Week 2

Andrew Yao (姚期智)

2024-3-5

Random Variable and its Expectation

- A random variable X is a function X: U→R,

  its expectation is defined as $E(X)=\sum_{u \in U} p(u)X(u)$.

  * An "event" is a simple case when X takes on value 0/1.

  Definition: Sum of random variables

  For real a, b, define $Z=aX+bY$ by $Z(u) = aX(u)+bY(u)$.

- Essential Probability Tool #3 *Law of Linear Expectation:*

  If $X=C_1X_1+ C_2X_2+ \ldots C_nX_n$ *, then* $E(X)= \sum_i C_i E(X_i)$

  Proof. $E(X) = \sum_{u \in U} p(u)X(u) = \sum_{u \in U} p(u) \sum_i C_i X_i(u_i) = \sum_i C_i \sum_{u \in U} p(u)X_i(u)$.

- Example 1. Throw n coins X_i each of bias b, that is, $\Pr\{X_i =1\} = b$. Let $X = \sum_i X_i$

  (number of coins with outcome 1). By Linear Expectation, $E(X)= \sum_i E(X_i)= bn$

Random Variable and its Expectation (continued)

- Note that Linear Expectation $E(X) = \sum_i C_i E(X_i)$ holds even if the X_i are highly correlated: for example, when $\Pr\{X_i=1\ \forall i\} = \Pr\{X_i=0\ \forall i\} = ½$.

  Such generality is very useful in the analysis of algorithms; we look at an example.

- A permutation σ of {1, 2, …, n} can be represented in several different ways.

  For example, let n=5 and σ = (3 5 4 1 2):

1) σ as an array of length n, with σ [1]=3, σ [2] = 5 etc.

2) $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{pmatrix}$ as an 2 x n array

3) σ as a graph G_σ :

4) cycle representation: (2, 5) (3, 4, 1) is one such representation

   * normal-form cycle representation (1,3,4) (2,5) is unique

What's the expected number of cycles in a permutation?

Example 2. $\mathcal{P}$=(U, p) where U is the set of all n! permutations, p(σ)=1/|U| for all $\sigma \in$ U.

Let X be the random variable X(σ)= # of cycles in σ's cycle representation.

What is E(X)?

For each $1 \leq i \leq n$, let $L_i(\sigma)$ = length of cycle containing i.

- For example, σ = (3 5 4 1 2) has cycle representation (2, 5) (3, 4, 1),

then $L_1(\sigma) = 3$ and $L_5(\sigma) = 2$

Note that $\sum_{i=1}^{n} \frac{1}{L_i(\sigma)} = \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 2$ (the # cycles in σ)

- That is, $X = \sum_{i=1}^{n} 1/L_i$ as random variables.

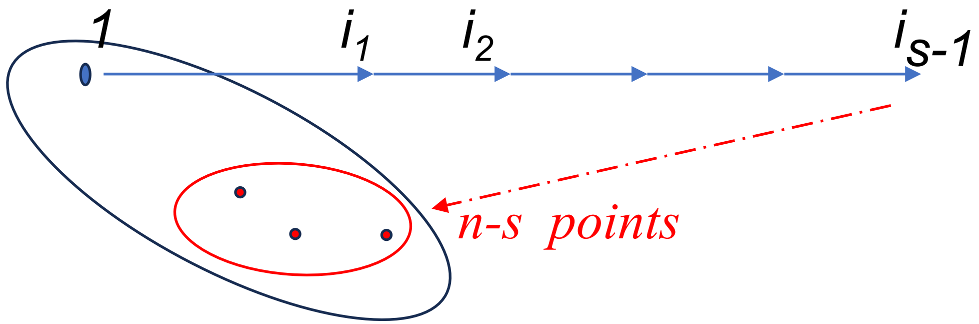
- By linearity of expectation, $E(X) = \sum_{i=1}^{n} E(1/L_i) = n\, E(1/L_1)$

Thus it remains to analyze $E(1/L_1)$.

