Course Notes

Part V

Probabilistic Combinatorics

and

Algorithms

J. A. Verstraete
Department of Mathematics
University of California San Diego
9500 Gilman Drive
La Jolla California 92037-0112

jacques@ucsd.edu

1 Martingale Inequalities

A filter (or filtration) in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family of increasing σ-algebras in $\mathcal{F}$, namely, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$. An example of a filter is the natural filter given by a random process $(X_i)_{i\geq 1}$, where $\mathcal{F}_i = \sigma(X_1, X_2, \ldots, X_i)$ for all i. Relative to a filter $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a process $(X_i)_{i\geq 1}$ is called a martingale if almost surely

$$\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) = X_{i-1}$$

for all $i \geq 1$. Here almost surely has the same meaning as with probability one. The classical example of a martingale is the fair game: the σ-algebra $\mathcal{F}_i$ is the game up to time i, and X_i is the amount won so far. Then almost surely

$$\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) = X_{i-1}$$

so we have a martingale. More generally, if $X_1, X_2, \ldots, X_n$ are independent random variables with zero mean and S_i is the sum of the first i random variables. Then

$$\begin{aligned}
\mathbb{E}(S_i|\mathcal{F}_{i-1}) &= \mathbb{E}(S_{i-1}|\mathcal{F}_{i-1}) + \mathbb{E}(X_i|\mathcal{F}_{i-1}) \\
&= S_{i-1} + \mathbb{E}(X_i) \\
&= S_{i-1},
\end{aligned}$$

as required. As an exercise, one can verify that $P_n = X_1 X_2 \ldots X_n$ is a martingale, where $\mathbb{E}(X_n) = 1$ for all n and the random variables $X_1, X_2, \ldots$ are independent, and $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$.

Relative to a filter $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a process $(X_i)_{i\geq 1}$ is called a supermartingale if almost surely

$$\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) \leq X_{i-1}$$

and a submartingale if almost surely

$$\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) \geq X_{i-1}.$$

1.1 Combinatorial Martingales

A useful martingale in graph theory is the edge-exposure martingale. For a function f on the set of graphs on n vertices and an ordering of the $N = \binom{n}{2}$ edges in K_n, we may expose the edges of a random graph one by one in order and let X_i be the expected value of f when the first i edges have been exposed. Then $(X_i)_{i\geq 0}$ is a martingale with $X_N = f(G)$ and $X_0 = \mathbb{E}(f(G))$. A subsequence of the edge-exposure martingale is the martingale $(X'_i)_{i\geq 1}$ where X'_i is the expectation of f when all the edges at the first i vertices are exposed. This is called the vertex-exposure martingale. Both of these martingales are particular finite cases of Doob martingales. In general, we may define a martingale by considering the expected value of an $\mathcal{F}_n$-measurable function conditional on $X_1, X_2, \ldots, X_i$ for $i \in [n]$.

1.2 Hoeffding-Azuma Inequality

Based on the difference sequence $Y_i = X_i - X_{i-1}$ of a martingale $(X_i)_{i=1}^n$, one can gather information about the concentration of X_n about its mean. For this reason, the use of the inequalities in this section is often referred to as the method of bounded differences. First we need a technical lemma concerning bounded real valued random variables with zero mean.

Lemma 1 *Let $a \in \mathbb{R}$, and let Y be any random variable with zero mean such that $-a \leq Y \leq 1-a$. Then for any $t \geq 0$,*

$$\mathbb{E}(e^{tY}) \leq e^{\frac{1}{8}t^2}.$$

Proof ▷ By convexity of the exponential function, for any $y \in \mathbb{R}$ such that $0 \leq a+y \leq 1$, we have

$$e^{ty} \leq (1-a-y)e^{-ta} + (a+y)e^{t(1-a)}.$$

Using that Y has zero mean, we obtain

$$\mathbb{E}(e^{tY}) \leq (1-a)e^{-ta} + ae^{t(1-a)} \leq \cosh(\tfrac{t}{2}).$$

Using Taylor's Theorem, we find $\cosh(t) \leq e^{\frac{1}{2}t^2}$. The lemma follows. ▮

One of the first fundamental inequalities with regard to bounded differences was proved by Hoeffding, generalizing an inequality of Azuma.

