TTIC 31150/CMSC 31150
Mathematical Toolkit (Spring 2023)

Avrim Blum and Ali Vakilian

Lecture 11: Tail inequalities 1

Hwk3 deadline is now on **Friday April 28**

Hwk4 out today. **Due May 8**

Recap

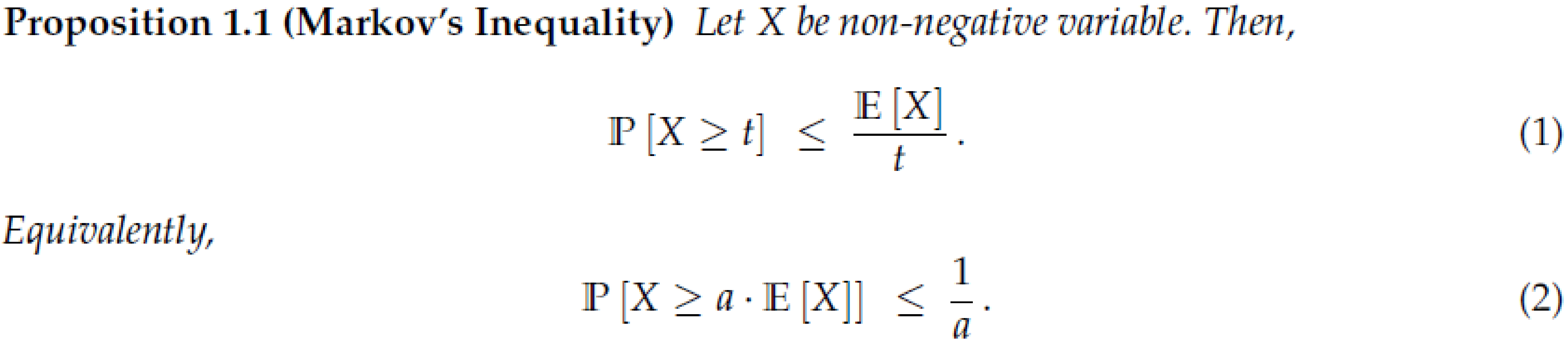
- Definitions of sample space Ω, events, random variables, expectation, conditional probability, conditional expectation, linearity of expectation, independence of events and R.Vs, mutual vs pairwise independence, properties of independence, Bernoulli, Binomial, and Geometric RVs.
- The Probabilistic Method. Examples.
- The Coupon Collector Problem.
- The DeMillo-Lipton-Schwartz-Zippel lemma. Polynomial identity testing.
- Application of DLSZ to finding perfect matchings in general graphs.

Tail inequalities

Bounds on the probability mass in the tail of a distribution. Use to show that it's unlikely a given R.V. X will take on a value too far from $\mathbb{E}[X]$.

Markov's inequality

The most basic. For non-negative R.V.s. Uses nothing about it except its expectation.

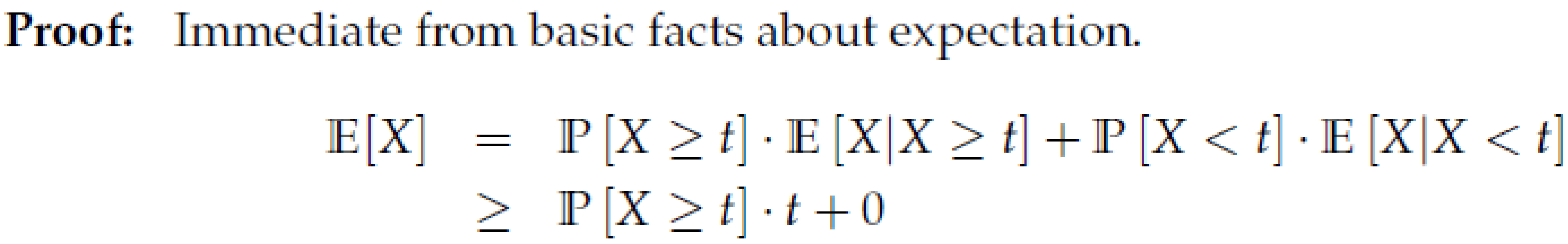


Proposition 1.1 (Markov's Inequality) *Let X be non-negative variable. Then,*

$$\mathbb{P}\left[X \geq t\right] \;\leq\; \frac{\mathbb{E}\left[X\right]}{t}\,. \tag{1}$$

