Lecture 5: "The Second Moment Method"

Sotiris Nikoletseas
Professor

Computer Engineering and Informatics Department
2020 - 2021

Summary of previous lecture

Variation of Linearity of Expectation method: Deletion Method

- Method's basic idea: prove existence of a structure "similar" to the desired one and modify it accordingly.
- Examples
  1. An improved lower bound for Ramsey numbers.
  2. Independent sets (Turán's Theorem, 1941).

Summary of this lecture

The Second Moment

i. The Variance of a random variable
ii. The Chebyshev Inequality
iii. The Second Moment method
iv. Covariance
v. Alternative techniques of estimation of the variance of a sum of indicator variables.
vi. Example - Cliques of size 4 in random graphs.

Variance

Variance:

- is the most vital statistic of a r.v. beyond expectation.
- is defined as $Var[X] = E\left[(X - E[X])^2\right]$
- properties:
  - $Var(X) = E[X^2] - E^2[X]$
  - $Var(cX) = c^2 Var(X)$, c constant
  - X, Y independent $\Rightarrow Var[X + Y] = Var[X] + Var[Y]$

Standard deviation:

$$\sigma = \sqrt{Var[X]} \Rightarrow Var[X] = \sigma^2$$

Chebyshev Inequality

Theorem 1 (Chebyshev Inequality)

Let X be a random variable with expected value μ. Then for any $t > 0$:

$$\Pr\left[|X-\mu| \geq t\right] \leq \frac{Var[X]}{t^2}$$

Proof:

$$\Pr[|X-\mu| \geq t] = \Pr\left[(X-\mu)^2 \geq t^2\right]$$

$$\underset{Markov}{\leq} \frac{E\left[(X-\mu)^2\right]}{t^2} = \frac{Var[X]}{t^2}$$

□

Chebyshev Inequality

Alternative Proof:

$$Var[X] = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 \Pr\{X=x\}$$

$$\geq \sum_{|x-\mu|\geq t} (x-\mu)^2 \Pr\{X=x\}$$

$$\geq \sum_{|x-\mu|\geq t} t^2 \Pr\{X=x\}$$

$$= t^2 \sum_{|x-\mu|\geq t} \Pr\{X=x\} = t^2 \Pr\{|X-\mu|\geq t\}$$

$$\Rightarrow \Pr\{|X-\mu|\geq t\} \leq \frac{Var[X]}{t^2}$$

□

Chebyshev Inequality - application

if $t = \sigma$ then $\Pr[|X - \mu| \geq \sigma] \leq \frac{\sigma^2}{\sigma^2} = 1$ (trivial bound)
if $t = 2\sigma$ then $\Pr[|X - \mu| \geq 2\sigma] \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4}$
$\vdots$
if $t = k\sigma$ then $\Pr[|X - \mu| \geq k\sigma] \leq \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$

In other words, this inequality bounds the concentration of a random variable around its mean.
A small variance implies high concentration.

The Second Moment Method

Theorem 2

For any random variable X it holds that:

if $E[X] \to \infty$ *and* $Var[X] = o(E^2[X])$ *then* $\Pr\{X = 0\} \to 0$

Proof: Since

$$|X - E[X]| \geq E[X] \Rightarrow \begin{cases} X \geq 2E[X] & \text{or} \\ X \leq 0 \end{cases}$$

$$\Pr\{X = 0\} \leq \Pr\{|X - E[X]| \geq E[X]\} \underset{t=E[X]}{\leq} \frac{Var[X]}{E^2[X]}$$

if $\dfrac{Var[X]}{E^2[X]} \to 0 \Leftrightarrow Var[X] = o(E^2[X])$ then $\Pr\{X = 0\} \to 0$ □

So, we need to estimate the variance. Actually, we need to properly bound it in terms of the mean.

Covariance

Covariance

Let X and Y be random variables. Then

$$\mathrm{Cov}(X,Y) = E[XY] - E[X] \cdot E[Y]$$

Remark:

- Covariance is a measure of association between two random variables.
- $\mathrm{Cov}(X,X) = Var[X]$
- if X, Y are independent r.v. then $\mathrm{Cov}(X,Y) = 0$
- $|Cov(X,Y)| \uparrow \quad \Rightarrow$ stochastic dependence of $X, Y \uparrow$

Theorem 3

Consider a sum of n random variables $X = X_1 + X_2 + \cdots + X_n$*. It holds that:*

$$Var[X] = \sum_{1 \le i,j \le n} Cov(X_i, X_j)$$

Remark: The sum is over ordered pairs, i.e. we take both $Cov(X_i, X_j)$ and $Cov(X_j, X_i)$.

