most $l$.
n $l$ and

$\varepsilon$. ■

# 4

# The Second Moment

You don't have to believe in God but you should believe in The Book.
– Paul Erdős

## 4.1 BASICS

After the expectation the most vital statistic for a random variable $X$ is the *variance*. We denote it $\mathrm{Var}[X]$. It is defined by

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

and measures how spread out $X$ is from its expectation. We shall generally, following standard practice, let $\mu$ denote expectation and $\sigma^2$ denote variance. The positive square root $\sigma$ of the variance is called the *standard deviation*. With this notation, here is our basic tool.

**Theorem 4.1.1 [Chebyshev's Inequality]** *For any positive* $\lambda$,

$$\Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

**Proof.** $\sigma^2 = \mathrm{Var}[X] = \mathrm{E}\left[(X-\mu)^2\right] \geq \lambda^2\sigma^2 \Pr[|X - \mu| \geq \lambda\sigma]$. ■

The Probabilistic Method, Third Edition
By Noga Alon and Joel Spencer

The use of Chebyshev's Inequality is called the *second moment method.*

Chebyshev's Inequality is most possible when no additional restrictions are placed on $X$ as $X$ may be $\mu+\lambda\sigma$ and $\mu-\lambda\sigma$ with probability $1/2\lambda^2$ and otherwise $\mu$. Note, however, that when $X$ is a normal distribution with mean $\mu$ and standard deviation $\sigma$ then

$$\Pr[|X-\mu| \geq \lambda\sigma] = 2\int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

and for $\lambda$ large this quantity is asymptotically $\sqrt{2/\pi}e^{-\lambda^2/2}/\lambda$, which is significantly smaller than $1/\lambda^2$. In Chapters 7 and 8 we shall see examples where $X$ is the sum of "nearly independent" random variables and these better bounds can apply.

Suppose we have a decomposition

$$X = X_1 + \cdots + X_m .$$

Then $\mathrm{Var}[X]$ may be computed by the formula

$$\mathrm{Var}[X] = \sum_{i=1}^{m} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j] .$$

Here the second sum is over ordered pairs and the *covariance* $\mathrm{Cov}[Y,Z]$ is defined by

$$\mathrm{Cov}[Y,Z] = \mathrm{E}[YZ] - \mathrm{E}[Y]\,\mathrm{E}[Z] .$$

In general, if $Y, Z$ are independent then $\mathrm{Cov}[Y,Z] = 0$. This often simplifies variance calculations considerably. Now suppose further, as will generally be the case in our applications, that the $X_i$ are indicator random variables; that is, $X_i = 1$ if a certain event $A_i$ holds and otherwise $X_i = 0$. If $X_i$ is one with probability $p_i = \Pr[A_i]$ then

$$\mathrm{Var}[X_i] = p_i(1-p_i) \leq p_i = \mathrm{E}[X_i] ,$$

and so

$$\mathrm{Var}[X] \leq \mathrm{E}[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j] .$$

## 4.2 NUMBER THEORY

The second moment method is an effective tool in number theory. Let $\nu(n)$ denote the number of primes $p$ dividing $n$. (We do not count multiplicity though it would make little difference.) The following result says, roughly, that "almost all" $n$ have "very close to" $\ln\ln n$ prime factors. This was first shown by Hardy and Ramanujan in 1920 by a quite complicated argument. We give a remarkably simple proof of Turán (1934), a proof that played a key role in the development of probabilistic methods in number theory.



