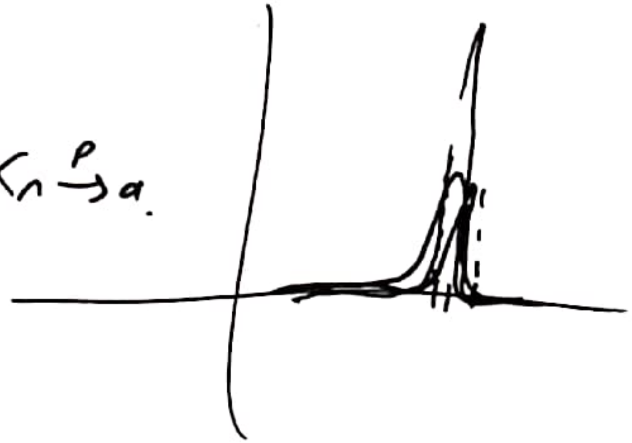
Defn Let (X_n) be a sequence of real-valued RVs and $a \in \mathbb{R}$ a constant. X_n converges to a in probability, $X_n \xrightarrow{P} a$ if $\forall \varepsilon > 0$ $\mathbb{P}(|X_n - a| \geq \varepsilon) \to 0$ as $n \to \infty$



Lemma 5.7 If $\mathbb{E}X_n \to a$ & $\mathbb{E}(X_n^2) \to a^2$ then $X_n \xrightarrow{P} a$.

Proof $Var(X_n) = \mathbb{E}X_n^2 - (\mathbb{E}(X_n))^2 \to a^2 - a^2 = 0$

Chebyshev $\mathbb{P}(|X_n - \mathbb{E}X_n| \geq \varepsilon/2) \leq \dfrac{Var(X_n)}{(\varepsilon/2)^2} \to 0$

Also for large n $|\mathbb{E}X_n - a| < \frac{\varepsilon}{2}$ $\square$

$Var = o(\mu^2) \Rightarrow \dfrac{X_n}{\mu_n} \xrightarrow{P} 1$

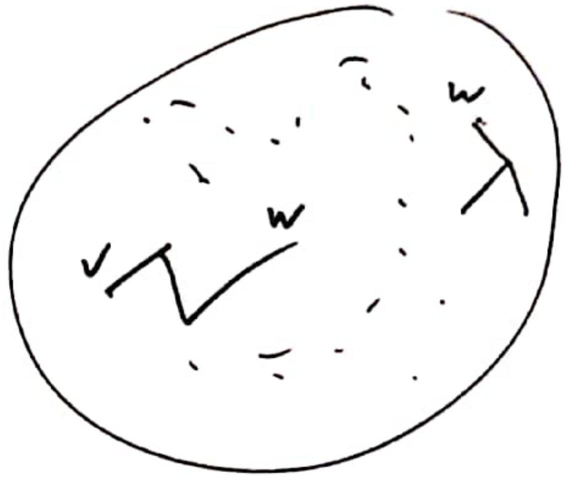
Lemma 5.8 Let $c>0$ and $k\geq 1$ be constant. Let $N_k = N_k(G(n, c/n))$.

Then $\frac{N_k}{n} \xrightarrow{p} \rho_k(c) = \mathbb{P}(|\mathcal{X}_{Po(c)}| = k)$.

Proof Shown $\mathbb{E}\left[\frac{N_k}{n}\right] \to \rho_k(c)$ already.

Recall $N_k = \sum_v I_v$ where $I_v = 1$ if $|C_v| = k$

Then $N_k^2 = \sum_v \sum_w I_v I_w = A + B$



where

$$A = \sum_v \sum_w I_v I_w 1_{\{C_v = C_w\}} \qquad \sum_v \sum_{w : C_w = C_v} I_v I_w$$

$$X+Y \qquad X(\omega)+Y(\omega)$$

$$B = \sum_v \sum_w I_v I_w 1_{\{C_v \neq C_w\}} \qquad \sum_{i\in R} X_i = \sum_i I_{\{i\in R\}} X_i$$

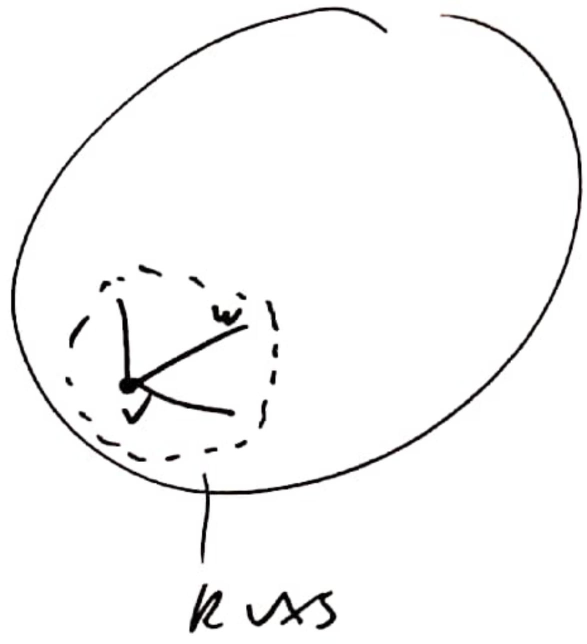
$$\sum_v I_v \sum_w I_w 1_{\{C_v = C_w\}}$$

If $I_v = 1$ then $\sum_w I_w 1_{\{C_v = C_w\}} = k$

(all k vertices in C_v count)

$\therefore A = \sum_v I_v k = k N_k \leq kn$

$\therefore \mathbb{E}[A] \leq kn.$



$$\mathbb{E}[B] = \mathbb{E}\Big[\sum_v I_v \sum_w I_w I_{\{C_v \neq C_w\}}\Big] \qquad \mathbb{E}[X] = \mathbb{P}(I=1)\,\mathbb{E}[X \mid I=1] + 0 \quad 0$$

$$= n\,\mathbb{E}\Big[I_v \sum_w I_w I_{\{C_v \neq C_w\}}\Big]$$

$$= n\,\underset{|C_v| = k}{\mathbb{P}(I_v = 1)}\;\mathbb{E}\Big[\underbrace{\sum_w I_w I_{\{C_v \neq C_w\}}}_{=\,N_k(G - C_v)} \;\Big|\; |C_v| = k\Big]$$

Given the cpt containing v, the edges outside C_v are still present indp with prob p, ie. $G - C_v \sim G(n', p)$ where $n' = n - |C_v|$.

