Chernoff bounds

These bounds are a variation of Markov's inequality that have meaningful applications.

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $I \equiv I(X)$ be an interval such that

Moment generating function:
$M_X : I \to \mathbb{R}_+$ given by
$M_X(r) = E[e^{rX}]$ is well-defined.

Let, $r \in I$; $Y = e^{rX}$. The Markov inequality implies

$$\mathbb{P}(Y \geq a) \leq \frac{E[Y]}{a} \quad \forall\, a > 0$$

$$\Rightarrow \quad \mathbb{P}(e^{rX} \geq a) \leq \frac{M_X(r)}{a} \quad \forall\, a > 0, \ \forall\, r \in I$$

Now let $a = e^{rb}$ (say)

$r > 0,\ r \in I \Rightarrow \mathbb{P}(X \geq b) \leq e^{-rb} M_X(r)$

$r < 0,\ r \in I \Rightarrow \mathbb{P}(X \leq b) \leq e^{-rb} M_X(r)$ } (c)

- Chernoff bounds -

These provide a family of upper bounds on the tails of the random variable X.

In (C) for $r > 0$, $r \in I(x)$

$\mathbb{P}(X \geq b)$ — decays exponentially to 0 at rate e^{-rb}

(Note: not true for all X : require $E[e^{rX}] < \infty$)

Example 1 : let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d X

for $n \geq 1$ $S_n = \sum_{i=1}^{n} X_i$

$r \in I(x)$ $E[e^{rS_n}] = \prod_{i=1}^{n} E[e^{rX_i}] = (M_X(r))^n$

(C) with $b = na$ gives

$$\mathbb{P}(S_n \geq na) \leq (M_X(r))^n e^{-rna}$$

$$\mathbb{P}(S_n < na) < (M_X(r))^n e^{-rna}$$

let $\gamma_X(r) = \log(M_X(r))$, for $r \in I(x)$.

$$\mathbb{P}(S_n \geq na) \leq \exp(n(\gamma_X(r) - ra)) \qquad r > 0,\ r \in I(x)$$

$$\mathbb{P}(S_n \leq na) \leq \exp(n(\gamma_X(r) - ra)) \qquad r < 0,\ r \in I(x)$$

Lemma 1: Let $X_1, X_2, \ldots, X_n$ be i.i.d X. Assume $0 \in I(X)^\circ$. Then,

$$\mu_X(a) = \inf_{r \in I(X)} [r_X(r) - ra] < 0, \quad \forall \; a \neq E[X]$$

Also,

$$P(S_n \geq na) \leq e^{n \mu_X(a)} \quad \text{for } a > E[X]$$

$$P(S_n \leq na) \leq e^{n \mu_X(a)} \quad \text{for } a < E[X]$$

Proof:-

let $I(X) = [r_-, r_+]$. Fix $a \in \mathbb{R}$.

$h : I \to \mathbb{R}$ be given by

$$h(r) = r_X(r) - ra$$

$$h'(r) = r_X'(r) - a$$

$\therefore$ h attains its minimum at

- r_-,
- $r_0 : r_X'(r_0) = a$
- or r_+.

let $\mu_X(a) := \min(h(r_-), h(r_0), h(r_+))$

$$= \inf_r [r_X(r) - ra)$$

Now: $r_X'(r) = \frac{d}{dr} \ln(E(e^{rX}))$

$$\left(\text{to be covered soon}\right) \doteq \frac{1}{E[e^{rX}]} E[X e^{rX}]$$

$\Rightarrow$ $\boxed{h'(0) = E[X] - a}$ - (1)

$$h''(\gamma) = r_X''(\gamma) = \frac{E[X^2 e^{\gamma X}] - \left(E[X e^{\gamma X}]\right)^2}{\left(E[e^{\gamma X}]\right)^2}$$

Ex :- $r_X''(\gamma) \geq 0$

$\Rightarrow$ $\boxed{h''(\gamma) \geq 0}$ - (2)

Keys :-

- $a < E[X] \Rightarrow h'(0) > 0 \; ; \; h(0) = 0$

($h'(.)$ is continuous) $\Rightarrow \exists \; \delta > 0 \quad -\delta < \gamma < 0 \Rightarrow h(\gamma) < 0$

