Lecture 6: Random Networks III - The Second Moment Method

Prof. Sotiris Nikoletseas

University of Patras
and CTI

ΥΔΑ ΜΔΕ, Patras
2019 - 2020
Summary of this lecture

The Second Moment

1. The Variance of a random variable
2. The Chebyshev Inequality
3. The Second Moment method
4. Covariance
5. Alternative techniques of estimation of the variance of a sum of indicator variables.
6. Example - Cliques of size 4 in random graphs.
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:

$$\sigma = \sqrt{Var[X]} \Rightarrow Var[X] = \sigma^2$$
Theorem 1 (Chebyshev Inequality)

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

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

Proof:

$$\Pr[|X - \mu| \geq t] = \Pr[(X - \mu)^2 \geq t^2] \leq \frac{E[(X - \mu)^2]}{t^2} = \frac{\text{Var}[X]}{t^2}$$

$\square$
Chebyshev Inequality

Alternative Proof:

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

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

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

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

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

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

**Proof:** Since

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

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

if $\frac{\text{Var}[X]}{E^2[X]} \to 0 \iff \text{Var}[X] = o(E^2[X])$ then $\Pr\{X = 0\} \to 0$ \(\Box\)

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

Let $X$ and $Y$ be random variables. Then

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

Remark:

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

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

Remark: The sum is over ordered pairs, i.e. we take both \( \text{Cov}(X_i, X_j) \) and \( \text{Cov}(X_j, X_i) \).
Proof of theorem 3

The proof is by induction on \( n \).

We show the case \( n = 2 \):

\[
\sum_{1 \leq i, j \leq 2} \text{Cov}(X_i, X_j) = \text{Cov}(X_1, X_1) + \text{Cov}(X_1, X_2) + \text{Cov}(X_2, X_1) + \text{Cov}(X_2, X_2) = \\
E[X_1^2] - E^2[X_1] + E[X_1 X_2] - E[X_1]E[X_2] + E[X_2 X_1] - E[X_2]E[X_1] + \\
+ E[X_2^2] - E^2[X_2] = \\
= E[X_1^2] + E[X_2^2] + 2E[X_1 X_2] - (E^2[X_1] + E^2[X_2] + 2E[X_1]E[X_2]) = \\
= E \left[ X_1^2 + X_2^2 + 2X_1 X_2 \right] - \left( E[X_1] + E[X_2] \right)^2 \\
= E \left[ (X_1 + X_2)^2 \right] - E^2 \left[ (X_1 + X_2) \right] = \\
= \text{Var}[X_1 + X_2]
\]
Covariance
An upper bound of the sum of indicator r.v.

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:

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

Proof:

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

$$\text{Cov}(X_i, X_i) = \text{E}[X_iX_i] - \text{E}[X_i]\text{E}[X_i] = \text{E}[(X_i)^2] - \text{E}^2[X_i] = \text{Var}[X_i]$$
Proof of theorem 4

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

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

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

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

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

\[
\Delta = \sum_{i \sim j} \Pr\{A_i \land A_j\}
\]

The sum is over ordered pairs.

\[
\text{Cov}(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j] \leq E[X_iX_j] = \Pr\{A_i \land A_j\}
\]

\[\Rightarrow \text{Var}[X] \leq E[X] + \Delta\]
The Basic Theorem

Theorem 5

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

Proof:

\[
\Pr\{X = 0\} \leq \frac{Var[X]}{E^2[X]} \leq \frac{E[X] + \Delta}{E^2[X]} = \frac{1}{E[X]} + \frac{\Delta}{E^2[X]} \to 0
\]
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).
We define
\[ \Delta^* = \sum_{j \sim i} \Pr\{A_j|A_i\} \]

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

Proof:
\[
\Delta = \sum_{i \sim j} \Pr\{A_i \land A_j\} = \sum_{i \sim j} \Pr\{A_i\} \Pr\{A_j|A_i\} \\
= \sum_i \sum_{j \sim i} \Pr\{A_i\} \Pr\{A_j|A_i\} \\
= \sum_i \Pr\{A_i\} \sum_{j \sim i} \Pr\{A_j|A_i\} \\
= \Delta^* \cdot \sum_i \Pr\{A_i\} \\
\Rightarrow \Delta = \Delta^* \cdot E[X]
\]
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 7**

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

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

**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 8**

Let $A$ be the property of existence of $K_4$ cliques in $G_{n,p}$. The threshold function for $A$ is $p_0(n) = n^{-2/3}$.

**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 8

By Linearity of expectation

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

- \( E[X] = n^4 p^6 \ll 1 \iff p \ll n^{-2/3} \)
  - If \( p \ll n^{-2/3} \Rightarrow E[X] \to 0 \Rightarrow \) non-existence w.h.p.
  - Also, clearly \( p \gg n^{-2/3} \Rightarrow E[X] \to \infty. \)
Proof of theorem 8

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 8

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 \(|S_i \cap S_j| = 3\).

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

\[ = O(n^2)p^5 + O(n)p^3 = O(1/n) \]
When $p = n^{-2/3}$ we have:

$$\Delta^* \sim n^2p^5 + np^3 \sim n^{-4/3} + n^{-1}$$

and $E[X] \to 1$

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.