Conditional on $|C_v| = k$ $\quad G - C_v \sim G(n-k, p)$. $\qquad p = \frac{c}{n}$

Now $(n-k)p \sim np = c$ so by ~~[illegible]~~ 18. Cor 5.6

$$\mathbb{E}\left[N_k(G - C_v) \mid |C_v| = k\right] = \mathbb{E}\left[N_k(G(n-k, p))\right]$$

$$\sim (n-k)\rho_k(c) \sim n\rho_k(c).$$

$$\therefore \mathbb{E}[B] \approx n\,\mathbb{P}(|C_v| = k)\, n\rho_k(c)$$

$$\sim n^2 \rho_k(c)^2$$

$$\mathbb{E}\left[\left(\frac{N_k}{n}\right)^2\right] = \frac{1}{n^2}\mathbb{E}[N_k^2] = \frac{1}{n^2}\big(\overset{o(n^2)}{\mathbb{E}[A]} + \mathbb{E}[B]\big) = \rho_k(c)^2 + o(1)$$

$$\to \rho_k(c)^2$$

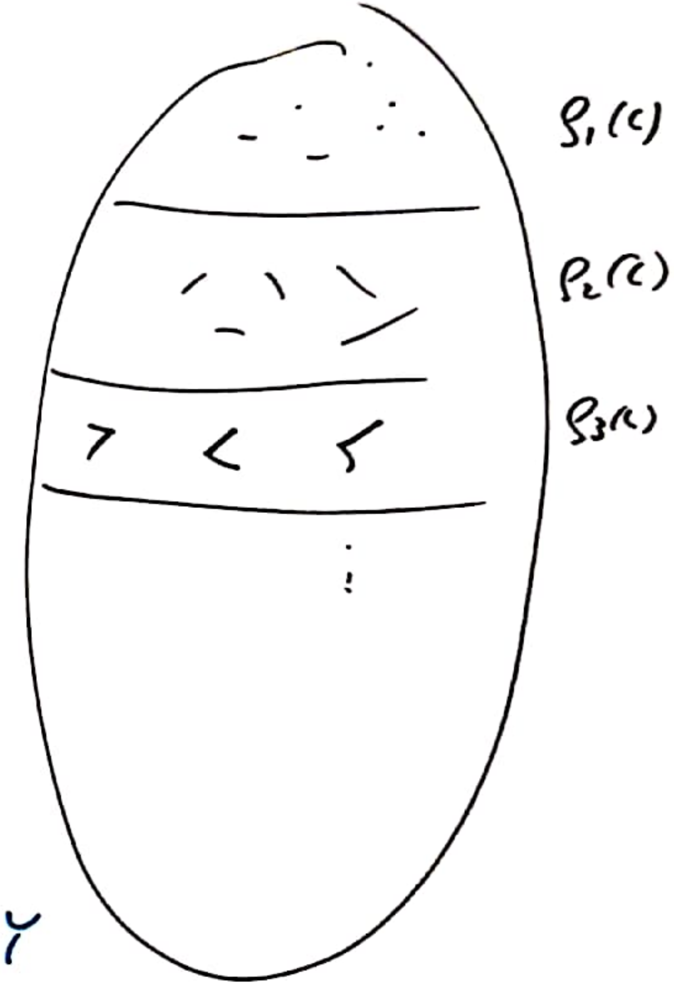
$$\therefore \frac{N_k}{n} \xrightarrow{p} \rho_k(c) \text{ by Lemma 5.7.}$$

For fixed k $\quad \frac{1}{n} N_k \xrightarrow{p} \rho_k(c)$

$N_k(G(n, \frac{c}{n}))$

$\forall k \, \forall \varepsilon \quad \exists n_0 \quad \forall n \geq n_0 \qquad \mathbb{P}\left(\left|\frac{1}{n} N_k - \rho_k(c)\right| > \varepsilon\right) \leq \varepsilon$

$n_0(k, \varepsilon)$



For fixed K. ALSO FOR $\omega(n) \to \infty$ SUFFICIENTLY SLOWLY

$$\mathbb{P}\left(\forall k \leq K_{\omega(n)} \quad \left|\frac{1}{n} N_k - \rho_k(c)\right| \leq \frac{\varepsilon}{K_{\omega(n)}} \right) \to 1$$

as $n \to \infty$

$\Downarrow$

$\left|\frac{1}{n} N_{\leq K} - \rho_{\leq K}(c)\right| < \varepsilon$

Let $N_{\leq K} = \sum_{k=1}^{K} N_k$ = # vxs in cpts size $\leq k$

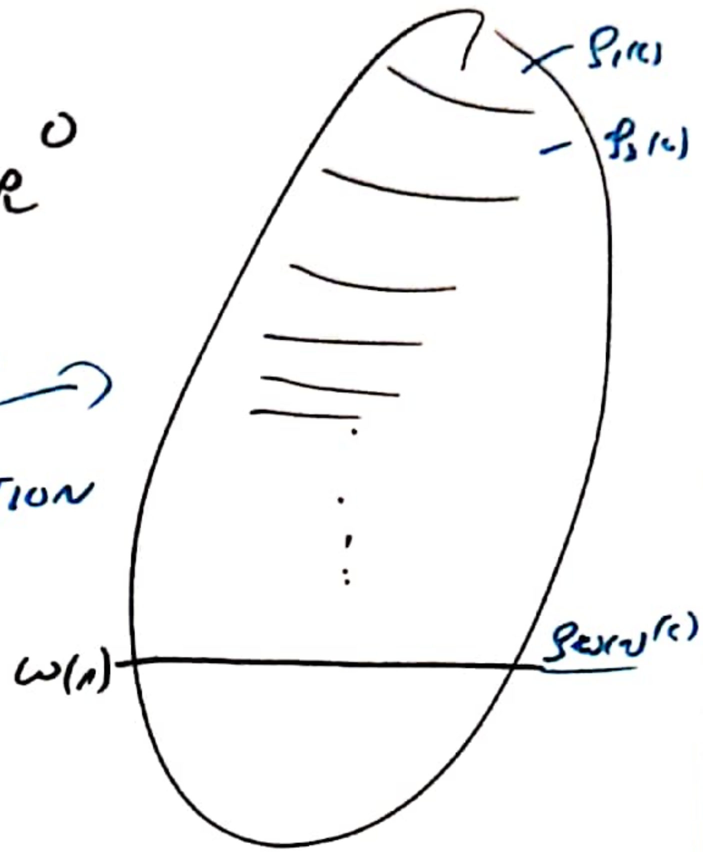
$\rho_{\leq K}(c) = \mathbb{P}(|X| \leq K)$