**Theorem 4.2.1** *Let $\omega(n) \to \infty$ arbitrarily slowly. Then the number of $x$ in $\{1,\ldots,n\}$ such that*

$$|\nu(x) - \ln\ln n| > \omega(n)\sqrt{\ln\ln n}$$

*is $o(n)$.*

**Proof.** Let $x$ be randomly chosen from $\{1,\ldots,n\}$. For $p$ prime set

$$X_p = \begin{cases} 1 & \text{if } p|x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$ and set $X = \sum X_p$, the summation over all primes $p \leq M$. As no $x \leq n$ can have more than ten prime factors larger than $M$ we have $\nu(x) - 10 \leq X(x) \leq \nu(x)$ so that large deviation bounds on $X$ will translate into asymptotically similar bounds for $\nu$. [Here 10 could be any (large) constant.] Now

$$\mathrm{E}[X_p] = \frac{\lfloor n/p \rfloor}{n} .$$

As $y - 1 < \lfloor y \rfloor \leq y$,

$$\mathrm{E}[X_p] = 1/p + O(1/n) .$$

By linearity of expectation,

$$\mathrm{E}[X] = \sum_{p \leq M} \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = \ln\ln n + O(1) ,$$

where here we used the well-known fact that $\sum_{p \leq x}(1/p) = \ln\ln x + O(1)$, which can be proved by combining Stirling's formula with Abel summation.

Now we find an asymptotic expression for

$$\mathrm{Var}[X] = \sum_{p \leq M} \mathrm{Var}[X_p] + \sum_{p \neq q} \mathrm{Cov}[X_p, X_q] .$$

As $\mathrm{Var}[X_p] = (1/p)(1-1/p) + O(1/n)$,

$$\sum_{p \leq M} \mathrm{Var}[X_p] = \left(\sum_{p \leq M} \frac{1}{p}\right) + O(1) = \ln\ln n + O(1) .$$

With $p, q$ distinct primes, $X_p X_q = 1$ if and only if $p|x$ and $q|x$, which occurs if and only if $pq|x$. Hence

$$\begin{aligned} \mathrm{Cov}[X_p, X_q] &= \mathrm{E}[X_p X_q] - \mathrm{E}[X_p]\,\mathrm{E}[X_q] \\ &= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\ &\leq \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right) \\ &\leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right) . \end{aligned}$$

Thus

$$\sum_{p\neq q}\mathrm{Cov}\left[X_p,X_q\right]\leq\frac{1}{n}\sum_{p\neq q}\left(\frac{1}{p}+\frac{1}{q}\right)\leq\frac{2M}{n}\sum\frac{1}{p}.$$

Thus

$$\sum_{p\neq q}\mathrm{Cov}\left[X_p,X_q\right]\leq O(n^{-9/10}\ln\ln n)=o(1),$$

and similarly

$$\sum_{p\neq q}\mathrm{Cov}\left[X_p,X_q\right]\geq -o(1).$$

That is, the covariances do not affect the variance, $\mathrm{Var}\left[X\right]=\ln\ln n+O(1)$ and Chebyshev's Inequality actually gives

$$\Pr\left[|X-\ln\ln n|>\lambda\sqrt{\ln\ln n}\right]<\lambda^{-2}+o(1)$$

for any constant $\lambda>0$. As $|X-\nu|\leq 10$ the same holds for $\nu$. ■

In a classic paper Erdős and Kac (1940) showed, essentially, that $\nu$ does behave like a normal distribution with mean and variance $\ln\ln n$. Here is their precise result.

**Theorem 4.2.2** *Let $\lambda$ be fixed, positive, negative or zero. Then*

$$\lim_{n\to\infty}\frac{1}{n}\left|\left\{x:1\leq x\leq n,\nu(x)\geq\ln\ln n+\lambda\sqrt{\ln\ln n}\right\}\right|=\int_{\lambda}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt.$$