Equivalently,

$$\mathbb{P}\left[X \geq a \cdot \mathbb{E}\left[X\right]\right] \;\leq\; \frac{1}{a}\,. \tag{2}$$

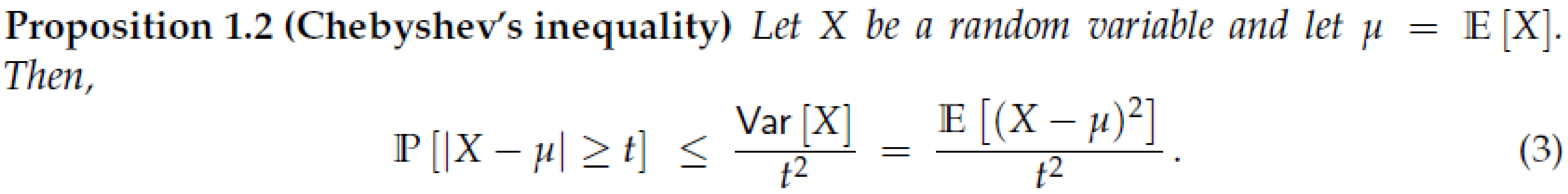


Proof: Immediate from basic facts about expectation.

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{P}\left[X \geq t\right] \cdot \mathbb{E}\left[X | X \geq t\right] + \mathbb{P}\left[X < t\right] \cdot \mathbb{E}\left[X | X < t\right] \\ &\geq \mathbb{P}\left[X \geq t\right] \cdot t + 0 \end{aligned}$$

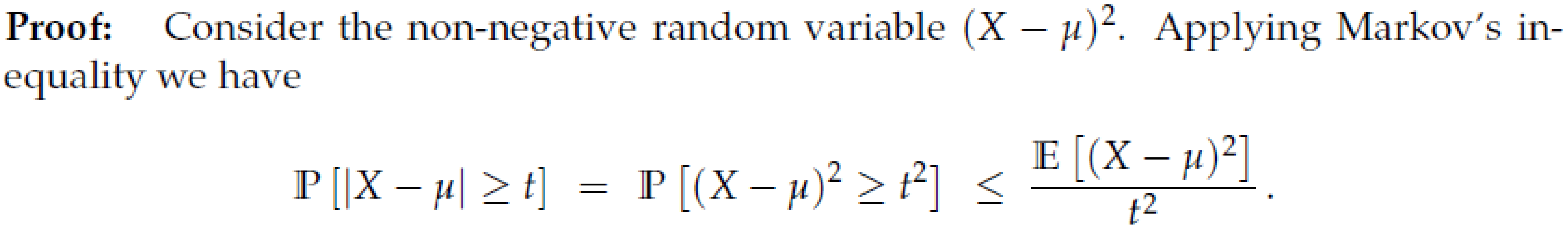
Chebyshev’s inequality

Stronger guarantee when we have a good bound on variance.



Proposition 1.2 (Chebyshev's inequality) *Let X be a random variable and let $\mu = \mathbb{E}[X]$. Then,*

$$\mathbb{P}[|X-\mu| \geq t] \;\leq\; \frac{\mathsf{Var}[X]}{t^2} \;=\; \frac{\mathbb{E}[(X-\mu)^2]}{t^2}. \tag{3}$$



Proof: Consider the non-negative random variable $(X-\mu)^2$. Applying Markov's inequality we have

$$\mathbb{P}[|X-\mu| \geq t] \;=\; \mathbb{P}[(X-\mu)^2 \geq t^2] \;\leq\; \frac{\mathbb{E}[(X-\mu)^2]}{t^2}.$$

Variance

- Definition: $Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Can simplify as: $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Example: Let X be an indicator R.V. for a coin of bias p.

- $\mathbb{E}[X] = p$.
- $Var[X] = p - p^2 = p(1 - p)$.

What if we flip n coins?

Variance

Proposition 1.3 *Let* $X = X_1 + \ldots + X_n$ *where the* X_i *are pairwise independent. Then* $\mathrm{Var}[X] = \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n]$.