If $\omega(n) \to \infty$ slowly enough

$$\left| \frac{1}{n} N_{\le \omega(n)} - \rho_{\le \omega(n)}(c) \right| \xrightarrow{P} 0$$

$$\sum_{k=1}^{\omega(n)} \rho_k(c) \longrightarrow \sum_{k=1}^{\infty} \rho_k(c)$$

UNDERSTAND UP TO $o(1)$ FRACTION OF VERTICES



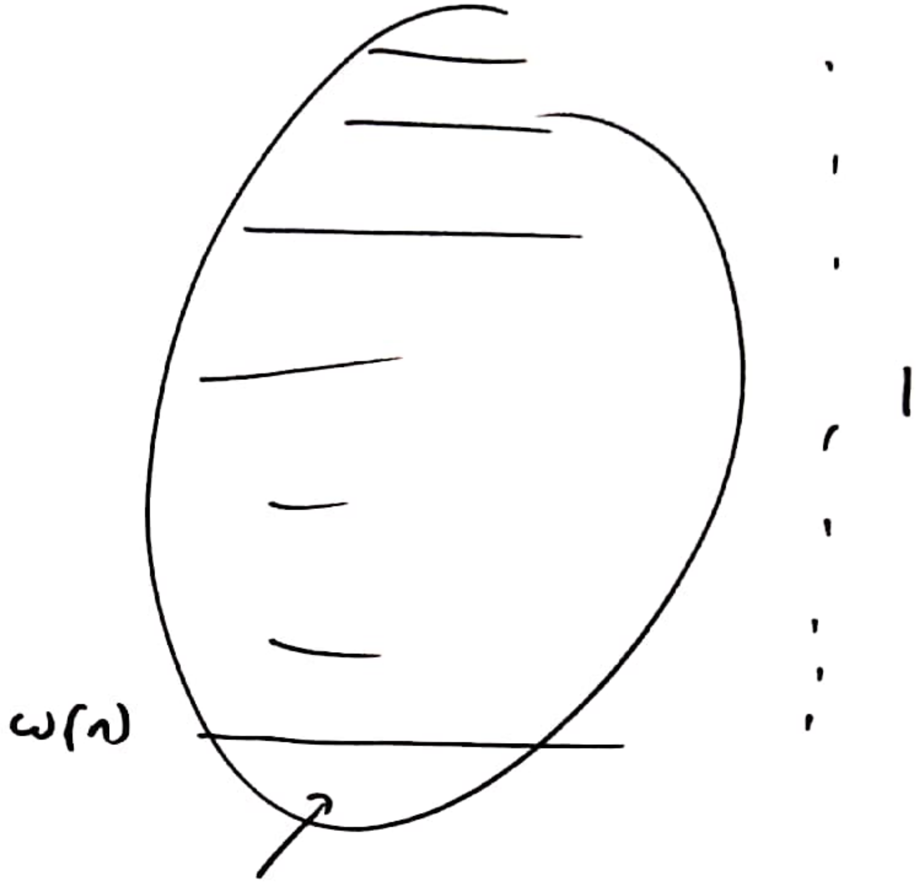
$$\frac{1}{n} N_{\le \omega(n)} \xrightarrow{P} \sum_{k=1}^{<\infty} \rho_k(c)$$

$$\text{RHS} = \sum_{k=1}^{<\infty} \mathbb{P}\left(|\underset{\sim}{X}_{Po(c)}| = k\right) = \mathbb{P}\left(|\underset{\sim}{X}| \text{ is finite}\right)$$

$$= \eta(c).$$

Case 1: $c \le 1$ $\quad \eta(c) = 1$

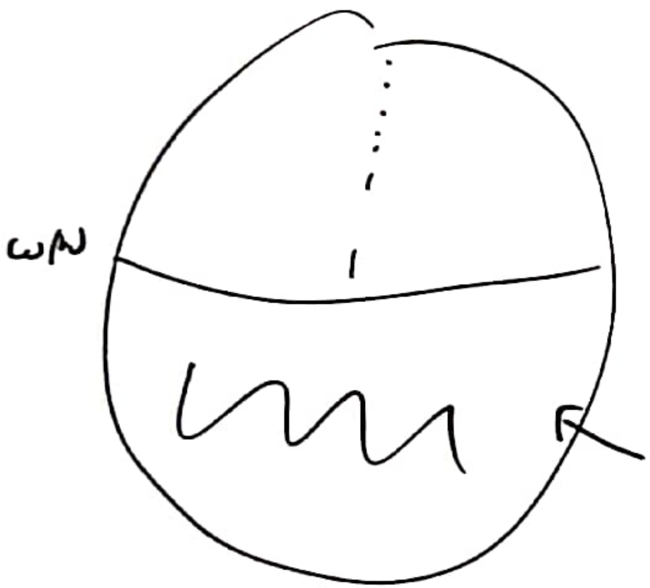
Understand component sizes for all but $o(n)$ vertices



$\frac{n}{\log n}$ $\quad$ $o(n)$ vertices in one cpt?

in lots?

<u>Case 2</u> $c > 1$ $\quad \eta(c) < 1$



$\approx (1-\eta(c))n$ vertices in ONE giant cpt?

All size $2\omega(n) \simeq 2\log n$?

Lemma 5.9. Let $c>0$ be constant, and suppose $k^- = k^-(n)$ satisfies $k^- \to \infty$ and $k^- \le n^{1/4}$. Then $N_{\le k^-} = N_{\le k^-}(G(n, \zeta_n))$ satisfies $\frac{1}{n} N_{\le k^-} \xrightarrow{P} \eta(c)$.