**Proof.** We outline the argument, emphasizing the similarities to Turán's proof. Fix a function $s(n)$ with $s(n)\to\infty$ and $s(n)=o\left((\ln\ln n)^{1/2}\right)$ — for example, $s(n)=\ln\ln\ln n$. Set $M=n^{1/s(n)}$. Set $X=\sum X_p$, the summation over all primes $p\leq M$. As no $x\leq n$ can have more than $s(n)$ prime factors greater than $M$ we have $\nu(x)-s(n)\leq X(x)\leq\nu(x)$ so that it suffices to show Theorem 4.2.2 with $\nu$ replaced by $X$. Let $Y_p$ be independent random variables with $\Pr\left[Y_p=1\right]=1/p$, $\Pr\left[Y_p=0\right]=1-1/p$ and set $Y=\sum Y_p$, the summation over all primes $p\leq M$. This $Y$ represents an idealized version of $X$. Set

$$\mu=\mathrm{E}\left[Y\right]=\sum_{p\leq M}\frac{1}{p}=\ln\ln n+o\left((\ln\ln n)^{1/2}\right)$$

and

$$\sigma^2=\mathrm{Var}\left[Y\right]=\sum_{p\leq M}\frac{1}{p}\left(1-\frac{1}{p}\right)\sim\ln\ln n$$

and define the normalized $\tilde{Y}=(Y-\mu)/\sigma$. From the Central Limit Theorem $\tilde{Y}$ approaches the standard normal $N$ and $\mathrm{E}\left[\tilde{Y}^k\right]\to\mathrm{E}\left[N^k\right]$ for every positive integer $k$. Set $\tilde{X}=(X-\mu)/\sigma$. We compare $\tilde{X},\tilde{Y}$.



For any distinct primes $p_1,\ldots,p_s\leq M$,

$$\mathrm{E}\left[X_{p_1}\cdots X_{p_s}\right]-\mathrm{E}\left[Y_{p_1}\cdots Y_{p_s}\right]=\frac{1}{n}\left\lfloor\frac{n}{p_1\cdots p_s}\right\rfloor-\frac{1}{p_1\cdots p_s}=O\left(\frac{1}{n}\right).$$

We let $k$ be an arbitrary fixed positive integer and compare $\mathrm{E}\left[\tilde{X}^k\right]$ and $\mathrm{E}\left[\tilde{Y}^k\right]$. Expanding, $\tilde{X}^k$ is a polynomial in $X$ with coefficients $n^{o(1)}$. Further expanding each $X^j=(\sum X_p)^j$ — always reducing $X_p^a$ to $X_p$ when $a\geq 2$ — gives the sum of $O(M^k)=n^{o(1)}$ terms of the form $X_{p_1}\cdots X_{p_s}$. The same expansion applies to $\tilde{Y}$. As the corresponding terms have expectations within $O(1/n)$ the total difference

$$\mathrm{E}\left[\tilde{X}^k\right]-\mathrm{E}\left[\tilde{Y}^k\right]=n^{-1+o(1)}=o(1).$$

Hence each moment of $\tilde{X}$ approaches that of the standard normal $N$. A standard, though nontrivial, theorem in probability theory gives that $\tilde{X}$ must therefore approach $N$ in distribution. ■

We recall the famous quotation of G. H. Hardy:

> 317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but *because it is so*, because mathematical reality is built that way.

How ironic — though not contradictory — that the methods of probability theory can lead to a greater understanding of the prime factorization of integers. Additional results applying information about the moments of a distribution in order to determine it appear in Chapter 8; see also Billingsley (1995).

## 4.3 MORE BASICS

Let $X$ be a nonnegative integral valued random variable and suppose we want to bound $\Pr\left[X=0\right]$ given the value $\mu=\mathrm{E}\left[X\right]$. If $\mu<1$ we may use the inequality

$$\Pr\left[X>0\right]\leq\mathrm{E}\left[X\right]$$

so that if $\mathrm{E}\left[X\right]\to 0$ then $X=0$ almost always. (Here we are imagining an infinite sequence of $X$ dependent on some parameter $n$ going to infinity.) But now suppose $\mathrm{E}\left[X\right]\to\infty$. It does *not* necessarily follow that $X>0$ almost always. For example, let $X$ be the number of deaths due to nuclear war in the twelve months after reading this paragraph. Calculation of $\mathrm{E}\left[X\right]$ can make for lively debate but few would deny that it is quite large. Yet we may believe — or hope — that $\Pr\left[X\neq 0\right]$ is very close to zero. We can sometimes deduce $X>0$ almost always if we have further information about $\mathrm{Var}\left[X\right]$.