Expected number of cycles in a permutation *(continued)*

Lemma $\Pr\{L_1 = s\} = \frac{1}{n}$ for any $s \in \{1, \ldots, n\}$.

Proof. Observe that for $1 \le s \le n$, $\Pr\{L_1 > s \mid L_1 > s-1\} = \frac{n-s}{n-s+1}$



By Chain rule, $\Pr\{L_1 = s\} = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-(s-1)}{n-(s-2)} \cdot \frac{1}{n-(s-1)} = \frac{1}{n}$

- Using the Lemma, we obtain

Harmonic number

$$E\left(\frac{1}{L_1}\right) = \sum_{s=1}^{n} \Pr\{L_1 = s\} \cdot \frac{1}{s} = \frac{1}{n}\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right) = \frac{1}{n} H_n$$

- Hence the expected number of cycles $E(X) = n\, E(1/L_1) = H_n$. *QED*

Example 3: *Finding my Pet*

n pets (tagged) are put in n random rooms, how can an owner find his/her pet?

- Allowed to open only $n/2$ doors! Hence success probability 1/2.
- But two owners can achieve better prob than 1/4 (without communication) using

Cycle Search: Person j starts with door j, and trace out a cycle

random permutation

Event A: *Person 1 succeeds, i.e. cycle* $|C_1| \le n/2$

Event B: 2 $|C_2| \le n/2$. Event T: $C_1 = C_2$

- Then, $\Pr\{A \cap B\} = \Pr\{A \cap B \cap T\} + \Pr\{A \cap B \cap \overline{T}\}$

Finding my Pet (continued)

Recall: Event A, B: $|C_1|, |C_2| \leq n/2$; Event T: $C_1 = C_2$

$\Pr\{A \cap B\} = \Pr\{A \cap B \cap T\} + \Pr\{A \cap B \cap \overline{T}\}$

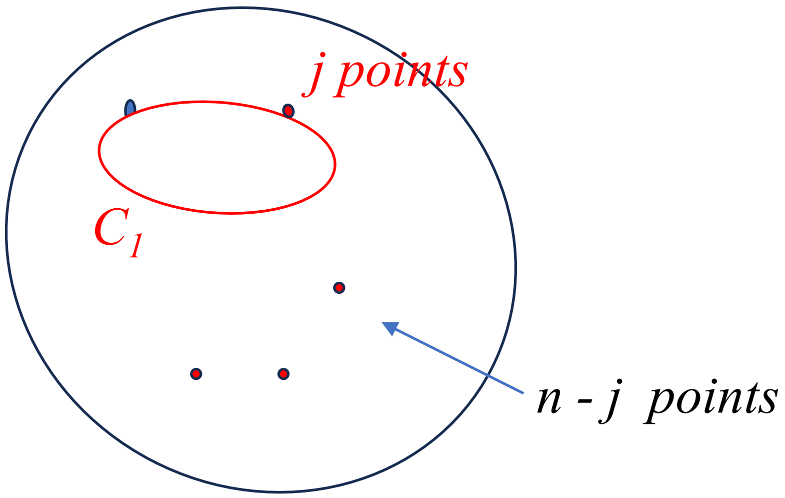
1) $\Pr\{A \cap B \cap T\} = \Pr\{A\} \cdot \Pr\{B \cap T \mid A\}$

$$= \sum_{j=1}^{n/2} \Pr\{|C_1| = j\} \cdot \Pr\{2 \epsilon C_1 \mid |C_1| = j\}$$

$$= \sum_{j=1}^{n/2} \frac{1}{n} \cdot \frac{j-1}{n-1}$$

$$= \frac{1}{n(n-1)} \cdot \frac{1}{2}\,\frac{n}{2}\left(\frac{n}{2} - 1\right)$$

$$= \frac{1}{4(n-1)} \cdot \left(\frac{n}{2} - 1\right) \approx \frac{1}{8}$$



2) Will show $\Pr\{A \cap B \cap \overline{T}\} \approx \frac{1}{4}$

Finding my Pet (continued)