For fixed k $\frac{1}{n} N_k \xrightarrow{P} \rho_k(c)$ $\qquad \omega(n) \le n^{1/4}$

Proof (Non-examinable)

key $\mathbb{P}(|\underset{\sim}{X}| \le k^-) \to \eta(c)$

show $\mathbb{P}(|C_v| \le k^-) \sim \mathbb{P}(|\underset{\sim}{X}| \le k^-)$

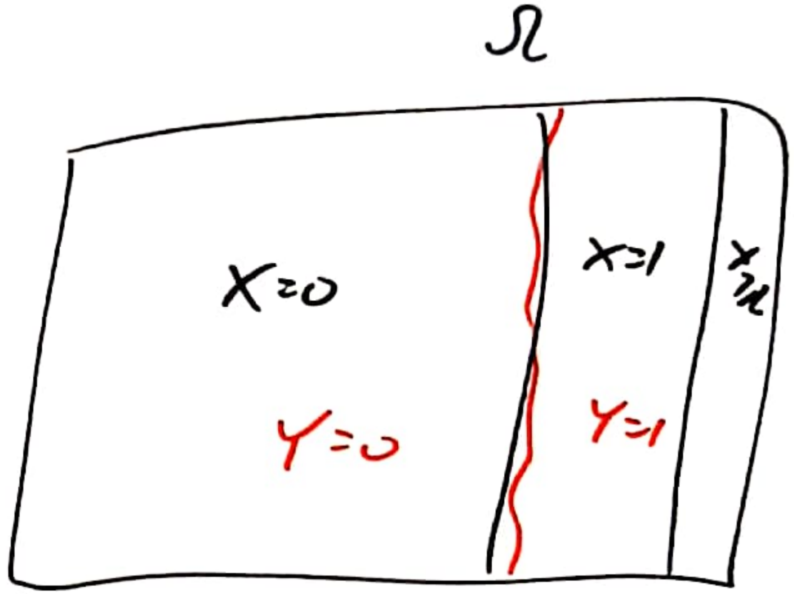
Use eg $\mathbb{P}(Bi(n-m, \zeta_n) = k)$ and $\mathbb{P}(Po(c)=k)$ agree

agree within a factor $1 \pm O\left(\frac{(k+m+1)^2}{n}\right)$ $\forall k, m \le \frac{n}{4}$ $\square$

Easy check $X \sim Po(p)$ & $Y \sim Bin(1,p)$ can be coupled to ~~agree~~ disagree with prob. $O(p^2)$.

$Po(c+d) \stackrel{d}{=} Po(c) + Po(d)$ if INDEP

$$\mathbb{P}(X=0) = e^{-p} = 1 - p + O(p^2)$$
$$\mathbb{P}(X=1) = pe^{-p} = p + O(p^2)$$
$$\mathbb{P}(X \geq 2) = O(p^2)$$



Follows can couple

$Po(\overset{np}{c})$ $Bin(n, \overset{p}{\tfrac{c}{n}})$ to disagree

with prob. $O(np^2) = O(\frac{1}{n})$

Coupling RVs / distributions

RV $\neq$ distribution

Can have different RVs with the same distribution

eg toss n coins

let $X = \#$ heads $\sim Bin(n, \frac{1}{2})$

$Y = \#$ tails $\sim Bin(n, \frac{1}{2})$

Note $X + Y = n$

$$Bin(n, \tfrac{1}{2}) + Bin(n, \tfrac{1}{2}) \stackrel{?}{=} Bin(2n, \tfrac{1}{2})$$

$$X + Y$$

Once For one RV X if ask

$\mathbb{P}(X=k)=?$, $\mathbb{P}(X\in S)$? only distribution matters. The data $p_k := \mathbb{P}(X=k)$ for all $k\in\mathbb{Z}$ is enough.

Consequence: can replace X by X' with same distribution

$$\mathbb{P}(X'=k) = \mathbb{P}(X=k).$$

Fact $\forall n, p, k$ if $X \sim Bin(n-1, p)$ $Y \sim Bin(n, p)$

$$\mathbb{P}(X \geq k) \leq \mathbb{P}(Y \geq k)$$

In general DO NOT have $X \leq Y$ always

BUT $\exists X', Y'$ with same distributions st. $X' \leq Y'$ does always hold

(a "coupling" of X and Y).

Here: consider n indep trials each success with prob p

Define $X' =$ # of successes in first $n-1$

$Y' =$ " " in all n

Then $X' \leq Y'$

$$\underbrace{\overbrace{SFFSFFS}^{X'}S}_{Y'}$$

$$
\begin{array}{ccc}
\mathbb{P}(X \geq k) & & \mathbb{P}(Y \geq k) \\
\parallel & & \parallel \\
\mathbb{P}(X' \geq k) & \leq & \mathbb{P}(Y' \geq k)
\end{array}
$$

$$
\{X' \geq k\} \subseteq \{Y' \geq k\}
$$

What if want to consider (X_1, X_2, X_3)

When does (X_1', X_2', X_3') have same distribution?

Need $\underbrace{\mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3)}_{\text{defines distribution}} = \mathbb{P}(\text{for } X_i')$

Equiv. $\mathbb{P}(X_1 = x_1)$

$\mathbb{P}(X_2 = x_2 \mid X_1 = x_1)$

$\mathbb{P}(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2)$

$\mathbb{P}(\text{first hits } 0 \text{ at } 3)$

Defn An event E ($=E_n$) holds with high probability / whp if $\mathbb{P}(E_n) \to 1$ as $n \to \infty$

Thm 5.10 (Phase transition, subcritical case)

Let $0<c<1$ be constant. There is a constant $A = A_c$ s.t. whp every component of $G(n, \frac{c}{n})$ has $\leq A \log n$ vxs.