Proof:

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[\sum_i \sum_j X_i X_j\right] - \left(\sum_i \mathbb{E}[X_i]\right)^2 \\
&= \sum_i \mathbb{E}[X_i^2] + \sum_i \sum_{j \neq i} \mathbb{E}[X_i X_j] - \sum_i \mathbb{E}[X_i]^2 - \sum_i \sum_{j \neq i} \mathbb{E}[X_i]\mathbb{E}[X_j] \\
&= \sum_i \mathrm{Var}[X_i] \quad \text{(using pairwise independence)}
\end{aligned}$$

So, if we flip n coins of bias p, we have $Var[X] = np(1-p)$. Standard deviation $\sigma = \sqrt{Var[X]} = \sqrt{np(1-p)}$.

Markov vs Chebyshev for coin flips

Flip n coins of bias $\frac{1}{2}$. Let X_i be indicator for ith toss, and let $X = X_1 + \cdots + X_n$.

- $\mathbb{E}[X_i] = \frac{1}{2}, Var[X_i] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}[X_i]^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$
- $\mathbb{E}[X] = \frac{n}{2}, Var[X] = \frac{n}{4}.$

Markov's inequality: $\mathbb{P}[X \geq 3n/4] \leq \frac{\mathbb{E}[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}.$

Chebyshev's inequality: $\mathbb{P}\left[\left|X - \frac{n}{2}\right| \geq t\right] \leq \frac{Var[X]}{t^2}$

- ➢ Using $t = \frac{n}{4}$, get $\mathbb{P}\left[X \geq \frac{3n}{4}\right] \leq \frac{n/4}{n^2/16} = \frac{4}{n}.$
- ➢ Using $t = \sqrt{n}$, get $\mathbb{P}\left[\left|X - \frac{n}{2}\right| \geq \sqrt{n}\right] \leq \frac{n/4}{n} = \frac{1}{4}.$

Markov vs Chebyshev for coin flips

So, by using pairwise independence, we can get much sharper concentration.

Later, we'll see even stronger concentration bounds we can get using mutual independence.

Markov's inequality: $\mathbb{P}[X \geq 3n/4] \leq \frac{\mathbb{E}[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}$.

Chebyshev's inequality: $\mathbb{P}\left[\left|X - \frac{n}{2}\right| \geq t\right] \leq \frac{Var[X]}{t^2}$

- Using $t = \frac{n}{4}$, get $\mathbb{P}\left[X \geq \frac{3n}{4}\right] \leq \frac{n/4}{n^2/16} = \frac{4}{n}$.
- Using $t = \sqrt{n}$, get $\mathbb{P}\left[\left|X - \frac{n}{2}\right| \geq \sqrt{n}\right] \leq \frac{n/4}{n} = \frac{1}{4}$.

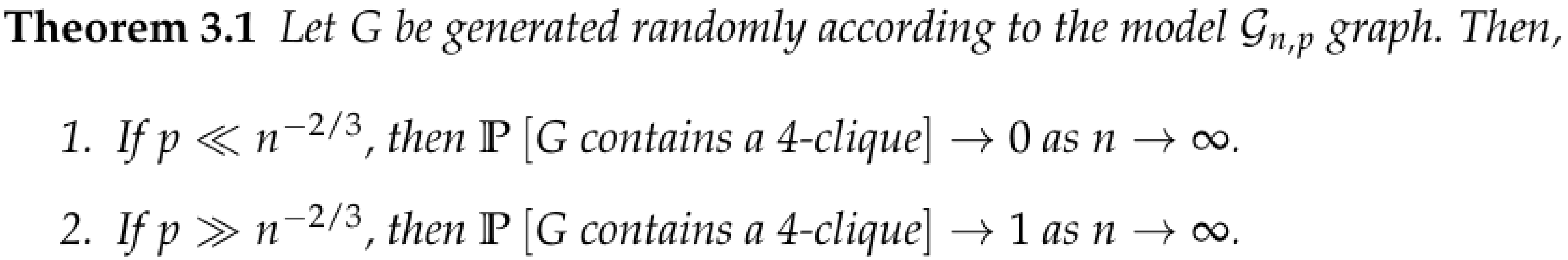
Threshold phenomena in Random Graphs

Consider a graph G on n vertices where each possible edge is placed into the graph independently with probability p. This is called the $G_{n,p}$ random graph model.

It turns out that many graph properties have “threshold phenomena”: for some function $f(n)$, for $p \ll f(n)$ the graph will almost surely not have the property and for $p \gg f(n)$ the graph almost surely will have the property (or vice-versa).