Recall: Event A, B: $|C_1|, |C_2| \le n/2$; Event T: $C_1 = C_2$

2. $\Pr\{A \cap B \cap \overline{T}\} = \Pr\{A\} \cdot \Pr\{B \cap \overline{T} \mid A\}$

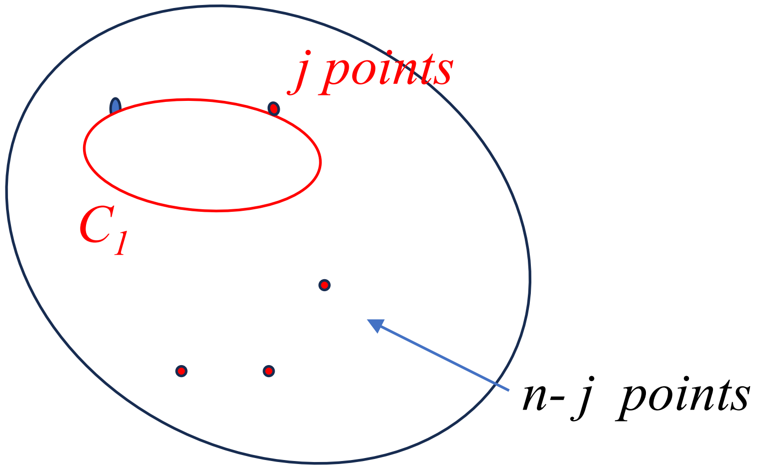
$$= \sum_{j=1}^{n/2} \frac{1}{n} \Pr\{2 \notin C_1, |C_2| \le \frac{n}{2} \;\mid\; |C_1| = j\}$$

$$= \sum_{j=1}^{n/2} \frac{1}{n} \Pr\{2 \notin C_1\} \cdot \Pr\{|C_2| \le \frac{n}{2} \;\mid\; |C_1| = j, 2 \notin C_1\}$$

$$= \frac{1}{n} \sum_{j=1}^{n/2} \frac{n-j}{n-1} \frac{n/2}{n-j}$$

$$= \frac{1}{n(n-1)} \cdot \frac{n}{2} \cdot \frac{n}{2}$$

$$= \frac{1}{4} \frac{n^2}{n(n-1)} \approx \frac{1}{4}$$



Putting together 1) and 2), we obtain $\Pr\{A \cap B\} = \frac{3n-2}{8(n-1)}$.

QED

More generally, what's the probability r_n for all n people to find their pets?

Clearly, this happens iff *permutation σ has no cycles of length > n/2.*

- Let T_j be the event that the longest cycle of σ has length j. Then by union bound,

$$r_n = \sum_{j > n/2}^{n} \frac{1}{n} \Pr\{T_j\} = \sum_{j > n/2}^{n} \frac{1}{n!} \binom{n}{j} (j-1)!\,(n-j)!$$

$$= \sum_{j > n/2}^{n} \frac{1}{n!} \frac{n!}{j!(n-j)!} (j-1)!\,(n-j)!$$

$$= \sum_{j > n/2}^{n} \frac{1}{j}$$

QED

- For large n,

$$r_n = H_n - H_{\lfloor n/2 \rfloor + 1} \approx \ln 2 \approx 38\%$$

What a surprise!

Here is another useful formula for computing expectation.

For any random variable X and event T, define *conditional expectation*

$$E(X \mid T) = \left(\sum_{u \text{ in } T} p(u)\, X(u)\right)/\Pr\{T\} \text{ if } \Pr(T)>0, \text{ and } 0 \text{ otherwise}$$

Essential Probability Tools #4:

Distributive Law for Expectation (Law of Total Expectation)

X: random variable

Universe U is the disjoint union of $W_1, W_2, \ldots, W_m$

Then $E(X) = \sum_i \Pr(W_i)\, E(X \mid W_i)$

More essential concepts from probability theory

- For a random variables X, E(X) alone may not be sufficient to make decisions.
- Consider a lottery ticket costing 50 ¥, whose payoff is a random variable X.
- Assume E(X)=100¥, is it reasonable to buy a ticket?