Proof of theorem 3

The proof is by induction on n.
We show the case $n = 2$:

$$\sum_{1 \le i,j \le 2} Cov(X_i, X_j) = Cov(X_1, X_1) + Cov(X_1, X_2) +$$

$$+ Cov(X_2, X_1) + Cov(X_2, X_2) =$$

$$E[X_1^2] - E^2[X_1] + E[X_1 X_2] - E[X_1]E[X_2] + E[X_2 X_1] - E[X_2]E[X_1] +$$

$$+ E[X_2^2] - E^2[X_2] =$$

$$= E[X_1^2] + E[X_2^2] + 2E[X_1 X_2] - (E^2[X_1] + E^2[X_2] + 2E[X_1]E[X_2]) =$$

$$= E\left[X_1^2 + X_2^2 + 2X_1 X_2\right] - (E[X_1] + E[X_2])^2$$

$$= E\left[(X_1 + X_2)^2\right] - E^2\left[(X_1 + X_2)\right] =$$

$$= Var[X_1 + X_2]$$

Theorem 4

Let X_i $1 \leq i \leq n$ be indicator random variables.

$$X_i = \begin{cases} 1 & p_i \\ 0 & 1 - p_i \end{cases}$$

Let X be their sum: $X = X_1 + X_2 + \cdots + X_n$.
It holds that:

$$Var[X] \leq E[X] + \sum_{1 \leq i \neq j \leq n} Cov(X_i, X_j)$$

Proof:
$Var[X] = \sum_{1 \leq i,j \leq n} Cov(X_i, X_j)$

$\mathrm{Cov}(X_i, X_i) = E[X_i X_i] - E[X_i]E[X_i] = E\left[(X_i)^2\right] - E^2[X_i] = Var[X_i]$

Covariance

Proof of theorem 4

$$Var[X_i] = (1-p_i)^2 \cdot p_i + (0-p_i)^2 \cdot (1-p_i) = p_i(1-p_i) \leq p_i = E[X_i]$$

$$\begin{aligned}
Var[X] &= \sum_{1\leq i\leq n} Cov(X_i, X_i) + \sum_{1\leq i\neq j\leq n} Cov(X_i, X_j) \\
&= \sum_{1\leq i\leq n} Var[X_i] + \sum_{1\leq i\neq j\leq n} Cov(X_i, X_j) \\
&\leq \sum_{1\leq i\leq n} E[X_i] + \sum_{1\leq i\neq j\leq n} Cov(X_i, X_j) \\
&= E[X] + \sum_{1\leq i\neq j\leq n} Cov(X_i, X_j)
\end{aligned}$$

$\square$

Bounding the Variance

- Suppose that $X = X_1 + X_2 + \cdots + X_n$ where X_i is the indicator r.v. for event A_i.
- For indices i, j we define the operator $\sim$ and write $i \sim j$ if $i \neq j$ and the events A_i and A_j are not independent. (non-trivial dependence)
- We define

$$\Delta = \sum_{i \sim j} \Pr\{A_i \wedge A_j\}$$

The sum is over ordered pairs.

$$\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] \leq E[X_i X_j] = \Pr\{A_i \wedge A_j\}$$

$$\Rightarrow Var[X] \leq E[X] + \Delta$$

The Basic Theorem

Theorem 5

If $E[X] \to \infty$ *and* $\Delta = o(E^2[X])$ *then* $\Pr\{X = 0\} \to 0$

Proof:

$$\Pr\{X = 0\} \leq \frac{Var[X]}{E^2[X]} \leq \frac{E[X] + \Delta}{E^2[X]} = \frac{1}{E[X]} + \frac{\Delta}{E^2[X]} \to 0$$

□

A variation (I)

Symmetric events:
Events A_i and A_j are symmetric if and only if

$$Pr\{X_i|X_j = 1\} = \Pr\{X_j|X_i = 1\}$$

- In other words, the conditional probability of a pair of events is independent of the "order" of conditioning.
- Symmetry applies in almost all graphotheoretical properties because of symmetry of corresponding subgraphs which are set of vertices (i.e. the conditioning affects the intersection and depends on its size).