Recall $|C_v| = m \iff$ cpt exploration walk first hits zero at $t = 0$

$|C_v| > k \iff$ walk above zero $t = 0, 1, \ldots, k$

Here $Y_0 = 1$

$$Y_t = Y_{t-1} + R_t - 1 \quad (\text{if } Y_{t-1} > 0)$$

where conditional dist of R_t given history is

$$\text{Bin}(U_{t-1}, p)$$

$$U_{t-1} = n - (t-1) - Y_{t-1} \leq n$$

Easier RW

$$Y_0^+ = 1$$

$$Y_t^+ = Y_{t-1}^+ + R_t^+ - 1$$

where conditional dist of R_t^+ given history is

$$\text{Bin}(n, p)$$

$\Leftrightarrow$ $R_1^+, R_2^+, \ldots$ iid $\text{Bin}(n,p)$

$$\boxed{\text{s.t. } R_t \leq R_t^+}$$

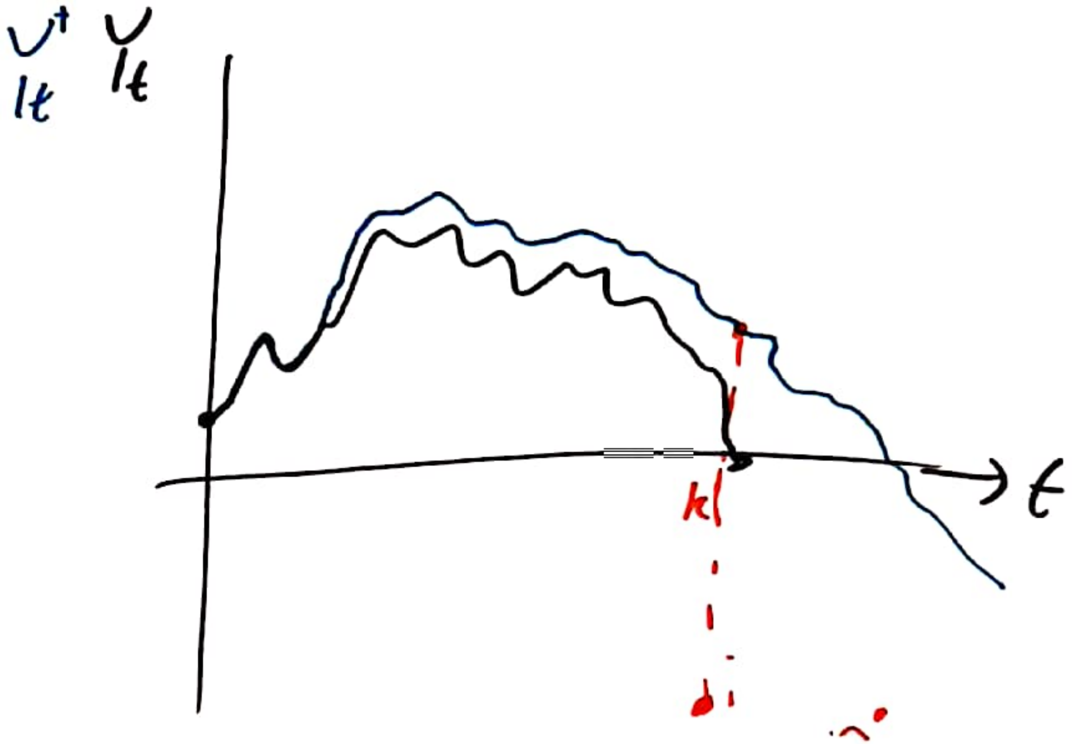
$$\Rightarrow Y_t \leq Y_t^+$$

$$\mathbb{P}(|C_v| > k)$$

$$= \mathbb{P}(Y_0, \dots, Y_k > 0)$$

$$\le \mathbb{P}(Y^+_0, \dots, Y^+_k > 0)$$

$$\le \mathbb{P}(Y^+_k > 0)$$



Key $\mathbb{E}[R^+_t - 1] = np - 1 = c - 1 < 0$

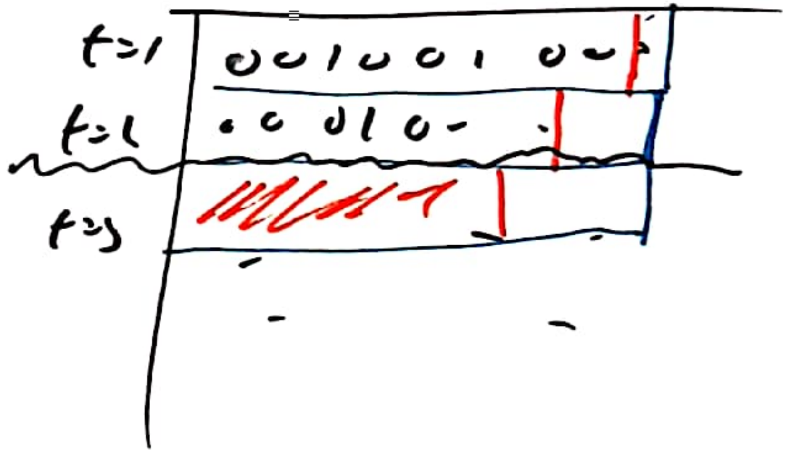
~~Formal~~ Construction of Coupling:

let $(B_{t,i})_{t=1,\, i=1}^{n \quad n}$ be indep. Bernoulli RVs, each $\begin{cases} 1 & \text{prob } p \\ 0 & 1-p \end{cases}$

Set $R_t^+ = \sum_{i=1}^{n} B_{t,i} \sim \mathrm{Bin}(n,p)$ ✓

For Y_t / R_t set $Y_0 = 1$

Having defined Y_{t-1} let



$$R_t = \sum_{i=1}^{n-(t-1)-Y_{t-1}} B_{t,i} \qquad (\text{if } Y_{t-1} > 0)$$

Conditional on history up to time $t-1$ involves rows $1, \ldots, t-1$ of table

Then know $Y_{t-1} = y_{t-1}$ & R_t has conditional dist $\mathrm{Bin}(n-(t-1)-y_{t-1}, p)$ ✓

Set $Y_t = Y_{t-1} + R_t - 1$ NB $R_t \le R_t^+$

Now $\mathbb{P}(|C_v| > k) = \mathbb{P}(Y_1, \dots, Y_k > 0)$

$$\le \mathbb{P}(Y_1^+, \dots, Y_k^+ > 0) \le \mathbb{P}(Y_k^+ > 0).$$

Now $Y_k^+ = 1 - k + \sum_{t=1}^{k} R_t^+$ $\sim \mathrm{Bin}(kn, p)$ $\qquad \mathrm{Bin}(n,p)$

mean $\mu = knp = kc$

$$\mathbb{P}(Y_k^+ > 0) = \mathbb{P}(\mathrm{Bin}(kn, p) \ge k)$$

$$\mu + (1-c)k = \mu(1+\varepsilon')$$

$$\le \mathbb{P}(\mathrm{Bin}(kn,p) \ge \mu(1+\varepsilon)) \text{ where } \varepsilon = \min\left\{\frac{1-c}{c}, 1\right\}$$