Theorem 2 *Let $(X_i)_{i=1}^n$ be a martingale with difference sequence $(Y_i)_{i=1}^n$, where $-a_i \leq Y_i \leq -a_i + c_i$, where a_i is a function on $(\Omega, \mathcal{F}_{i-1})$ and $c_i \in \mathbb{R}$. Then for $t \geq 0$ and $c := \sum c_i^2$,*

$$\mathbb{P}(X_n > \mathbb{E}(X_n) + \lambda) \leq e^{-\frac{2\lambda^2}{c}} \quad \text{and} \quad \mathbb{P}(X_n < \mathbb{E}(X_n) - \lambda) \leq e^{-\frac{2\lambda^2}{c}}.$$

Proof ▷ First apply Lemma 1 to $Y = \frac{Y_i}{c_i} | \mathcal{F}_{i-1}$ to obtain, for any non-negative real number t,

$$\mathbb{E}(e^{tY_i} | \mathcal{F}_{i-1}) \leq e^{\frac{1}{8}c_i^2 t^2}.$$

We now prove the theorem. Let $Z_n = Y_1 + Y_2 + \cdots + Y_n$. For any non-negative real number t, we have

$$\begin{aligned}
\mathbb{E}(\mathrm{e}^{tZ_n}) &= \mathbb{E}(\mathbb{E}(\mathrm{e}^{tZ_n} \mid \mathcal{F}_{n-1}]) && \text{by the tower property} \\
&= \mathbb{E}(\mathrm{e}^{tZ_{n-1}}) \cdot \mathbb{E}(\mathrm{e}^{tY_n} \mid \mathcal{F}_{n-1}) \\
&= \mathbb{E}(\mathrm{e}^{tZ_{n-1}}) \cdot \mathbb{E}(\mathrm{e}^{c_n t(\frac{Y_n}{c_n})} \mid \mathcal{F}_{n-1}) \\
&\leq \mathrm{e}^{\frac{1}{8}c_n^2 t^2} \cdot \mathbb{E}(\mathrm{e}^{tZ_{n-1}}) && \text{by the first inequality.}
\end{aligned}$$

Repeating the above calculations, we find $\mathbb{E}(\mathrm{e}^{tZ_n}) \leq \mathrm{e}^{ct^2/8}$. Finally, by Markov's inequality,

$$\begin{aligned}
\mathbb{P}(X_n > \mathbb{E}(X_n) + \lambda) &= \mathbb{P}(Z_n > \lambda) \\
&= \mathbb{P}(\mathrm{e}^{tZ_n} > \mathrm{e}^{\lambda t}) \\
&\leq \mathbb{E}(\mathrm{e}^{tZ_n}) \cdot \mathrm{e}^{-\lambda t} \\
&\leq \mathrm{e}^{-\lambda t}\mathrm{e}^{\frac{1}{8}ct^2} \\
&\leq \mathrm{e}^{-\frac{2\lambda^2}{c}}.
\end{aligned}$$

In the last line, we chose $t = \frac{4\lambda}{c}$. ∎

A special case of Hoeffding's inequality is Azuma's inequality: if $(X_i)_{i\geq 1}$ is a martingale and $|X_{i+1} - X_i| \leq 1$ for all i and $X_0 = 0$, then

$$\mathbb{P}(X_n > \lambda\sqrt{n}) \ < \ \mathrm{e}^{-\lambda^2/2}.$$

1.3 Concentration of Lipschitz Functions

There is a special name for real-valued functions on a Cartesian product $\prod_{i=1}^n A_i$ which satisfy $|f(x) - f(y)| \leq c$ whenever x and y differ in one co-ordinate: we say that f is c-Lipschitz or f has Lipschitz coefficient c. The next theorem says that as we expose the random variables of a c-Lipschitz function, the function is highly concentrated around its expected value. More precisely:

Theorem 3 *Let $X_1, X_2, \dots, X_n$ be independent random variables with values in $A_1, A_2, \dots, A_n \subset \mathbb{R}$, respectively. Suppose $f : \prod_{i=1}^n A_i \to \mathbb{R}$ is Lipschitz in the sense that $|f(x)-f(y)| \leq c_i$ whenever x and y differ only on the ith co-ordinate. Then with $Z = f(X_1, X_2, \dots, X_n)$,*

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > \lambda) \ \leq \ 2e^{-2\lambda^2/c}.$$