$h''(\gamma) \geq 0 \Rightarrow \mu_X(a) < 0$ and

$$\mu_X(a) = \inf_{\gamma < 0} \{r_X(\gamma) - \gamma a\}$$

- $a > E[X] \Rightarrow h'(0) < 0 \; ; \; h(0) = 0$

($h'(.)$ is continuous) $\Rightarrow \exists \; \eta > 0 \quad 0 < \gamma < \eta \Rightarrow h(\gamma) < 0$

$h''(\gamma) \geq 0 \Rightarrow \mu_X(a) < 0$ and

$$\mu_X(a) = \inf_{\gamma > 0} \{r_X(\gamma) - \gamma a\}$$

From (a) result follows $\square$

Example 1: $X_1, X_2, \ldots, X_n, \ldots$ i.i.d $X \sim \text{Bernoulli}(p)$

$$M_X(r) = pe^r + 1-p \qquad \forall\, r \in (-\infty, \infty)$$

$$h(r) = r_X(r) - ra \qquad a \in \mathbb{R}$$

$$= \ln(pe^r + 1-p) - ra$$

$$h'(r) = 0 \iff \frac{pe^r}{pe^r + 1-p} = a$$

$$\iff r = \ln\left(\frac{a(1-p)}{p(1-a)}\right)$$

$$\Rightarrow \mu_X(a) = \ln\left(p\,\frac{a(1-p)}{p(1-a)} + (1-p)\right) - a\ln\left(\frac{a(1-p)}{p(1-a)}\right)$$

$$= (1-a)\ln\left(\frac{1-p}{1-a}\right) + a\ln\left(\frac{p}{a}\right)$$

$\therefore$

$$a > p \Rightarrow \mathbb{P}(S_n \geq na) \leq \exp\left(n\left[(1-a)\ln\left(\frac{1-p}{1-a}\right) + a\ln\left(\frac{p}{a}\right)\right]\right)$$

$$a < p \Rightarrow \mathbb{P}(S_n \leq na) \leq \exp\left(n\left[(1-a)\ln\left(\frac{1-p}{1-a}\right) + a\ln\left(\frac{p}{a}\right)\right]\right) \quad -(3)$$

(Markov) $\mathbb{P}(S_n \geq na) \leq \frac{p}{a} \quad -(4)$

(Chebyshev) $\mathbb{P}(S_n \geq na) = \mathbb{P}((S_n - np) \geq n(a-p))$

$$\leq \mathbb{P}(|S_n - np| \geq n(a-p))$$

$$\leq \frac{np(1-p)}{n^2(a-p)^2} = \frac{p(1-p)}{n(a-p)^2}$$

Proposition 3 (Cauchy-Bunyakovsky-Schwarz inequality)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a Probability space. Let X, Y be random variables such that $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[Y^2] < \infty$. Then

$$\mathbb{E}\,|XY| \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}$$

Proof:- $Z = \dfrac{|X|}{\sqrt{\mathbb{E}[X^2]}}$, $W = \dfrac{|Y|}{\sqrt{\mathbb{E}[Y^2]}}$

Both Z, W are measurable functions of X, Y $\Rightarrow$ Z, W are random variables.

$$(Z-W)^2 \geq 0$$

$$\Leftrightarrow \quad Z^2 + W^2 \geq 2ZW$$

$$\Rightarrow \mathbb{E}[ZW] \leq \frac{1}{2}\,\mathbb{E}[Z^2 + W^2]$$

$$\Rightarrow \quad \mathbb{E}\,|XY| \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]} \qquad \square$$

Proposition 4 (Jensen's Inequality):

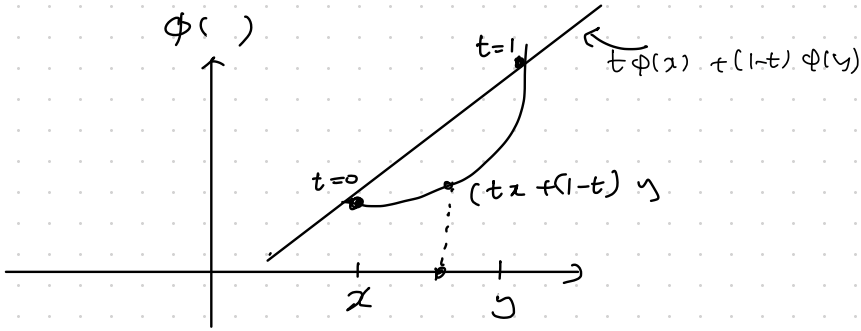
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a Probability space and X be a random variable with finite mean $E[X]$. Let $\phi: \mathbb{R} \to \mathbb{R}$ be a convex function such that $E[\phi(X)]$ exists.