We will see one example here: the property of containing a 4-clique.

Threshold phenomena in Random Graphs



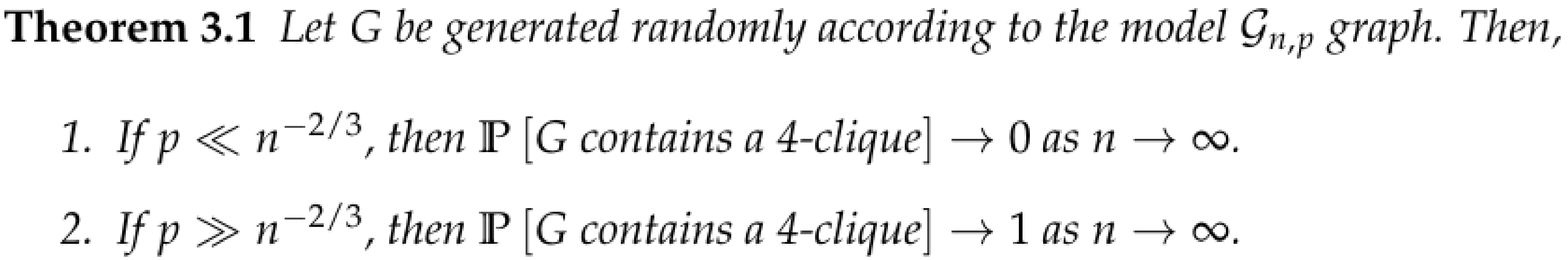
Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 1$ as $n \to \infty$.*

(1) Is the easier case, so let’s start with that:

- For each set S of 4 vertices, define indicator R.V. X_S for the event that S is a clique.
- Let $X = \sum_S X_s$ denote the number of 4-cliques in the graph.
- We have $\mathbb{E}[X] = \sum_S \mathbb{E}[X_S] = O\left(n^4 p^6\right) = o(1)$ for $p \ll n^{-2/3}$.
- So, by Markov’s inequality, $\mathbb{P}[X \geq 1] \leq \mathbb{E}[X]/1 = o(1)$.

Threshold phenomena in Random Graphs



Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 1$ as $n \to \infty$.*

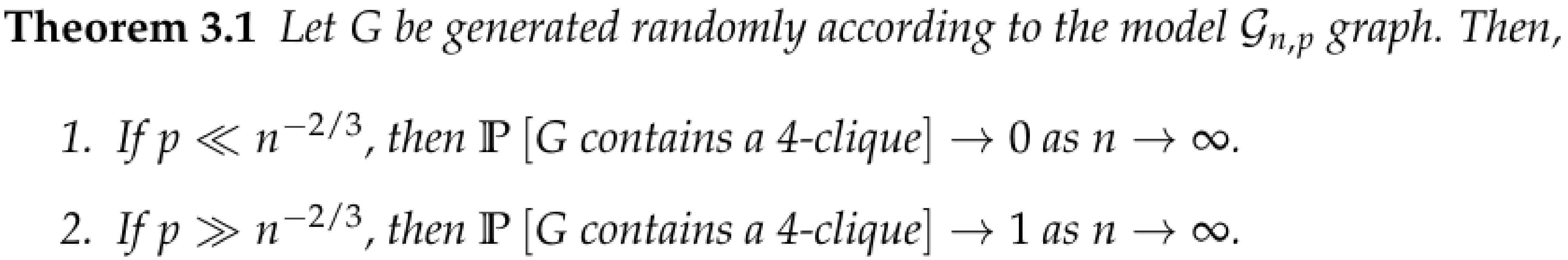
For (2), we have $\mathbb{E}[X] = \Theta\left(n^4 p^6\right) \to \infty$, but this is not sufficient to get $\mathbb{P}[X = 0] = o(1)$.