Proof ▷ Let $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ be the following filter on $\Omega = \prod_{i=1}^n A_i$: the σ-field $\mathcal{F}_k$ is a partition of Ω into $\left|\prod_{i=1}^k A_i\right|$ parts according to the first i co-ordinates of the vectors in Ω. Then $Z_i = \mathbb{E}(Z \mid \mathcal{F}_i)$ is a martingale. The Lipschitz condition on f implies that

$$\min Z_{i+1} \ \leq \ \mathbb{E}(Z \mid \mathcal{F}_i) \ \leq \ \max Z_{i+1} \leq \min Z_{i+1} + c_{i+1}.$$

where the maxima and minima are taken over the $(i+1)$th co-ordinates in Ω. It follows that the difference sequence $Y_i = Z_i - Z_{i-1}$ is bounded, namely $-a_i \leq Y_i \leq c_i - a_i$ for some measurable function a_i on $(\Omega, \mathcal{F}_i)$. An application of Hoeffding's inequality completes the proof. ∎

The above inequality is one of the simplest concentration inequalities for Lipschitz functions, and there are many others in the literature. Later on we will consider the case of polynomial concentration: clearly an arbitrary multivariate polynomial $f(X_1, X_2, \dots, X_n)$ need not be Lipschitz, while we shall see that concentration follows from certain conditions on partial derivatives of f.

1.4 A simple example

Let $v_1, v_2, \dots, v_n \in \mathbb{R}^n$ be unit vectors and let $s_1, s_2, \dots, s_n \in \{-1, 1\}$ be uniformly and independently chosen. We wish to prove that the function

$$f(s_1, s_2, \dots, s_n) = \|s_1 v_1 + s_2 v_2 + \cdots + s_n v_n\|$$

is highly concentrated at $\sqrt{n}$. We know that $\mathbb{E}(f^2) = n$ using linearity of expectation. Unfortunately f is not a sum of independent random variables, so concentration does not follow from Chernoff-type bounds. However we can show that f is 2-Lipschitz as follows: let $\vec{s}$ and $\vec{t}$ differ in one co-ordinate, which we may assume is the nth co-ordinate. Then

$$|f(\vec{s}) - f(\vec{t})| \leq \|\textstyle\sum_{i<n} s_i v_i\| + \|s_n v_n\| - |\|\sum_{i<n} s_i v_i\| - \|t_n v_n\|| \leq 2.$$

So f is 2-Lipschitz, and by the Lipschitz inequality,

$$\mathbb{P}(|f - \mathbb{E}(f)| > \lambda) \leq 2e^{-\lambda^2/2n}.$$

In particular, using the Cauchy-Schwartz inequality, we see that with high probability, a random signing of unit vectors gives $f \ll \sqrt{n}$.

1.5 Concentration of chromatic number

We can now prove that the chromatic number of a random graph is highly concentrated at its expected value, whilst computing the expected value is actually a separate and non-trivial issue. We denote buy $\omega(n)$ an arbitrary function of n tending to infinity as n tends to infinity.

Theorem 4 *Let $G_{n,p}$ denote the random graph with uniform edge-probability p, and let $\omega(n)$ be a function tending arbitrarily slowly to infinity. Then asymptotically almost surely*

$$|\chi(G_{n,p}) - \mathbb{E}(\chi(G_{n,p}))| \leq \sqrt{n/2} + \omega(n).$$

Proof $\triangleright$ We apply the Lipschitz inequality to the vertex exposure martingale. The chromatic number of a graph changes by at most one on adding a vertex, so the conditions of the last lemma are satisfied with $c_i = 1$ for all i. In other words, the function $f(G) = \chi(G)$ is 1-Lipschitz. Therefore

$$\mathbb{P}[|\chi(G_{n,p}) - \mathbb{E}(\chi(G_{n,p}))| > \lambda) \;\; \leq \;\; \mathrm{e}^{-2\lambda^2/n}.$$

The result follows on choosing $\lambda = \sqrt{n/2} + \omega(n)$. ∎

In fact, one can do a lot more using martingales: Bollobás proved that if $G = G(n,p)$ is the random graph with edge-probability $p > n^{-5/6}$, then $\chi(G)$ is one of four consecutive integers asymptotically almost surely. In other words, there exists a function $u = u(n,p)$ such that $u \leq \chi(G) \leq u + 3$ with probability $1 - o(1)$ as n tends to infinity. We will give the proof in the section on random graphs. A recent breakthrough was achieved by Achlioptas and Naor, who showed that the chromatic number of a random graph is concentrated on two values.