Then $$E[\phi(X)] \geq \phi(E[X])$$

Proof:- Now ϕ is Convex $\Leftrightarrow$

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$$

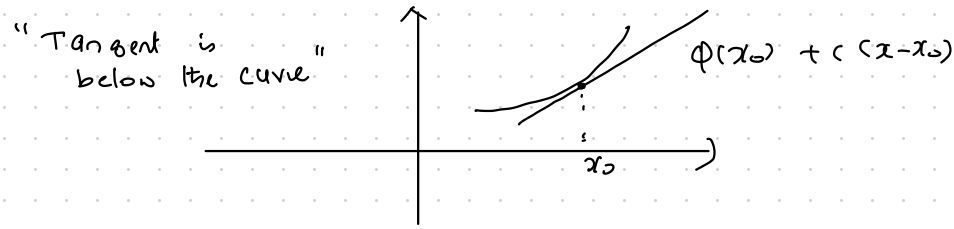
$\forall\, t \in (0,1) \quad \forall\, x, y \in \mathbb{R}.$



lemma 1: Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex. Then $\forall x_0 \in \mathbb{R}$

$\exists\, c \equiv \partial\phi(x_0)$ s.t.

$$\phi(x) \geq \phi(x_0) + c(x - x_0)$$



Proof of lemma1: let $x < y < z$

$$\Rightarrow \quad y = tx + (1-t)z \quad \text{for } t \in [0,1]$$

$$\Rightarrow \quad t = \frac{y-z}{x-z} \in (0,1)$$

ϕ – convex $\Rightarrow$

$$\phi(tx + (1-t)z) \leq t\phi(x) + (1-t)\phi(z)$$

$(\Leftarrow)$

$$\frac{\phi(x) - \phi(y)}{x-y} \leq \frac{\phi(z) - \phi(y)}{z-y} \quad -(1)$$

$\forall \quad x < y < z$

$\therefore \quad k > 1 \qquad a_k = \dfrac{\phi(x_0 - \frac{1}{k}) - \phi(x_0)}{-1/k}$

$m \geq 1$ $\qquad b_m = \dfrac{\phi(x_0 + \frac{1}{m}) - \phi(x_0)}{\frac{1}{m}}$

By (1) :

$$a_k \leq a_{k+1} \quad \forall\ k \geq 1$$

$$b_m \geq b_{m+1} \quad \forall\ m \geq 1$$

$$a_k \leq b_m \quad \forall\ \begin{matrix} k \geq 1 \\ m \geq 1 \end{matrix}$$

$\Rightarrow$ $\alpha := \lim_{k \to \infty} a_k$ $\qquad \beta := \lim_{m \to \infty} b_m$

exists and

$$\alpha \leq \beta$$

$\therefore$ If $\alpha = \beta$ then $c = \alpha = \beta$ (if f is differentiable at x_0)

otherwise choose any $c \in (\alpha, \beta)$

Also, $x < x_0$ by (1)

$$\frac{\phi(x) - \phi(x_0)}{x - x_0} \leq \alpha \leq c$$

Similarly, $x > x_0$ by (1)

$$\frac{\phi(x) - \phi(x_0)}{x - x_0} \geq \beta \geq c$$

$\Rightarrow$ $\forall\ x \in \mathbb{R}$ $\qquad \phi(x) \geq c(x - x_0) + \phi(x_0)$ $\square$

Proof:- Let $x_0 = E[X]$ By lemma 1

$\exists\, c \in \mathbb{R}$ s.t

$$\phi(X) \geq \phi(x_0) + c(X - x_0)$$

$$\Rightarrow E[\phi(X)] \geq \phi(E[X]) + 0$$

□

9 First and Second Moment Method.

In general the Moment method is a method to control the probability that a random variable takes values far away from its expectation by its moments (i.e. by $E[X^k]$ $k \geq 1$).

