} \leq \frac{n^r}{r!}$$

$$\leq \cdot$$

\# "Stirling's formular"

$$r! \sim \sqrt{2\pi r}\,(r/e)^r$$

Homework

$$\longrightarrow 0$$

\# Homework은 강의 중간, 중간 Proof 완성:

명확하게 — 여러분의 note