1.6 Counting Sums of Sets

The point of this section is that Fourier transforms of characteristic functions of random sets lend themselves very readily to martingale inequalities. We'll use this to show that the number of subsets of $\mathbb{Z}_n$ of the form $A - A = \{a - b : a, b \in A\}$ for some $A \subset \mathbb{Z}_n$ is small relative to the total 2^n subsets of $\mathbb{Z}_n$. We call such sets $A - A$ difference sets. For a set A, we denote by $A(r)$ its characteristic function, that is, $A(r) = 1$ if $r \in A$ and $A(r) = 0$ otherwise. The Fourier transform of a function $f : \mathbb{Z}_n \to \mathbb{C}$ is defined by

$$\hat{f}(x) \;=\; \sum_{r=0}^{n-1} f(r)\omega^{rx}.$$

Here and throughout what follows, $\omega = \mathrm{e}^{2\pi i/n}$. Thus f is a function from $\mathbb{Z}_n$ to $\mathbb{C}$, the set of complex numbers. In particular, if $f(r) = A(r)$, then we note that $\hat{A}(0) = |A|$. The following is one of the most important identities in Fourier analysis:

Lemma 5 *Let $f : \mathbb{Z}_n \to \mathbb{C}$ be a function. Then*

$$\sum_{r=0}^{n-1} |\hat{f}(r)|^2 = n \sum_{r=0}^{n-1} |f(r)|^2.$$

The identity is known as Parseval's identity.

Lemma 6 *Suppose $A, S \subset \mathbb{Z}_n$, and $S \supset A - A$, and let M be the maximum value of $|\hat{T}(r)|$ for $r \in \mathbb{Z}_n$, where $T = S^c \cup (-S^c)$. Then $M \geq |T||A|/(n - |A|)$.*

Proof ▷ Since $S \supset A - A$, $S^c \cap (A - A) = \emptyset$ and $-S^c \cap (A - A) = \emptyset$. Therefore $T \cap (A - A) = \emptyset$. Therefore

$$\sum_{r=0}^{n-1}\sum_{t=0}^{n-1}\sum_{a=0}^{n-1}\sum_{b=0}^{n-1} A(a)A(b)T(t)\omega^{r(t-a+b)} = 0.$$

It follows by definition of $\hat{A}(r)$ and $\hat{T}(r)$ that

$$\sum_{r=0}^{n-1} \hat{T}(r)|\hat{A}(r)|^2 = 0.$$

Now $|\hat{T}(0)||\hat{A}(0)| = |T||A|^2$. Therefore, by Parseval's identity,

$$|T||A|^2 = \Big|\sum_{r=1}^{n-1} \hat{T}(r)|\hat{A}(r)|^2\Big| \leq M \sum_{r=1}^{n-1} |\hat{A}(r)|^2 = Mn|A| - M|A|^2.$$

Since $|A| \geq 1$, we find $M \geq |T||A|/(n - |A|)$ as required. ∎

Using Theorem 2 and Lemma 6, it is relatively straightforward to give an upper bound on the number of subsets of $\mathbb{Z}_n$ containing large difference sets:

Theorem 7 *Let n be a positive odd integer. The number of subsets of $\mathbb{Z}_n$ containing a difference set $A - A$ with $|A| \geq m$ is at most*

$$n2^{n+1} \cdot \exp\Big(-\frac{9m^2 n}{8[n + (\sqrt{3} - 1)m]^2}\Big).$$

Proof ▷ Note that may assume $m \leq \frac{1}{2}n$, otherwise $A - A = \mathbb{Z}_n$ since $a - A$ intersects $b - A$ for all $a, b \in \mathbb{Z}_n$. Let S be a random set whose elements are chosen uniformly with probability $\frac{1}{2}$ from $\mathbb{Z}_n$. Let $X_r = S(r)$, so that $\sum_{i=0}^{n-1} X_i = |S|$. Noting that $S^c(i) = 1 - X_i$, we have

$$f(X_0, X_1, \ldots, X_{n-1}) = \hat{T}(r) = \sum_{s=0}^{n-1} (1 - X_s)(\omega^{rs} + \omega^{-rs}).$$

It follows that $\hat{T}(r)$ is real. We now apply Theorem 3 to $\hat{T}(r)$. Note that we have a martingale, since $\mathbb{E}(\hat{T}(r)) = 0$, and the Lipschitz coefficient of f is one (that is, every c_i is equal to one in Theorem 3). Therefore

$$\mathbb{P}(|\hat{T}(r)| \geq \lambda) \ \leq \ \mathbb{P}(\hat{T}(r) \geq \lambda) \ \leq \ \mathrm{e}^{-2\lambda^2/n}.$$