Already seen

$X \geq 0$ r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ then (Markov's Inequality)

$$\mathbb{P}(X \geq \alpha) \leq \frac{E[X]}{\alpha} \quad - (1)$$

The first moment method, is essentially Markov's inequality for discrete random variables. More precisely, $X \geq 0$ integer valued then

$$\mathbb{P}(X > 0) = \mathbb{P}(X \geq 1) \overset{(1)}{\leq} E[X]$$

$$\Rightarrow \boxed{\mathbb{P}(X > 0) \leq E[X] \quad - \text{ 1st Moment Method}} \quad - (2)$$

The second moment method controls the $\mathbb{P}(X > 0)$ by its first ($E[X]$) and second ($E[X^2]$) moment.

More precisely,

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X \geq 0$, $0 < E[X] < \infty$ and $0 < E[X^2] < \infty$.

Then

$$E[X^2] = E[X \cdot X 1_{X>0}]$$

(C.B.S. inequality) $\Leftarrow \leq \sqrt{E[X^2]}\sqrt{E[1_{X>0}]}$

$$\Rightarrow \quad \mathbb{P}(X>0) \geq \frac{(E[X])^2}{E[X^2]}$$ (Second moment method)

These seemingly elementary inequalities are given their respective names because they are very useful in many applications

Example 1 :- (Coupon Collector Problem) Fix $n \geq 2$. $Y_1, Y_2, \ldots$ be i.i.d. uniform $\{1, \ldots, n\}$. Let

$$T_n = \inf\{k \geq 1 \mid \{Y_1, \ldots, Y_k\} = \{1, \ldots, n\}\}$$

Coupon Collector problem

- think of $1, 2, \ldots, n$ as coupons in a Tinkuazad company chips packet.
- each packet has a randomly chosen coupon placed in it
- T_n = # of chips packets you need to buy to collect all n coupons.

Has many variations

Question :- What is the order of T_n?

Answer(s) :- · Fix $m \geq 1$ - to denote # of days

· let $k \in \{1, \ldots, n\}$

and $X_{m,k} = \begin{cases} 1 & \text{if } k^{th} \text{ coupon has not been picked until } m^{th} \text{ day} \\ 0 & \text{otherwise} \end{cases}$

Note $\mathbb{P}(X_{m,k}=1) = \left(1-\frac{1}{n}\right)$ $\forall\, m \geq 1$, $1 \leq k \leq n$

$\Rightarrow \mathbb{E}[X_{m,k}] = \left(1-\frac{1}{n}\right)^m$ ——— (1)

Also $\mathbb{P}(X_{m,k}=1, X_{m,\ell}=1) = \left(1-\frac{2}{n}\right)^m$ $\forall\, m \geq 1$, $1 \leq k \neq \ell \leq n$

$\Rightarrow \mathbb{E}[X_{m,k}\, X_{m,\ell}] = \left(1-\frac{2}{n}\right)^m$ ——— (2)

$\mathbb{P}(T_n > m) = \mathbb{P}(X_{m,1} + \cdots + X_{m,n} \geq 1)$

1st moment method $\leq \mathbb{E}[X_{m,1} + \cdots + X_{m,n}]$

$\overset{(1)}{=} n\left(1-\frac{1}{n}\right)^m$

Ex: $1-x \leq e^{-x}$ $\forall\, x > 0$ $\leq n e^{-\frac{m}{n}}$ ——— (3)

If $m \cong n\log n$ then $n e^{-\frac{m}{n}} \cong 1$.

From this intuition take $m = n\log n + n\theta_n$ for some sequence of real numbers θ_n, $0 < \theta_n < \log n$.

Therefore,

$\mathbb{P}(T_n > n\log n + n\theta_n)$

Ex: $\overset{(3)}{\leq} n\, e^{-\frac{\lfloor n\log n + n\theta_n \rfloor}{n}}$

$x - 1 < \lfloor x \rfloor \leq x$ $\forall\, x \in \mathbb{R}$ $\leq e^{-\theta_n}$ ——— (4)

Now on the other hand,

$$\mathbb{P}(T_n > m) = \mathbb{P}(X_{m,1} + \cdots + X_{m,n} \geq 1)$$

2nd moment method $\Rightarrow$

$$\geq \frac{\left(E[X_{m,1} + \cdots + X_{m,n}]\right)^2}{E[(X_{m,1} + \cdots + X_{m,n})^2]}$$

$$\overset{(1)}{=} \frac{n^2\left(1-\frac{1}{n}\right)^{2m}}{\left(\sum_{k=1}^{n} E[X_{m,k}^2] + \sum_{\substack{k,l=1 \\ k \neq l}}^{n} E[X_{m,k} X_{m,l}]\right)}$$