**Theorem 4.3.1** $\Pr\left[X=0\right]\leq\dfrac{\mathrm{Var}\left[X\right]}{\mathrm{E}\left[X\right]^2}$.

**Proof.** Set $\lambda = \mu/\sigma$ in Chebyshev's Inequality. Then

$$\Pr[X = 0] \le \Pr[|X - \mu| \ge \lambda\sigma] \le \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}.$$

■

We generally apply this result in asymptotic terms.

**Corollary 4.3.2** *If* $\mathrm{Var}[X] = o(\mathrm{E}[X]^2)$ *then* $X > 0$ *almost always.*

The proof of Theorem 4.3.1 actually gives that, for any $\epsilon > 0$,

$$\Pr[|X - \mathrm{E}[X]| \ge \epsilon \mathrm{E}[X]] \le \frac{\mathrm{Var}[X]}{\epsilon^2 \mathrm{E}[X]^2}$$

and thus in asymptotic terms we actually have the following stronger assertion.

**Corollary 4.3.3** *If* $\mathrm{Var}[X] = o(\mathrm{E}[X]^2)$ *then* $X \sim \mathrm{E}[X]$ *almost always.*

Suppose again $X = X_1 + \cdots + X_m$, where $X_i$ is the indicator random variable for event $A_i$. For indices $i, j$ write $i \sim j$ if $i \neq j$ and the events $A_i, A_j$ are not independent. We set (the sum is over ordered pairs)

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

Note that when $i \sim j$,

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

and that when $i \neq j$ and not $i \sim j$ then $\mathrm{Cov}[X_i, X_j] = 0$. Thus

$$\mathrm{Var}[X] \le \mathrm{E}[X] + \Delta.$$

**Corollary 4.3.4** *If* $\mathrm{E}[X] \to \infty$ *and* $\Delta = o(\mathrm{E}[X]^2)$ *then* $X > 0$ *almost always. Furthermore* $X \sim \mathrm{E}[X]$ *almost always.*

Let us say $X_1, \ldots, X_m$ are *symmetric* if for every $i \neq j$ there is a measure preserving mapping of the underlying probability space that sends event $A_i$ to event $A_j$. Examples will appear in the next section. In this instance we write

$$\Delta = \sum_{i \sim j} \Pr[A_i \wedge A_j] = \sum_i \Pr[A_i] \sum_{j \sim i} \Pr[A_j \mid A_i]$$

and note that the inner summation is independent of $i$. We set

$$\Delta^* = \sum_{j \sim i} \Pr[A_j \mid A_i],$$



where $i$ is any fixed index. Then

$$\Delta = \sum_i \Pr[A_i]\Delta^* = \Delta^* \sum_i \Pr[A_i] = \Delta^* \mathrm{E}[X].$$

**Corollary 4.3.5** *If* $\mathrm{E}[X] \to \infty$ *and* $\Delta^* = o(\mathrm{E}[X])$ *then* $X > 0$ *almost always. Furthermore* $X \sim \mathrm{E}[X]$ *almost always.*

The condition of Corollary 4.3.5 has the intuitive sense that conditioning on any specific $A_i$ holding does not substantially increase the expected number $\mathrm{E}[X]$ of events holding.

## 4.4 RANDOM GRAPHS

The random graph $G(n,p)$ is, informally, the graph on $n$ labeled vertices, obtained by selecting each pair of vertices to be an edge, randomly and independently, with probability $p$. A property of graphs is a family of graphs closed under isomorphism. A function $r(n)$ is a *threshold function* for some property $P$, if whenever $p = p(n) \ll r(n)$ then $G(n,p)$ does not satisfy $P$ almost always, and whenever $p \gg r(n)$ then $G(n,p)$ satisfies $P$ almost always. For more precise definitions of the random graph $G(n,p)$ and of threshold functions, see Section 10.1.