Finally, we apply Lemma 6. Suppose $S \supset A - A$ for some A with $|A| \geq m$. Now there exists $r \in \mathbb{Z}_n \backslash \{0\}$ such that $|\hat{T}(r)| \geq m|T|/(n-m)$, by Lemma 6. Let us write this event as $|\hat{T}| \geq m|T|/(n-m)$. It follows that for any $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(S \supset A - A) &\leq \mathbb{P}(|\hat{T}| \geq m|T|/(n-m)) \\ &\leq \mathbb{P}(|\hat{T}| \geq m\lambda/(n-m)) + \mathbb{P}(|T| \leq \lambda) \\ &\leq n\mathrm{e}^{-2m^2\lambda^2/n(n-m)^2} + \mathrm{e}^{-(1-4\lambda/3n)^2 3n/8}. \end{aligned}$$

We used the Chernoff Bound to bound $\mathbb{P}(|T| \leq \lambda]$. Since n is odd, $|T|$ is a sum of $(n-1)/2$ independent random variables, namely $X(i) + Y(i) - X(i)Y(i)$ where $X = S^c$ and $Y = -S^c$. Note that $\mathbb{E}(|T|) = 3n/4$. Choosing $\lambda = 3n(n-m)/[(\sqrt{3}-1)m + n]$ completes the proof. ▮

Corollary 8 *Let n be a positive odd integer. Then the number of subsets of $\mathbb{Z}_n$ which are difference sets is at most 2^{cn} for some constant $c < 1$.*

Proof ▷ Let α be a positive real number less than $\frac{1}{2}$. Then the number of sets of size at most αn is

$$\sum_{k \leq \alpha n} \binom{n}{\alpha n} \leq \alpha n \binom{n}{\alpha n} \leq n\exp[H(\alpha)],$$

where $H(\alpha)$ is the entropy of α. In particular, the number of sets of the form $A - A$ with $|A| \leq \alpha n$ is certainly at most $2n\exp[H(\alpha)]$. By Theorem 7, the number of $A - A$ with $|A| > \alpha n$ is at most

$$2n \cdot \exp\ n\Big[\log 2 - \frac{9\alpha^2}{8[1 + (\sqrt{3}-1)\alpha]^2}\Big].$$

Now the equation

$$H(\alpha) = \log 2 - \frac{9\alpha^2}{8[1 + (\sqrt{3}-1)\alpha]^2}$$

has exactly two roots $\alpha \in (0,1)$, but only one of these is less than $\frac{1}{2}$. Therefore we take α to be this solution. Numerical analysis shows $\alpha \approx 0.312$, in which case the bound is at most of order $\mathrm{e}^{0.617n} \ll 2^{0.892n}$. ▮

Similar arguments can be carried out if we wish to count the number of sets of the form $A + A$ in $\mathbb{Z}_n$. We also remark that a simple modification of these arguments can be used to count such sets in $\{1, 2, \ldots, n\}$. The question of the number of sets $A + A$ or $A - A$ in $\mathbb{Z}_p$ was completely answered by Green and Ruzsa (2001) if p is prime. They proved that the number of $A + A$ or $A - A$ is of order $f(p)2^{p/3}$ for some polynomial $f(p)$ (the asymptotic behaviour, in terms of the degree of $f(p)$, is not known). Green (2002) further solved a major conjecture of Erdős and Cameron, that the number of sum-free subsets of $\{1, 2, \ldots, n\}$ is asymptotic to $c2^{n/2}$ where c is one of two values, dependening on whether n is even or odd.

1.7 Polynomial Concentration

This recent topic was developed by Kim and Vu. The idea is that if a function $f(X_1, X_2, \ldots, X_n)$ of independent random variables is a smooth enough polynomial, then f enjoys concentration similar to Lipschitz functions. A test case is the number of triangles in a random graph $G_{n,p}$: each X_i is then a Bernouilli random variable attached to the ith edge, and the number $\triangle$ of triangles is a multivariable polynomial. Note that $\triangle$ is not a Lipschitz function under vertex or edge-exposure martingales. The setup is as follows: let $\mathcal{A}$ be a family of finite subsets of a given finite set V, where each set $A \in \mathcal{A}$ has a weight $\omega(A) \geq 0$. For each $i \in V$, let X_i be a Bernouilli random variable with mean p_i, such that $X_i : i \in V$ are mutually independent. Consider the random variable