Scenario A: X = $\begin{cases} 50 & \textit{with prob} & 10\% \\ 100 & \textit{with prob} & 80\% \\ 150 & \textit{with prob} & 10\% \end{cases}$

Scenario B: X = $\begin{cases} 0 & \textit{with prob} & 0.9999999 \\ 10^9 & \textit{with prob} & 0.0000001 \end{cases}$

In both scenarios, E(X)=100. In case A, the value distribution is concentrated near E(X), thus E(X) reflects the behavior of X pretty well.

In case B, the value distribution is far away from E(X).

- The info E(X)=100 is not a good scientific basis to decide whether to buy a ticket.

More essential concepts from probability theory

- To help capture important information on X, besides E(X), we'd like to know how spread-out X is around the value E(X).

Definitions. *Variance of X:* $Var(X) = E((X-E(X))^2)$

Standard Deviation of X: $\sigma(X) = \sqrt{Var(X)}$

Fact: $Var(X) = E(X^2) - (E(X))^2$

proof. $Var(X) = E(X^2 - 2\,E(X) \cdot X + (E(X))^2)$

$$= E(X^2) - 2\,E(X)\,E(X) + (E(X))^2)$$

$$= E(X^2) - (E(X))^2$$

QED

- Var(X), or equivalently $\sigma(X)$, provides valuable info regarding how spread-out X is around E(X), as expressed in Chebyshev's Inequality below.
- In fact, we discuss also two other inequalities of this nature, applicable in various situations.

Essential Probability Tools #5 Tail Estimates:

-- Markov's, Chebyshev's and Chernoff's Inequalities

- *Markov' Inequality:*

  Let X be a random variable taking on only non-negative values.

  Then for any c > 0, Pr {X > cE(X)} < 1/c.

- Proof. If E(X) = 0, then X $\equiv 0$, the inequality follows.

  If E(X) > 0, then E(X) $= \sum_{s=1}^{n} p(\mathrm{u})X(u)$

  > Pr{X > cE(X)} · cE(X)

  Cancelling out E(X), we obtain *Markov' Inequality.* QED

- Var(X), or equivalently the standard deviation $\sigma(X)$, is often regarded as the second most important feature (next to E(X)) about X. We have the following

- *Chebyshev's Inequality:*

  For any c > 0, Pr {|X – E(X)| > c $\sigma(X)$} < 1/c^2.

Essential Probability Tools #5 Tail Estimates:

-- Markov's, Chebyshev's and Chernoff's Inequalities

- *Chebyshev's Inequality:*

  For any $c > 0$, $\Pr\{|X - E(X)| > c\,\sigma(X)\} < 1/c^2$.

- Proof. $Var(X) = E((X-E(X))^2)$.

  Applying Markov's Inequalities to $((X-E(X))^2$, we obtain

  $\Pr\{(X-E(X))^2 > c^2\, Var(X)\} < E((X-E(X))2)/c^2\, Var(X) = 1/c^2$.

  The left hand side is equal to $\Pr\{|X - E(X)| > c\,\sigma(X)\}$. QED

- In cases when $\sigma(X) << |E(X)|$ (e.g. the value distribution of X is clustered around E(X) within a narrow band of $6\sigma(X)$), then 误差 $1/6^2 \sim 3\%$

  Let's look at an example.

Example: *Throw n independent coins $X_1, X_2, \ldots X_n$ of bias $0<b<1$, that is,*

Pr $\{X_i = 1\} = b$, Pr $\{X_i = 0\} = 1 - b$. Let $X = \sum_{i=1}^{n} X_i$ = # of 1's among all the outcomes.

$$E(X) = \sum_{i=1}^{n} E(X_i) = nb$$

$$\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - (E(X))^2 \\
&= E((X_1 \ldots + X_n)^2) - b^2n^2 \\
&= E\left(\sum_i X_i^2 + \sum_{i \neq j} X_i X_j\right) - b^2n^2 \\
&= n\,E(X_1^2) + n(n-1)E(X_1X_2) - b^2n^2
\end{aligned}$$