The results of this section are generally surpassed by those of Chapter 10 but they were historically the first results and provide a good illustration of the second moment. We begin with a particular example. By $\omega(G)$ we denote here and in the rest of the book the number of vertices in the maximum clique of the graph $G$.

**Theorem 4.4.1** *The property* $\omega(G) \ge 4$ *has threshold function* $n^{-2/3}$.

**Proof.** For every 4-set $S$ of vertices in $G(n,p)$ let $A_S$ be the event "$S$ is a clique" and $X_S$ its indicator random variable. Then

$$\mathrm{E}[X_S] = \Pr[A_S] = p^6$$

as six different edges must all lie in $G(n,p)$. Set

$$X = \sum_{|S|=4} X_S$$

so that $X$ is the number of 4-cliques in $G$ and $\omega(G) \ge 4$ if and only if $X > 0$. Linearity of expectation gives

$$\mathrm{E}[X] = \sum_{|S|=4} \mathrm{E}[X_S] = \binom{n}{4} p^6 \sim \frac{n^4 p^6}{24}.$$

When $p(n) \ll n^{-2/3}$, $\mathrm{E}[X] = o(1)$ and so $X = 0$ almost surely.

Now suppose $p(n) \gg n^{-2/3}$ so that $\mathrm{E}[X] \to \infty$ and consider the $\Delta^*$ of Corollary 4.3.5. (All 4-sets "look the same" so that the $X_S$ are symmetric.) Here $S \sim T$ if and only if $S \neq T$ and $S, T$ have common edges; that is, if and only if $|S \cap T| = 2$ or 3. Fix $S$. There are $O(n^2)$ sets $T$ with $|S \cap T| = 2$ and for each of these $\Pr[A_T \mid A_S] = p^5$. There are $O(n)$ sets $T$ with $|S \cap T| = 3$ and for each of these $\Pr[A_T \mid A_S] = p^3$. Thus

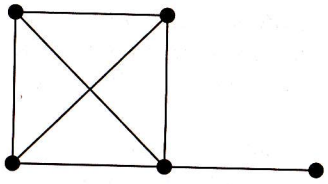
$$\Delta^* = O(n^2 p^5) + O(np^3) = o(n^4 p^6) = o(\mathrm{E}[X])$$

since $p \gg n^{-2/3}$. Corollary 4.3.5 therefore applies and $X > 0$; that is, there *does* exist a clique of size 4, almost always. ■

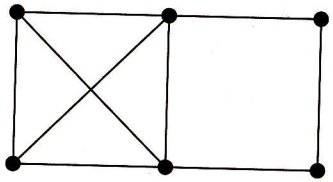
The proof of Theorem 4.4.1 appears to require a fortuitous calculation of $\Delta^*$. The following definitions pave the way for the more general Theorem 4.4.2.

**Definition 1** *Let $H$ be a graph with $v$ vertices and $e$ edges. We call $\rho(H) = e/v$ the* density *of $H$. We call $H$* balanced *if every subgraph $H'$ has $\rho(H') \leq \rho(H)$. We call $H$* strictly balanced *if every proper subgraph $H'$ has $\rho(H') < \rho(H)$.*

**Examples.** $K_4$ and, in general, $K_k$ are strictly balanced. The graph



is not balanced as it has density $7/5$ while the subgraph $K_4$ has density $3/2$. The graph



is balanced but not strictly balanced as it and its subgraph $K_4$ have density $3/2$.