$$X = X(\mathcal{A}) = \sum_{A \in \mathcal{A}} \omega(A) \prod_{i \in A} X_i$$

where $\prod_{i \in \emptyset} X_i = 1$ by convention. So X is a multivariate polynomial in $X_1, X_2, \ldots, X_n$. Now let $\partial_Y X$ denote the partial derivative of X with respect to each $X_i : i \in Y$ when $Y \subseteq X$. This is equivalent to considering $X(\mathcal{B})$ where

$$\mathcal{B} = \{A \backslash Y : A \in \mathcal{A} \wedge Y \subseteq A\}.$$

Therefore

$$\partial_Y X = \sum_{\substack{A \in \mathcal{A} \\ Y \subseteq A}} \omega(A) \prod_{i \in A \backslash Y} X_i.$$

Define the influence by

$$\inf(X) = \max_{|A| \geq 0} \mathbb{E}(\partial_A X)$$

and the non-empty influence by

$$\inf^*(X) = \max_{|A| \geq 1} \mathbb{E}(\partial_A X).$$

The main result proved by Kim and Vu is the following:

Theorem 9 *Let $\lambda > 1$ be a real number, and let $k = \max\{|A| : A \in \mathcal{A}\}$. Then*

$$\mathbb{P}(|X - \mathbb{E}(X)| > (2k)!\lambda^k \sqrt{\inf(X)\inf^*(X)}) \leq 16e^{-\lambda + (k-1)\log n}$$

The proof of Theorem 9 uses a clever induction, which we do not present here.

The most basic application is to triangles in a random graph $G_{n,p}$. We clearly expect $p^3\binom{n}{3}$ triangles, but what is the probability that we deviate from this, in particular, roughly what is the chance that there are no triangles? We already addressed this via Janson's Inequality when $pn^{1/2} \to 0$. Here we will consider small p, namely $p = n^{\theta - 1}$ where $0 < \theta < \frac{1}{3}$. The polynomial concentration approach applies, by defining $\triangle$ to be the number of triangles in $G_{n,p}$. Let $\mathcal{A}$ denote the family of all three-element subsets of $V = \binom{[n]}{2}$, the set of pairs of elements of $[n]$. For $\{i, j\} \in V$, let X_{ij} be a Bernouilli random variable with mean p, corresponding to the presence of the ith edge. Then clearly

$$\triangle = \sum_{i,j,k} X_{ij} X_{jk} X_{ki}$$

where the sum is over distinct $i, j, k \in [n]$. Now we compute the influences. First note

$$\mathrm{i}(\triangle) = \max\{p^3\binom{n}{3}, (n-2)p^2, p, 1\} \sim \frac{1}{6}n^{3\theta}$$

since $p^2(n-2) \leq 1$ triangles on a given edge is very small relative to the expected number of triangles $p^3\binom{n}{3}$. Therefore $\mathrm{i}^*(\triangle) \leq 1$, so we obtain from Theorem 9:

Theorem 10 *Let $\triangle$ denote the number of triangles in $G_{n,p}$ and μ the expected number of triangles, where $p = n^{\theta-1}$ and $0 < \theta < \frac{1}{3}$. Then*

$$\mathbb{P}(|\triangle - \mu| > \lambda(\log n)^3\sqrt{\mu}) < n^{-\lambda^{1/3}}.$$

In particular,

$$\mathbb{P}(\triangle = 0) \ll e^{-\frac{\mu^{1/6}}{(\log n)^{1/2}}}.$$

This is far better than what we would obtain from Chebyshev's Inequality, and is more in line with our concentration for martingales. A concentration for $\triangle = 0$ for all values of p is missing. We will fill in some more values when we get to Janson's Inequality. For now we make the following observation: assume there is a Steiner system $\mathcal{S}$ of triples on $[n]$ – a family of triples for which every pair of elements of $[n]$ is contained in exactly one triple in the system. A simple counting argument shows that a Steiner system has exactly $\frac{1}{3}\binom{n}{2}$ triples. Now for each $S \in \mathcal{S}$, let $\triangle(S)$ denote the Bernouilli variable corresponding to a triangle in $G_{n,p}$ with vertex set S. Then the variables $\triangle(S)$ are independent, so

$$\mathbb{P}(\triangle = 0) \leq \prod_{S\in\mathcal{S}}(1-p^3) \leq e^{-\frac{1}{3}p^3\binom{n}{2}}.$$

For $p > n^{-2/3}$, this gives a concentration result, complimenting our analysis above.