As $X_1^2 = X_1$, and X_1, X_2 are independent, we have

$$\begin{aligned}
\mathrm{Var}(X) &= n\,E(X_1) + n(n-1)E(X_1)\,E(X_2) - b^2n^2 \\
&= nb + n(n-1)\,b^2 - b^2n^2 \\
&= b(1-b)\,n
\end{aligned}$$

$$\sigma(X) = \sqrt{b(1-b)\,n}$$

Essential Probability Tools #5 Tail Estimates (continued)

Chebyshev's Inequality: For any c > 0, $\Pr\{|X - E(X)| > c\,\sigma(X)\} < 1/c^2$.

- Very general, but often not tight:

For example, toss a fair coin=10,000 times, E(X)=5000, $\sigma(X) = \frac{1}{2}\sqrt{n} = 50$.

Here *Chebyshev's Inequality* only says $\Pr\{|X - 5000| > 500\} \leq 1/10^2 = 1\%$,

But it can actually be shown that this probability $< 2\,e^{-17}$

We now introduce a powerful method for establishing the above bound, known as the *Chernoff Bound*

Essential Probability Tools #5 Tail Estimates (continued)

Let $X_1, X_2, \ldots X_n$ be independent coin tosses, and $X = \sum_{i=1}^{n} X_i$ where

$$\begin{cases} \Pr(X_i = 1\} = b_i \\ \Pr(X_i = 0\} = 1 - bi \end{cases}$$

Note $E(X) = \sum_{i=1}^{n} b_i = \mu$

Theorem (Chernoff's Bound)

$$\Pr\{X \geq (1+\delta)\mu\} \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \quad \text{for } \delta > 0, \qquad (1)$$

$$\Pr\{X \leq (1-\delta)\mu\} \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \quad \text{for } 0 \leq \delta < 1. \qquad (2)$$

Corollary1:

$$\Pr\{X \geq (1+\delta)\mu\} \leq e^{-\frac{1}{3}\delta^2\mu} \quad \text{for } \delta > 0,$$

$$\Pr\{X \leq (1-\delta)\mu\} \leq e^{-\frac{1}{2}\delta^2\mu} \quad \text{for } 0 \leq \delta < 1.$$

Corollary2:

$$\Pr\{X \geq c\} \leq 2^{-c} \quad \text{if } c \geq 7E(X)$$

Proof of Chernoff's Bound

- We only prove (1) $\Pr\{X \geq (1+\delta)\mu\} \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$ for $\delta > 0$,

  (2) will be left as exercise.

  Recall $E(X) = \sum_{i=1}^{n} \mathrm{b}_i = \mu$

- Let t > 0 be a parameter. As the *exponential function* is monotone, we have

(*) $\Pr\{X > (1+\delta)\mu\} = \Pr\{\mathrm{e}^{t\,X} > \mathrm{e}^{t\,(1+\delta)u}\} \leq \frac{E(\mathrm{e}^{t\,X})}{e^{t(1+\delta)\mu}}$ by Markov Inequality.

- We take two steps:

  Step 1: Get an explicit expression for $E(\mathrm{e}^{t\,X})$

  Step 2: Choose t optimally to get best (i.e. smallest) upper bound.

- Step 1: $E(e^{tx}) = E(e^{t\,\Sigma_i X_i}) = \Pi_i E(e^{t\,X_i})$

  $= \Pi_i(1+ b_i(e^t - 1)) \leq \Pi_i(e^{b_i(e^t-1)})$

  $= (e^{(e^t-1)\,\Sigma_i b_i}) \leq e^{\mu\,(e^t-1)}$

$e^{\mu\,(e^t-1)--t(1+\delta)u}$

Proof of Chernoff's Bound (continued)

Step 2: Pick $t = t_0$ to minimize $f(t) \equiv \mu\,(e^{t}-1) - t(1+\delta)u$

Answer: $t_0 = \ln(1+\delta)$ (homework problem)