$$\le e^{-\varepsilon^2 \mu / 4} \text{ by Chernoff}$$

$$= e^{-\varepsilon^2 ck/4}.$$

Now let $A = \frac{8}{\varepsilon^2 c}$ = const

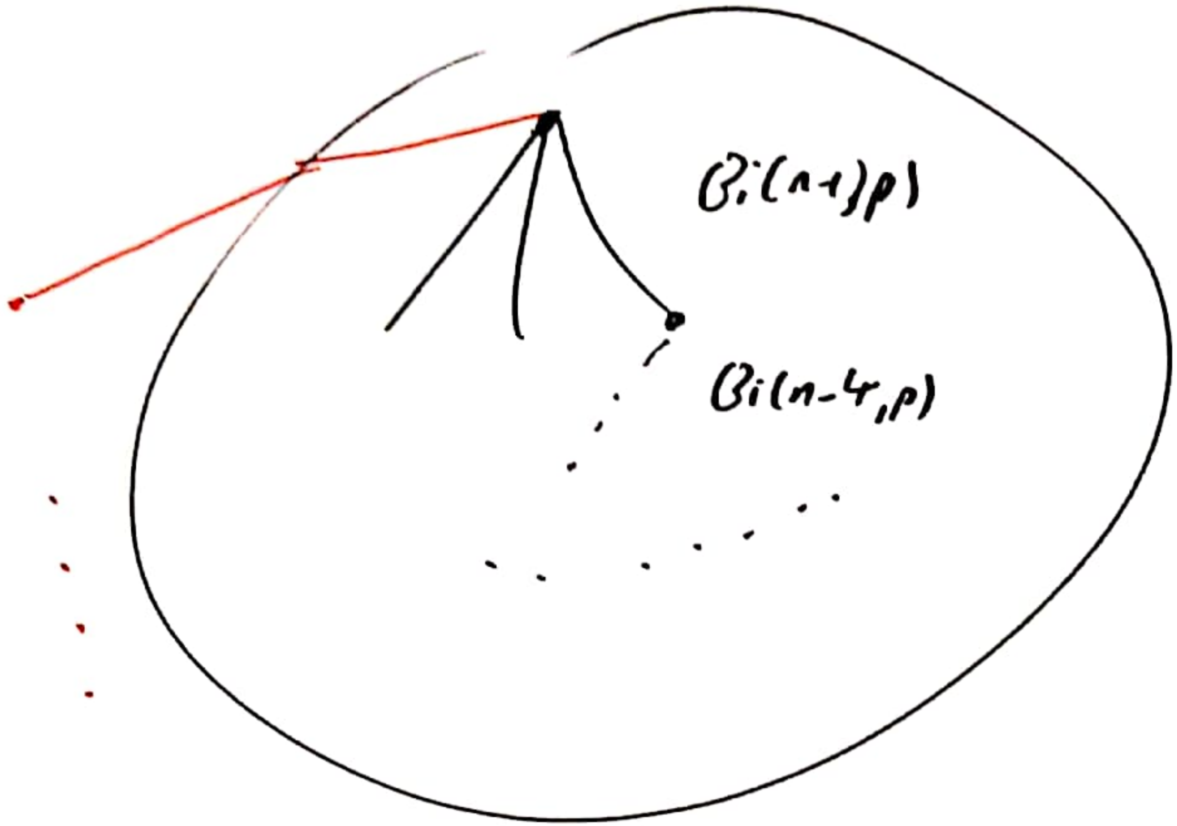
Take $k = A \log n$ $\lceil A \log n \rceil$

$$\mathbb{P}(|C_v| > k) \le e^{-2\log n} = \frac{1}{n^2}$$

By union bound $\mathbb{P}(\exists \text{ cpt size } > k)$

$$\le \sum_v \mathbb{P}(|C_v| > k) \le \frac{1}{n} \to 0$$

$\square$


Bi(n-1,p)
Bi(n-4,p)

Let $L_i(G)$ = # vxs in i^{th} largest cpt

<u>Thm 5.11</u> Let $c>1$ be constant, consider $G = G(n, \frac{c}{n})$.
Then $\frac{L_1(G)}{n} \xrightarrow{p} 1 - \eta(c)$. Also $\exists$ constant $A = A_c$
s.t. whp $L_2(G) \leq A \log n$

9, 8, 8, 5

$L_1 = 9$, $L_2 = 8$, $L_3 = 8$

$R_t \sim Bin(U_{t-1}, p)$

$\mathbb{E}(R_t - 1)$: drift $= U_{t-1}p - 1 = \frac{U_{t-1}c}{n} - 1$ $\geq c\delta$