For this, we will use Chebyshev's inequality with $t = \mathbb{E}[X]$, giving:

$$\mathbb{P}[X = 0] \leq \frac{Var[X]}{\mathbb{E}[X]^2}$$

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs



Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 1$ as $n \to \infty$.*

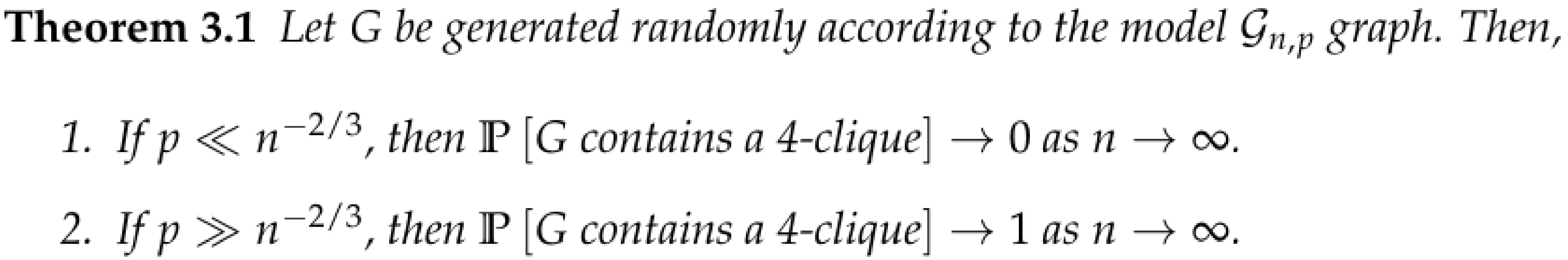
We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}\left[X_S X_{S'}\right] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- If S, S' share at most 1 vertex in common, then X_S and $X_{S'}$ are independent, so $\mathbb{E}\left[X_S X_{S'}\right] = \mathbb{E}[X_S]\mathbb{E}[X_{S'}]$ and the sum over all of these is at most $\mathbb{E}[X]^2$. We can therefore cover these using the $-\mathbb{E}[X]^2$ term.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs



Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 1$ as $n \to \infty$.*

We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}\left[X_S X_{S'}\right] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- If S, S' share 2 vertices in common, there are at most $O(n^6)$ such cases and each one has $\mathbb{E}\left[X_S X_{S'}\right] = p^{11}$. So, overall, we get $O\left(n^6 p^{11}\right) = o\left(n^8 p^{12}\right) = o\left(\mathbb{E}[X]^2\right)$.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 1$ as $n \to \infty$.*

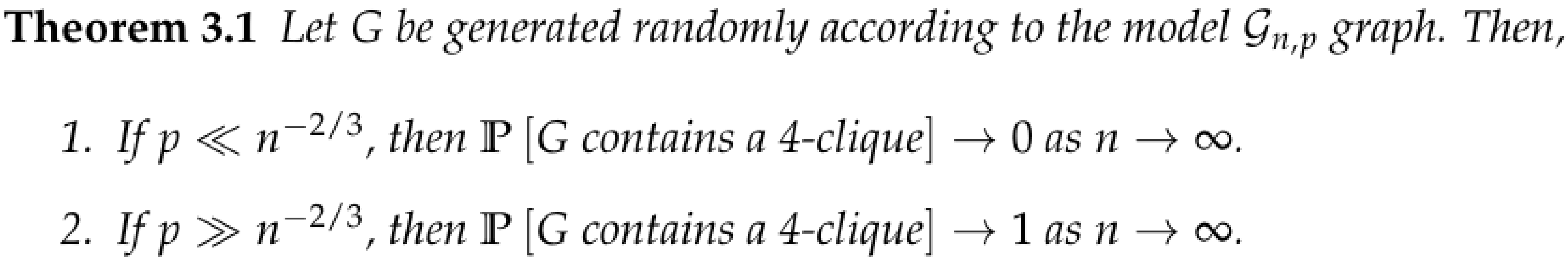
We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}\left[X_S X_{S'}\right] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- If S, S' share 3 vertices in common, there are at most $O(n^5)$ such cases and each one has $\mathbb{E}\left[X_s X_{S'}\right] = p^9$. So, overall, we get $O\left(n^5 p^9\right) = o\left(n^8 p^{12}\right) = o\left(\mathbb{E}[X]^2\right)$.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs



Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 1$ as $n \to \infty$.*

We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}\left[X_S X_{S'}\right] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- And finally, if S, S' share all 4 vertices in common, then the total is just $\mathbb{E}[X] \; = o\left(\mathbb{E}[X]^2\right)$.
- So, overall we have $Var[X] = o\left(\mathbb{E}[X]^2\right)$ as desired.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.