$E[X_{m,k}^2] = E[X_{m,k}]$ (1) & (2) $=$

$$\frac{n\left(1-\frac{1}{n}\right)^{2m}}{n\left(1-\frac{1}{n}\right)^{m} + n(n-1)\left(1-\frac{2}{n}\right)^{m}}$$

Ex:-

$1 - x \leq e^{-x}$

& $1 - x \geq e^{-x-x^2}$ $\Leftarrow$ $\geq$

for $-\frac{1}{2} \leq x \leq \frac{1}{2}$

$$\frac{n\, e^{-\frac{2m}{n} - \frac{2m}{n^2}}}{n\, e^{-\frac{m}{n}} + n(n-1)\, e^{-\frac{2m}{n}}}$$

$\therefore$ $\mathbb{P}(T_n > n\log n - n\theta_n)$

Ex. $\geq$

$$\frac{e^{2\theta_n - \frac{2}{n}\log n + \frac{2\theta_n}{n}}}{e^{\theta_n} + \left(n(n-1)e^{-2\log n}\right) e^{2\theta_n}}$$

$\geq$

$$\frac{e^{\theta_n - \frac{2}{n}(\log n - \theta_n)}}{1 + e^{\theta_n}}$$

$=$

$$\frac{e^{\theta_n\left(1+\frac{2}{n}\right)}}{n^{2/n}\left(1+e^{\theta_n}\right)} \qquad \text{—— (5)}$$

$$\mathbb{P}(|T_n - n\log n| > n\theta_n)$$

$$= \mathbb{P}(T_n > n\log n + n\theta_n)$$

$$1 - \mathbb{P}(T_n \geq n\log n - n\theta_n)$$

$$\overset{(4),(5)}{\leq} e^{-\theta_n} + 1 - \frac{e^{\theta_n(1+\frac{2}{n})}}{n^{2/n}(1+e^{\theta_n})} \qquad -(6)$$

$\therefore$ From (6)

if $0 \leq \theta_n \leq \log n$ and $\theta_n \to \infty$ as $n \to \infty$

then

$$\mathbb{P}(|T_n - n\log n| > n\theta_n) \longrightarrow 0 \quad \text{as} \quad n \to \infty. \qquad -(7)$$

$\Rightarrow$ T_n concentrates around $n\log n$ in a window of size n.

Corollary 1 :-

$$\frac{T_n}{n\log n} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty$$

Proof :- Given $0 < \varepsilon < 1$, take $\theta_n = \varepsilon \log n$

$$\Rightarrow \mathbb{P}\left(\left|\frac{T_n}{n\log n} - 1\right| > \varepsilon\right) = \mathbb{P}(|T - n\log n| > n\theta_n).$$

Ex:- Use (7) to complete the proof. $\square$

Example 2 : Bond - Percolation on $\mathbb{T}_2$

let $\mathbb{T}_2^0 = \{0\}$ & $\mathbb{T}_2^n = \{0,1\}^n$ $\forall n \geq 1$

(Vertices) $\mathbb{T}_2$ - binary tree $= \bigcup_{n=0}^{\infty} \mathbb{T}_2^n$

$\forall\, x \in \mathbb{T}_2^n$ let $a(x) = (x_1, x_2, \ldots, x_{n-1}) \in \mathbb{T}_2^{n-1}$ be the ancestor of x in $\mathbb{T}_2$.

(Edge)

$$E := \left\{ \{x, a(x)\} \;\middle|\; x \in \mathbb{T}_2 \setminus \mathbb{T}_2^0 \right\}$$

Fix $0 \leq p \leq 1$; declare each edge $e \in E$ to be

- open with probability p
- closed with probability $1-p$

[each $e \in E_d$ is independently open or closed with probability p] ← in an i.i.d. manner

(Path)
$x = x_0 \to x_1 \cdots \to x_n = y$ is path in $\mathbb{T}^2$ between x and y if $\{x_i, x_{i+1}\} \in E \quad \forall\, i = 0, 1, \ldots, n-1$.

