CS5314
Randomized Algorithms

Lecture 19: Probabilistic Method
(2nd Moment, Condition Expectation Inequality)

Objectives

- Introduce two further techniques (apart from Counting, Expectation, Sample-and-Modify) to show (non-)existence of a certain object

Second Moment Method

– Based on the Chebyshev inequality

Conditional Expectation Inequality

– Based on Binomial RV

Second Moment Method

The following is the core of this method :

Theorem: If X is a nonnegative integer-valued random variable, then

$$\Pr(X = 0) \leq \mathrm{Var}[X] \,/\, (E[X])^2$$

Proof: $\Pr(X = 0) \leq \Pr(|X - E[X]| \geq E[X])$

$$\leq \mathrm{Var}[X] \,/\, (E[X])^2$$

Remarks: (1) if RHS < 1 ➔ there is an object with $X > 0$;

(2) if RHS ≈ 0 ➔ a random object almost always has $X > 0$

Will K_4 exist in $G_{n,p}$?

Next, we shall show a bound f such that :

1. when $p \ll f$, a random graph from $G_{n,p}$ does not contain a 4-Clique (K_4) w.h.p.
2. when $p \gg f$, a random graph from $G_{n,p}$ contains a K_4 w.h.p.

We call f : threshold function for K_4 to occur in $G_{n,p}$

Will K_4 exist in $G_{n,p}$? (2)

Theorem: Suppose $p = o(n^{-2/3})$. Let G be a random graph from $G_{n,p}$. Then, for any $\varepsilon > 0$ and sufficiently large n,

$$\Pr(G \text{ contains } K_4) < \varepsilon$$

How to prove??

(By Basic Counting / Expectation Argument)

Proof

Let X_j be an indicator such that:

$X_j = 1$ if j^{th} subset of four vertices is a K_4 in G

$X_j = 0$ otherwise

➔ $E[X_j] = Pr(X_j = 1) = p^6$

Let X denote the number of K_4 in G

➔ $E[X] = C(n,4)\, p^6 = o(n^4\, n^{-4}) = o(1)$

Proof (2)

This implies that for large enough n,

$$E[X] < \varepsilon$$

Since X is a non-negative integer, we have :

$$E[X] = \Sigma_{j = 1,2, \ldots} j \Pr(X = j)$$

$$\geq \Sigma_{j = 1,2, \ldots} \Pr(X = j)$$

$$= \Pr(X \geq 1)$$

➔ Theorem follows

Will K_4 exist in $G_{n,p}$? (3)

Theorem: Suppose that $p = \omega(n^{-2/3})$. Let G be a random graph from $G_{n,p}$. Then, for any $\varepsilon > 0$ and sufficiently large n,

$$\Pr(G \text{ does not contain } K_4) < \varepsilon$$

Let X = # of K_4 in G

Proof Idea: By Second Moment Method

Compute $E[X]$ and $Var[X]$

Before that, we introduce a simple result …

A Simple Result

Lemma: Let $Y_1, Y_2, \ldots, Y_m$ be **m** indicators, and $Y = Y_1 + Y_2 + \ldots + Y_m$. Then,

$$\mathrm{Var}[Y] \leq E[Y] + \sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j)$$

Proof: Since Y_j is an indicator, $E[Y_j^2] = E[Y_j]$

$$\mathrm{Var}[Y_j] = E[Y_j^2] - (E[Y_j])^2$$
$$\leq E[Y_j^2] = E[Y_j]$$

The lemma thus follows, since

$$\mathrm{Var}[Y] = \sum_j \mathrm{Var}[Y_j] + \sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j)$$

Back to the Proof …

Let X_j be an indicator such that:

$X_j = 1$ if the j^{th} subset of four vertices is a K_4 in G

$X_j = 0$ otherwise

- We wish to bound $Var[X] = Var[\Sigma X_j]$ and apply second moment method
  - By the previous lemma, we can first consider the values of $Cov(X_i, X_j)$

Back to the Proof … (2)

- The value of Cov(X_i, X_j) depends on whether the i^{th} subset of four vertices share any vertex with the j^{th} subset

There are four cases:

Case 0: They share no vertex

Case 1: They share 1 vertex

Case 2: They share 2 vertices

Case 3: They share 3 vertices

Back to the Proof … (3)

For Case 0 (they share no vertex) :

➔ X_i and X_j are independent

➔ $\mathrm{Cov}(X_i, X_j) = 0$

For Case 1 (they share 1 vertex) :

➔ X_i and X_j are independent (why?)

➔ $\mathrm{Cov}(X_i, X_j) = 0$

Back to the Proof … (4)

For Case 2 (they share 2 vertices) :

$$E[X_iX_j] = Pr(X_iX_j = 1)$$
$$= Pr(X_i = 1 \mid X_j = 1)\ Pr(X_j = 1)$$
$$= p^5 * p^6 = p^{11}$$

$$\rightarrow \ Cov(X_i, X_j) = E[X_iX_j] - E[X_i]\,E[X_j]$$
$$\leq E[X_iX_j] = p^{11}$$

Back to the Proof … (5)

For Case 3 (they share 3 vertices),

$$E[X_iX_j] = Pr(X_iX_j = 1)$$
$$= Pr(X_i = 1 \mid X_j = 1)\, Pr(X_j = 1)$$
$$= p^3 * p^6 = p^9$$