[satisfying $f'(t_0)=0$, $f''(t_0) \geq 0$]

- $f(t_0) = \mu\,(e^{\ln(1+\delta)} - 1) - \mu\,(\ln(1+\delta))(1+\delta)$

  $= \mu\,\delta - \mu\,(1+\delta)(\ln(1+\delta))$

- Thus, $\Pr\{X > (1+\delta)\mu\} \leq e^{f(t_0)} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$ QED

Hoeffding's Inequality: If value of $X_i \in (a, b)$,

$$\Pr\{\,|\Sigma X_i - E(X)| \geq t\,\} \leq exp\left(\frac{-2t^2}{n(b-a)^2}\right)$$

For X_i i.i.d. and $t=\varepsilon n$: $\leq exp\left(\frac{-2\varepsilon^2 n}{(b-a)^2}\right)$

- We have finished presenting some essential tools of probability theory.
- Will look at some interesting research results obtained with these tools.

Greedy Clique Algorithm A

Input: a random graph G=(V, E), V={1, 2, …, n} and E a set of edges.

Step 1. S ← {1}

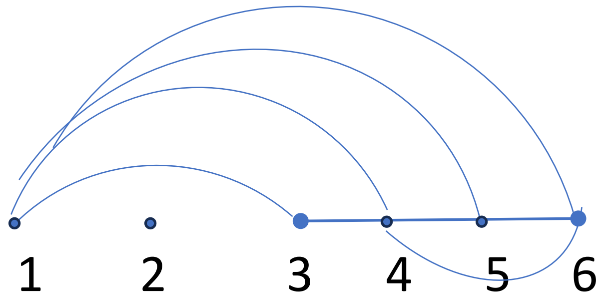
Step 2. for i = 2 to n:

    if {i, j} ∈ E for all j ∈ S

        then S ← S U {i }

    i ← i+1

Output: *Clique A(G)= S*



A outputs {1,3,4}, but max clique is {1,4,5,6}

Theorem For an input random graph on n vertices, the greedy algorithm returns a clique A(G) of size $\log_2 n - \log_2\log_2 n \le |A(G)| \le \log_2 n + \log_2\log_2 n$ with probability 1– o(1).

* Notation: o(1) stands for a function f(n) such that f(n) → 0 as n → ∞.

- We first prove the upper bound for |A(G)|.

<u>Upper bound</u> $\Pr\{|A(G)| > \log_2 n + \log_2\log_2 n\} = o(1)$

proof. Let $K = \log_2 n + \log_2\log_2 n$. For $2 \le i \le n$, let T_i be the event that the greedy algorithm selects vertex i as the K-th vertex to join S. Then by distributive law,

$$\Pr\{|S|>K\} = \sum_{2\le i\le n} \Pr\{T_i\} \cdot \Pr\{|S|>K \mid T_i\}$$

As only n-i vertices are available to extend S by 1, we have for each i by union bound

$$\Pr\{|S|>K \mid T_i\} \le (n-i)\frac{1}{2^K} \le \frac{n}{2^K}$$

We have thus

$$\Pr\{|S|>K\} \le \frac{n}{2^K}\sum_{2\le i\le n} \Pr\{T_i\} \le \frac{n}{2^{\ln n+\ln\ln n}} = \frac{1}{\ln n}$$

$$= o(1)$$

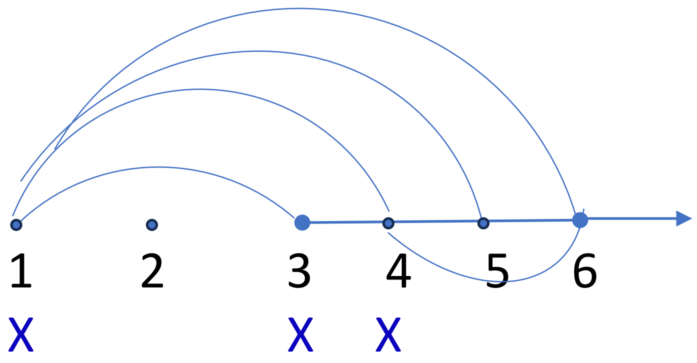
QED

➢ We next prove the lower bound for $|A(G)|$.