A variation (II)

We define

$$\Delta^* = \sum_{j \sim i} \Pr\{A_j | A_i\}$$

Lemma: $\Delta = \Delta^* \cdot E[X]$

Proof:

$$\Delta = \sum_{i \sim j} \Pr\{A_i \wedge A_j\} = \sum_{i \sim j} \Pr\{A_i\} \Pr\{A_j | A_i\}$$

$$= \sum_{i} \sum_{j \sim i} \Pr\{A_i\} \Pr\{A_j | A_i\}$$

$$= \sum_{i} \Pr\{A_i\} \sum_{j \sim i} \Pr\{A_j | A_i\}$$

$$= \Delta^* \cdot \sum_{i} \Pr\{A_i\}$$

$$\Rightarrow \Delta = \Delta^* \cdot E[X]$$

The basic theorem of the variation

Change of previous theorem's condition:

$$\Delta = o(E^2[X])$$

$$\Leftrightarrow \Delta^* \cdot E[X] = o(E^2[X])$$

$$\Leftrightarrow \Delta^* = o(E[X])$$

Theorem 6

If $E[X] \to \infty$ *and* $\Delta^* = o(E[X])$ *then* $\Pr\{X = 0\} \to 0$

Threshold functions in $G_{n,p}$

Definition 1

$p_o = p_o(n)$ is a threshold of property A iff

- $p >> p_o \Rightarrow \Pr\{G_{n,p}$ *has the property* A $\} \to 1$
- $p << p_o \Rightarrow \Pr\{G_{n,p}$ *has the property* A $\} \to 0$

Typical thresholds:

- giant component: $\frac{c}{n}$ (c constant)
- connectivity: $\frac{c \log n}{n}$
- hamiltonicity: $\frac{c \log n}{n}$

Example

Existence of complete subgraph of size 4 in $G_{n,p}$

Theorem 7

Let A be the property of existence of K_4 cliques in $G_{n,p}$. The threshold function for A is $p_o(n) = n^{-2/3}$.

Proof:

- Let S be any fixed set of 4 vertices.
- Define r.v. X that counts the number of cliques of size 4.
- $X = \sum_{S,|S|=4} X_S$ where X_S is an indicator variable:

$$X_S = \begin{cases} 1 & \text{S is clique} \\ 0 & \text{otherwise} \end{cases}$$

- $E[X_S] = p^6$

Proof of theorem 7

- By Linearity of expectation

$$E[X] = E\left[\sum_{S,|S|=4} X_S\right] = \sum_{S,|S|=4} E[X_S] = \binom{n}{4} p^6 \sim n^4 p^6$$

- $E[X] = n^4 p^6 << 1 \Leftrightarrow p << n^{-2/3}$
  - If $p << n^{-2/3} \Rightarrow E[X] \to 0 \Rightarrow$ non-existence w.h.p.
  - Also, clearly $p >> n^{-2/3} \Rightarrow E[X] \to \infty$.

Proof of theorem 7

- All the X_S are symmetric and so, these values $p >> n^{-2/3}$ must satisfy $\Delta^* = o(E[X])$ where $\Delta^* = \sum_{j \sim i} \Pr\{A_j | A_i\}$. The event A_i is defined as "the set S_i is a clique of size 4"
- $j \sim i$ means that A_i, A_j are not independent and $i \neq j$
- Here, $A_j \sim A_i$ if and only if A_j and A_i have common edges (but less than four edges).
- So, $A_j \sim A_i$ if and only if $|S_i \cap S_j| = 2$ or 3.

Proof of theorem 7

1. $|S_i \cap S_j| = 2$
   - There is only 1 common edge $\Rightarrow \Pr\{A_j|A_i\} = p^5$
   - There are $\binom{4}{2}\binom{n-4}{2} = O(n^2)$ different ways to choose the set S_j such that $|S_i \cap S_j| = 2$.
2. $|S_i \cap S_j| = 3$
   - There are 3 common edges so $\Pr\{A_j|A_i\} = p^3$
   - There are $\binom{4}{3}\binom{n-4}{1} = O(n)$ different ways to choose the set S_j such that $|S_i \cap S_j| = 3$.

$$\Delta^* = \sum_{2 \leq |S_i \cap S_j| \leq 3} \Pr\{A_j|A_i\} = \sum_{|S_i \cap S_j| = 2} \Pr\{A_j|A_i\} + \sum_{|S_i \cap S_j| = 3} \Pr\{A_j|A_i\}$$

$$\sim n^2 \cdot p^5 + n \cdot p^3$$

Proof of theorem 7

When $p = n^{-2/3}$ then:

$$\frac{\Delta^*}{E[X]} \sim \frac{n^2 \cdot p^5 + n \cdot p^3}{n^4 \cdot p^6} = \frac{1}{n^2 \cdot p} + \frac{1}{n^3 \cdot p^3} = \frac{1}{n^{\frac{4}{3}}} + \frac{1}{n} \to 0$$

So, indeed, for that value of p we have

$$\Delta^* = o(E[X])$$

and a K_4 exists w.h.p.

- This, obviously holds for larger p values too, because of monotonicity. □