➔ $Cov(X_i, X_j) = E[X_iX_j] - E[X_i]\,E[X_j]$
$$\leq E[X_iX_j] = p^9$$

Back to the Proof … (6)

➔ $\mathrm{Var}[X] \leq E[X] + \Sigma_{i \neq j} \mathrm{Cov}(X_i, X_j)$

$= C(n,4)\, p^6 + \Sigma_{i \neq j} \mathrm{Cov}(X_i, X_j)$

$\leq C(n,4)\, p^6 + (\#Case2)\, p^{11} + (\#Case3)\, p^9$

$= C(n,4)\, p^6 + O(C(n,6)\, p^{11}) + O(C(n,5)\, p^9)$

$= o(n^8 p^{12})$ … [why?? Recall: $p = \omega(n^{-2/3})$]

On the other hand, since

$(E[X])^2 = (C(n,4)\, p^6)^2 = \Theta(n^8 p^{12})$

➔ $\mathrm{Var}[X] / (E[X])^2 = o(1)$ ➔ theorem follows

Conditional Expectation Inequality

- Let X be a random variable such that

  $X > 0 \Leftrightarrow$ a certain object exists

  E.g., X = # of K_4 in a graph G chosen from $G_{n,p}$

  ➔ In this case, $X > 0$ implies the existence of K_4 in G

- If X can be expressed as a sum of indicators (which is true in many situations), we can usually get a simpler proof of existence via the next theorem

Conditional Expectation Inequality

Lemma: Let $X_1, X_2, \ldots, X_n$ be n indicators, and $X = X_1 + X_2 + \ldots + X_n$. Then,

$$\Pr(X > 0) \geq \sum_{j=1 \text{ to } n} \left(\Pr(X_j = 1) \,/\, E[X \mid X_j = 1] \right)$$

Note: We do not require X_j's to be independent

Proof: Let $Y = 1/X$ if $X > 0$

$Y = 0$ if $X = 0$

Note: XY is an indicator random variable!!

→ $\Pr(X > 0) = \Pr(XY = 1) = E[XY] \ldots$

Proof (cont)

So, $\Pr(X > 0) = E[XY] = \sum_{j=1 \text{ to } n} E[X_jY]$

$= \sum_{j=1 \text{ to } n} \big(E[X_jY \mid X_j = 1] \Pr(X_j = 1) +$

$E[X_jY \mid X_j = 0] \Pr(X_j = 0) \big)$

$= \sum_{j=1 \text{ to } n} \big(E[Y \mid X_j = 1] \Pr(X_j = 1) \big)$ ... [why?]

$= \sum_{j=1 \text{ to } n} \big(E[1/X \mid X_j = 1] \Pr(X_j = 1) \big)$

$\geq \sum_{j=1 \text{ to } n} \big(\Pr(X_j = 1) / E[X \mid X_j = 1] \big)$ ... [Jensen]

→ This completes the proof

Existence of K_4 in $G_{n,p}$ (revisited)

- Now, we revisit the theorem in Page 8 and give a simpler proof
- Recall: X = # of K_4 in a random graph G chosen from from $G_{n,p}$

And X_j be an indicator such that:

$X_j = 1$ if the j^{th} subset of four vertices is a K_4 in G

$X_j = 0$ otherwise

➔ $X = X_1 + X_2 + \dots + X_{C(n,4)}$

Existence of K_4 in $G_{n,p}$ (revisited)

First, recall that $\Pr(X_j = 1) = p^6$

- In order to apply conditional expectation inequality to prove existence of K_4 in $G_{n,p}$, we want to bound $E[X \mid X_j = 1]$

By linearity of expectation, we have

$$E[X \mid X_j = 1] = \sum_{k=1 \text{ to } C(n,4)} E[X_k \mid X_j = 1]$$

$$= \sum_{k=1 \text{ to } C(n,4)} \Pr(X_k = 1 \mid X_j = 1) \qquad \text{... [why?]}$$

Existence of K_4 in $G_{n,p}$ (revisited)

Question: What is $Pr(X_k = 1 \mid X_j = 1)$?

Ans. ... depends on the number of vertices shared by j^{th} and k^{th} subset

	value of Pr	# of k's
Share 0 vertex:	p^6	$C(n-4,4)$
Share 1 vertex:	p^6	$4\ C(n-4,3)$
Share 2 vertices:	p^5	$6\ C(n-4,2)$
Share 3 vertices:	p^3	$4\ C(n-4,1)$
Share 4 vertices:	1	1

Existence of K_4 in $G_{n,p}$ (revisited)

Thus, $E[X \mid X_j = 1]$

$$= \sum_{k=1 \text{ to } {}^nC_4} \Pr(X_k = 1 \mid X_j = 1)$$

$$= \; p^6 \times C(n-4,4) + p^6 \times 4\, C(n-4,3)$$

$$+\, p^5 \times 6\, C(n-4,2) + p^3 \times 4\, C(n-4,1) + 1$$

→ As $n \to \infty$ and $p = \omega(n^{-2/3})$

$$\Pr(X > 0) \geq \sum_{j=1 \text{ to } C(n,4)} \left(\Pr(X_j = 1) / E[X \mid X_j = 1]\right)$$

$$= C(n,4)\, p^6 / E[X \mid X_j = 1] \approx 1$$

→ This completes the proof of the theorem