>0 at start $(U_{t-1} \approx n)$

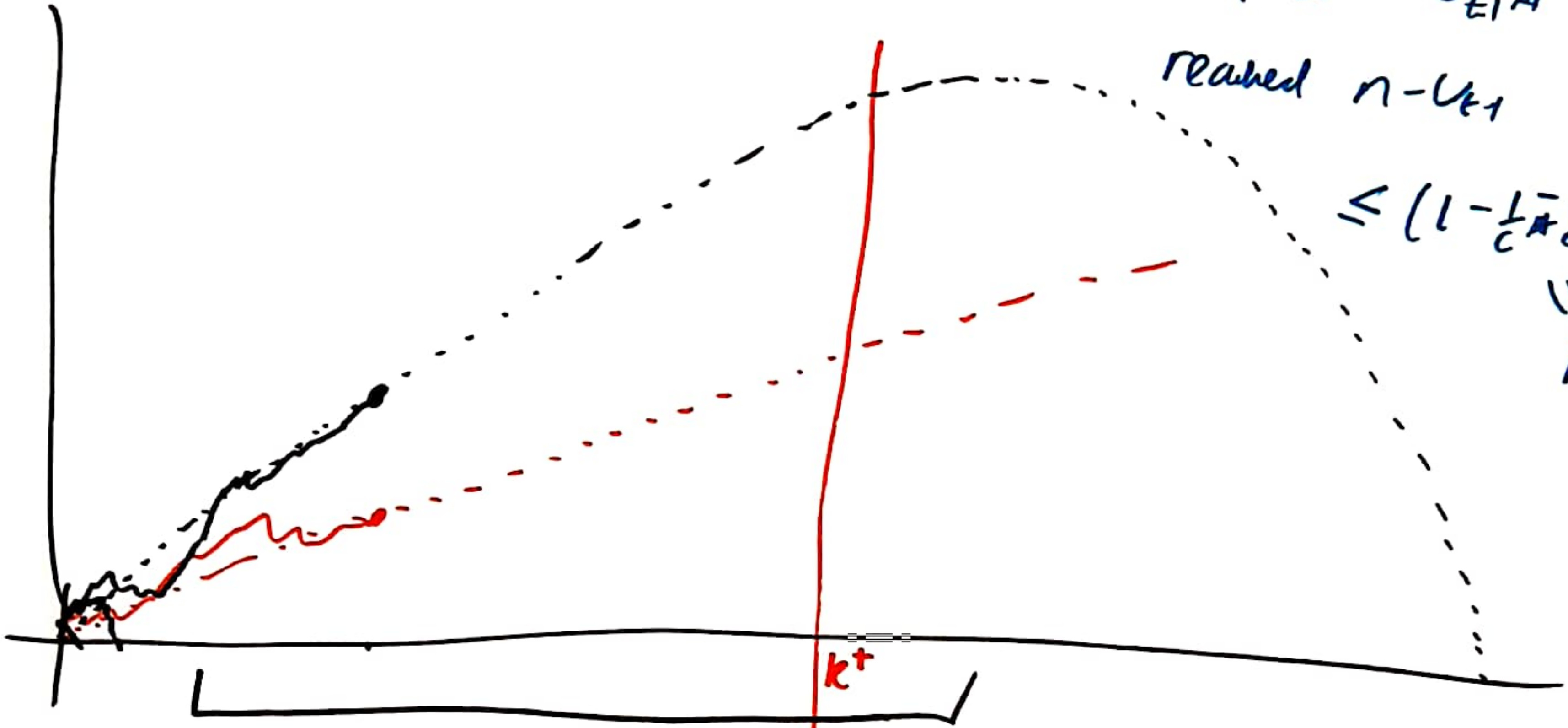
Take Y_t^-, R_t^- drift $\geq c\delta$ Need $U_{t-1} \geq (\frac{1}{c}+\delta)n$

reached $n - U_{t-1}$

$\leq (1 - \frac{1}{c} - \delta)n$

$\parallel$

k^+



Unlikely $Y_t = 0$

Fix $\delta > 0$

Let $k^+ = (1 - \frac{1}{c} - \delta) n$

Construct Y_t^- RW with $Y_t^- = Y_{t-1}^- + R_t^- - 1$

with iid $R_t^- \sim Bin(n - k^+, p)$

s.t. $R_t \geq R_t^-$ if 'used up' $\leq k^+$ vxs

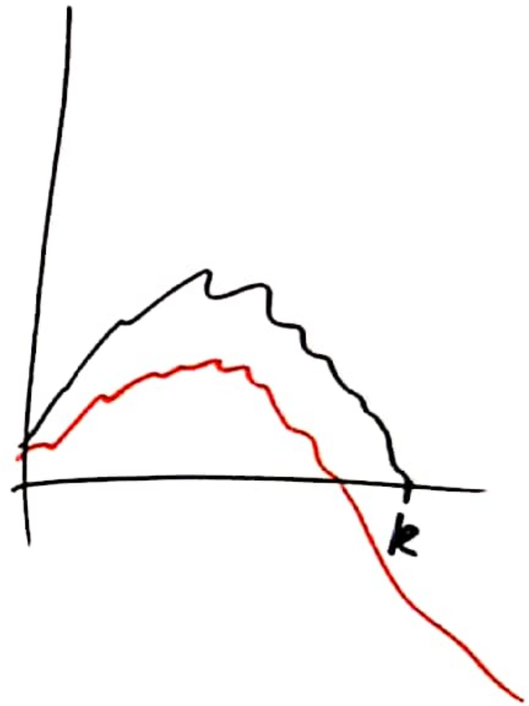
$U_{t-1} \geq n - k^+$

$$\left[\text{As before} \quad R_t^- = \sum_{i=1}^{n-k^+} B_{t,i} \right]$$

For $k \leq k^+$

$$\mathbb{P}(|C_v| = k) = \mathbb{P}(Y_1, \dots, Y_{k-1} > 0,\ Y_k = 0) \quad (*)$$

$$\leq \mathbb{P}(Y_k^- \leq 0)$$



[If $(*)$ holds after k steps

$U_k = n - k \geq n - k_+$

U_t is decreasing so in this case

For $t < k$ $\quad U_t \geq U_k \geq n - k_+$

$\therefore \quad R_t^- \leq R_t \quad$ for $\quad t \leq k$

$\therefore \quad Y_t^- \leq Y_t \quad t \leq k \qquad Y_k^- \leq Y_k = 0$]

$$Y_k^- = 1 - k + \sum_{t=1}^{k} R_t^-$$

$\mathrm{Bin}(kn(\tfrac{1}{c}+\delta), \tfrac{c}{n})$ mean $\mu = k + c\delta k$

$\mathrm{Bin}(n - k^s, p) = \mathrm{Bin}(n(\tfrac{1}{c}+\delta), \tfrac{c}{n})$

$$\mathbb{P}(Y_k^- \le 0) \le \mathbb{P}(\mathrm{Bin}(\quad) \le k)$$

$$k = \mu(1-\varepsilon) \qquad 1-\varepsilon = \frac{1}{1+c\delta}$$

By Chernoff $\mathbb{P}(\quad) \le e^{-\varepsilon^2\mu/2} \le e^{-\varepsilon^2 k/2}$.