Lower bound Pr $\{|A(G)| \geq K^{\sim}\} = 1-o(1)$ where $K^{\sim} = \log_2 n - \log_2\log_2 n$.

proof. For the purpose of analysis, consider running the greedy algorithm on an infinite sequence of vertices $\{1, 2, 3, \ldots\}$. Let $X_m(G)$ be the m-th vertex of G selected to join the clique.

➢ It turns out easy to characterize $Y_m \equiv X_{m+1} - X_m$, the *gap* that happens between two successive vertices chosen for the clique.



Observe: for any $j \geq 1$, Y_j is an independent geometric random variable with parameter $b_j = \frac{1}{2^j}$. That is, $\Pr\{Y_j = t\} = (1-b_j)^{t-1} b_j$ for all integers $t \geq 1$.

Lemma 1 $X_m = 1 + \sum_{1 \leq j \leq m-1} Y_j$ for all $m \geq 2$.

➢ Note that $A(G) \geq K^{\sim}$ if and only if $X_{K^{\sim}}(G) \leq n$. Hence our lower bound is to prove $\Pr\{X_{K^{\sim}}(G) \leq n\} = 1-o(1)$. Suffices to show $\Pr\{\sum_{1 \leq j \leq K^{\sim}} Y_j \leq n-1\} = 1-o(1)$.

- Prove lower bound $K^{\sim}$ for $|A(G)|$ by showing $\Pr\{\sum_{1\le j\le K^{\sim}} Y_j \le n-1\} = 1 - o(1)$.
- Denote $\sum_{1\le j\le K^{\sim}} Y_j$ by X', and estimate $E(X')$ and $Var(X')$:

Lemma 2 $E(X') \le \dfrac{2\,n}{\log_2 n}$

proof. Each geometric distribution Y_j satisfies $E(Y_j) = \dfrac{1}{b_j} = 2^j$. By Linear Expectation, $E(X') = \sum_{1\le j\le K^{\sim}} 2^j = 2^{1+K^{\sim}} - 2 \le \dfrac{2\,n}{\log_2 n}$.

Lemma 3 $Var(X') \le 2\left(\dfrac{n}{\log_2 n}\right)^2$

proof. It is well known $Var(Y_j) = \dfrac{1}{b_j^2} - \dfrac{1}{b_j} = 4^j - 2^j$. As Y_j's are independent,

$$Var(X') = \sum_{1\le j\le K^{\sim}} Var(Y_j) = \sum_{1\le j\le K^{\sim}} (4^j - 2^j) = \frac{4}{3}(4^{K^{\sim}} - 1) - (2^{1+K^{\sim}} - 2) \le 2\left(\frac{n}{\log_2 n}\right)^2 .$$

- We are now ready to plug the above estimates for $E(X')$ and $Var(X')$ into Chebyshev's inequality.

- If X' > n-1, then $X' - E(X') > n - 1 - \frac{2n}{\log_2 n} > \frac{n}{2}$ by Lemma 2.

  It implies that $\Pr\{X' > n-1\} \leq \Pr\{X' - E(X') > \frac{n}{2}\}$ (1)

- On the other hand, Chebyshev's inequality tells us

$$\Pr\{X' - E(X') > \frac{n}{2}\} \leq \frac{\mathrm{Var}(X')}{(\frac{n}{2})^2} \leq 8\frac{1}{(\log_2 n)^2} \quad \text{by Lemma 3.} \qquad (2)$$

- It follows from (1) and (2) that

$$\Pr\{X' > n-1\} \leq 8\frac{1}{(\log_2 n)^2} = o(1)$$

This proves the lower bound. QED

Open Problem: Design an efficient algorithm (i.e. polynomial running time) that,

For a random n-vertex graph G, outputs a clique of size > c $\log_2$ n with prob.

1-o(1) where c > 1.

End