Ex: Between $x, y \in \mathbb{T}^2$ ∃ a unique path in $\mathbb{T}^2$.

let $C_0 = \{x \in \mathbb{T}^2 \mid \text{path between } x \rightsquigarrow 0 \text{ is open}\}$

let $X_n = \sum_{x \in \mathbb{T}_2^n} \mathbb{1}(x \in C_0)$

(i.e X_n = # of vertices in $\mathbb{T}_2^n \cap C_0$)

$$\mathbb{P}(X_n > 0) \geq \frac{(E[X_n])^2}{E[X_n^2]}$$

(Second moment method)

$$= \frac{\left(\sum_{x \in \mathbb{T}_2^n} \mathbb{P}(x \in C_0)\right)^2}{\sum_{y \in \mathbb{T}_2^n} \sum_{x \in \mathbb{T}_2^n} \mathbb{P}(x \in C_0, y \in C_0)}$$

- $\sum_{x \in \mathbb{T}_2^n} \mathbb{P}(x \in C_0) = \sum_{x \in \mathbb{T}_2^n} \mathbb{P}($  $) = \sum_{x \in \mathbb{T}_2^n} p^n = |\mathbb{T}_2^n|\, p^n = 2^n p^n$

- (8)

- $\forall\, x, y \in \mathbb{T}_2^n \quad n \geq 1$, $x \neq y$

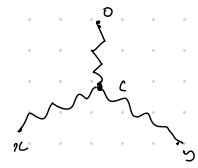
∃ a unique $c = c(x, y, 0) \in \mathbb{T}_2$ s.t $c \in \mathbb{T}_2^s$ $s \leq n-1$

& c lies on the path $x \rightsquigarrow 0$ and $y \overset{\text{path}}{\rightsquigarrow} x$.

- if $x = y$ $\quad c(x, y, 0) = x = y \in \mathbb{T}_2^n$

So

$$\mathbb{P}(x \in C_0, y \in C_0) = \mathbb{P}\left(\right) = p^{2n-s}$$



$$\Rightarrow \sum_{\substack{y \in \Pi_2^n \\ x \neq y}} \sum_{x \in \Pi_2^n} \mathbb{P}(x \in C_0, y \in C_0)$$

$$= \sum_{s=0}^{n-1} \sum_{y \in \Pi_2^n} \sum_{x \in \Pi_2^n} \mathbb{P}(x \in C_0, y \in C_0, c \in \Pi_2^s)$$

$$= \sum_{s=0}^{n-1} \sum_{y \in \Pi_2^n} \sum_{x \in \Pi_2^n} p^{2n-s} \, \mathbb{1}(x \in \Pi_2^n, y \in \Pi_2^n, c \in \Pi_2^s)$$

$$= \sum_{s=0}^{n-1} p^{2n-s} \; 2^s \cdot 2^{n-s} \; 2^{n-s-1}$$

$$= (2p)^{2n} \sum_{s=0}^{n-1} (2p)^{-s}$$

$$= \frac{1}{2} \cdot (2p)^{2n} \cdot \frac{1 - (\frac{1}{2p})^n}{1 - (\frac{1}{2p})} \leq (2p)^{2n} \frac{p}{2p-1} \quad \text{if } p > \frac{1}{2}$$

$$-(9)$$

$$\Rightarrow \sum_{\substack{y \in \Pi_2^n \\ x = y}} \sum_{x \in \Pi_2^n} \mathbb{P}(x \in C_0, y \in C_0) = p^n \cdot 2^n \quad -(10)$$

$$\therefore \mathbb{P}(X_1 > 0) \overset{(9),(10)}{\geq} \frac{(2^n p^n)^2}{(2p)^{2n} \cdot \frac{p}{2p-1} + (2p)^n} = \frac{1}{\frac{p}{2p-1} + (2p)^{-n}} \quad \text{if } p > \frac{1}{2}$$

$$\Rightarrow \quad \mathbb{P}(X_n > 0) \geq \frac{1}{\frac{p}{2p-1} + (2p)^{-n}} \quad \text{if} \quad p \in \left(\tfrac{1}{2}, 1\right] \qquad -(11)$$

$$\therefore \quad \mathbb{P}(|C_0| = \infty) \geq \mathbb{P}\left(\limsup_{n\to\infty} \{X_n > 0\}\right)$$

$$\text{Ex} \leftarrow \geq \limsup_{n\to\infty} \mathbb{P}(\{X_n > 0\})$$

$$\overset{\text{Ex.}}{\underset{(11)}{\geq}} \frac{2p-1}{p} > 0$$

So the second moment method implies that the probability that the cluster size at 0 is infinite is positive $\forall\ p \in \left(\tfrac{1}{2}, 1\right]$

□