Let $k^- = \frac{6}{\varepsilon^2} \log n$

For $k^- \le k \le k^+$ $\quad \mathbb{P}(|C_v| \ge k) \le e^{-\frac{\varepsilon^2 k}{2}} \le e^{-3\log n} = \frac{1}{n^3}$

By union bound over v, k

$$\mathbb{P}(\exists \text{ any cpt of size between } k^-, k^+) \le \frac{1}{n} \to 0$$

Whp ALL cpts are

- SMALL $< k^-$

or

- BIG $> k^+$

By Lemma 5.9

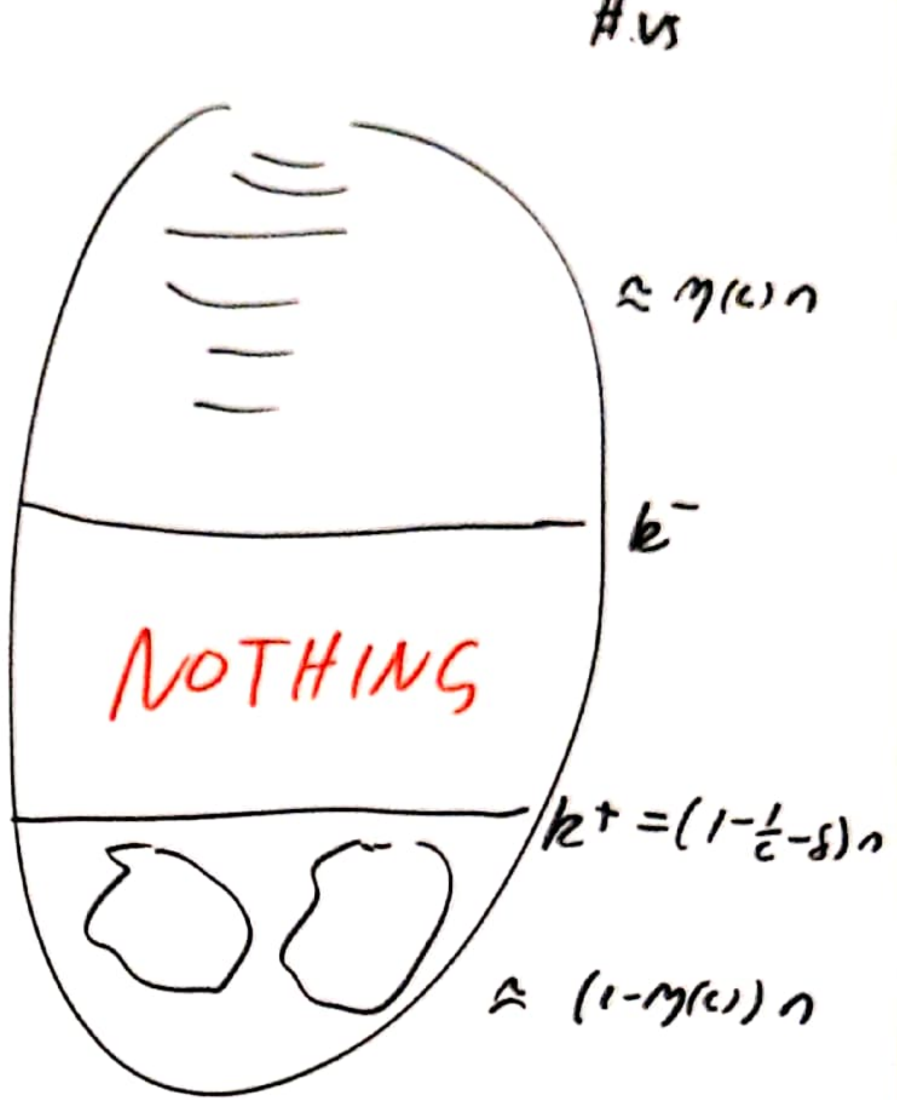
$\frac{1}{n} N_{\le k^-} \xrightarrow{P} \eta(c)$

Let N = # vs in large cpts

$\Rightarrow \quad \frac{1}{n} N \xrightarrow{P} 1-\eta(c)$

Suppose $2(1-\frac{1}{c}-\delta) > 1-\eta(c)$.

THEN DONE



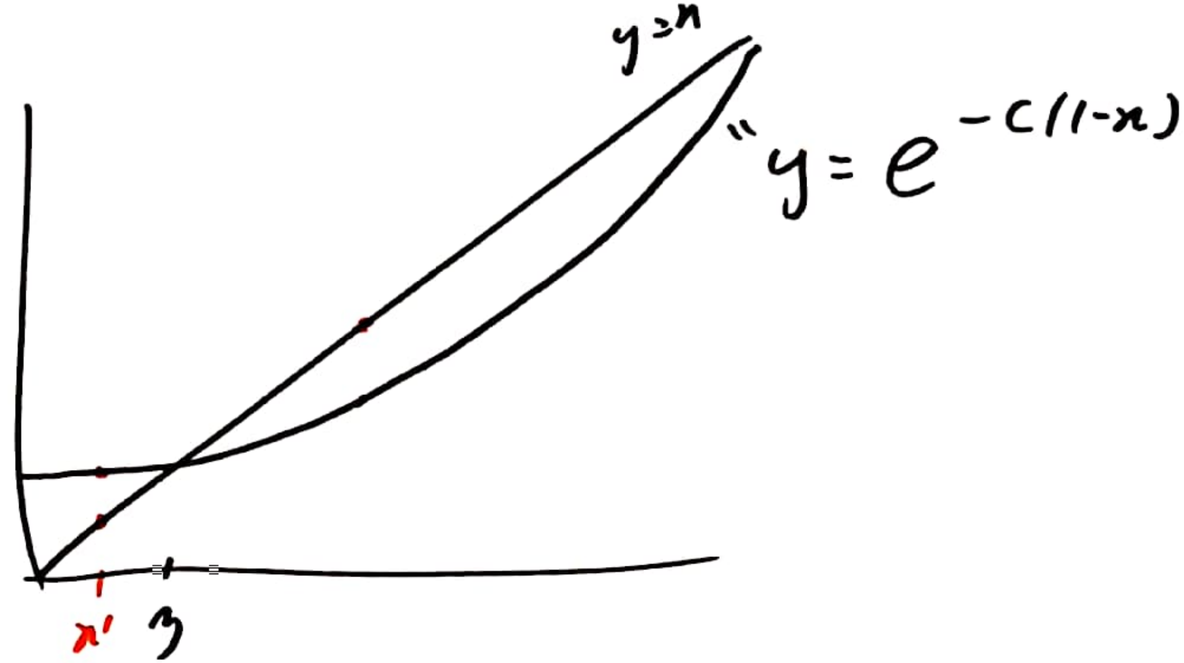
$\cdot \quad N \le (1-\eta(c)+\varepsilon)n$

$< 2k^+$

No room for two large cpts

want $\exists \delta$ s.t. $2(1-\frac{1}{c}-\delta) \geq 1-\eta(c)$

i.e. $2(1-\frac{1}{c}) \geq 1-\eta(c)$ $\quad\bullet\quad \eta > \overset{x'}{\delta(c)}$



$\square$