**Theorem 4.4.2** *Let $H$ be a balanced graph with $v$ vertices and $e$ edges. Let $A(G)$ be the event that $H$ is a subgraph (not necessarily induced) of $G$. Then $p = n^{-v/e}$ is the threshold function for $A$.*

**Proof.** We follow the argument of Theorem 4.4.1. For each $v$-set $S$ let $A_S$ be the event that $G|_S$ contains $H$ as a subgraph. Then

$$p^e \leq \Pr[A_S] \leq v! p^e .$$



(Any particular placement of $H$ has probability $p^e$ of occurring and there are at most $v!$ possible placements. The precise calculation of $\Pr[A_S]$ is, in general, complicated due to the overlapping of potential copies of $H$.) Let $X_S$ be the indicator random variable for $A_S$ and

$$X = \sum_{|S|=v} X_S$$

so that $A$ holds if and only if $X > 0$. Linearity of expectation gives

$$\mathrm{E}[X] = \sum_{|S|=v} \mathrm{E}[X_S] = \binom{n}{v} \Pr[A_S] = \Theta(n^v p^e) .$$

If $p \ll n^{-v/e}$ then $\mathrm{E}[X] = o(1)$, so $X = 0$ almost always.

Now assume $p \gg n^{-v/e}$ so that $\mathrm{E}[X] \to \infty$ and consider the $\Delta^*$ of Corollary 4.3.5. (All $v$-sets look the same so the $X_S$ are symmetric.) Here $S \sim T$ if and only if $S \neq T$ and $S, T$ have common edges; that is, if and only if $|S \cap T| = i$ with $2 \leq i \leq v-1$. Let $S$ be fixed. We split

$$\Delta^* = \sum_{T \sim S} \Pr[A_T \mid A_S] = \sum_{i=2}^{v-1} \sum_{|T \cap S| = i} \Pr[A_T \mid A_S] .$$

For each $i$ there are $O(n^{v-i})$ choices of $T$. Fix $S, T$ and consider $\Pr[A_T \mid A_S]$. There are $O(1)$ possible copies of $H$ on $T$. Each has — since, critically, $H$ is balanced — at most $ie/v$ edges with both vertices in $S$ and thus at least $e - (ie/v)$ other edges. Hence

$$\Pr[A_T \mid A_S] = O(p^{e-(ie/v)})$$

and

$$\begin{aligned} \Delta^* &= \sum_{i=2}^{v-1} O(n^{v-i} p^{e-(ie/v)}) \\ &= \sum_{i=2}^{v-1} O((n^v p^e)^{1-i/v}) \\ &= \sum_{i=2}^{v-1} o(n^v p^e) \\ &= o(\mathrm{E}[X]) \end{aligned}$$

since $n^v p^e \to \infty$. Hence Corollary 4.3.5 applies. ■

**Theorem 4.4.3** *In the notation of Theorem 4.4.2 if $H$ is* not *balanced then $p = n^{-v/e}$ is* not *the threshold function for $A$.*

**Proof.** Let $H_1$ be a subgraph of $H$ with $v_1$ vertices, $e_1$ edges and $e_1/v_1 > e/v$. Let $\alpha$ satisfy $v_1/e_1 < \alpha < v/e$ and set $p = n^{-\alpha}$. The expected number of copies of $H_1$

is then $o(1)$ so almost always $G(n,p)$ contains no copy of $H_1$. But if it contains no copy of $H_1$ then it surely can contain no copy of $H$. ■

The threshold function for the property of containing a copy of $H$, for general $H$, was examined in the original papers of Erdős and Rényi (1960). It still provides an excellent introduction to the theory of random graphs. Let $H_1$ be that subgraph with maximal density $\rho(H_1) = e_1/v_1$. (When $H$ is balanced we may take $H_1 = H$.) They showed that $p = n^{-v_1/e_1}$ is the threshold function. We do not show this here though it follows fairly straightforwardly from these methods.

We finish this section with two strengthenings of Theorem 4.4.2.

**Theorem 4.4.4** *Let $H$ be strictly balanced with $v$ vertices, $e$ edges and $a$ automorphisms. Let $X$ be the number of copies of $H$ in $G(n,p)$. Assume $p \gg n^{-v/e}$. Then almost always*

$$X \sim \frac{n^v p^e}{a}.$$

**Proof.** Label the vertices of $H$ by $1,\ldots,v$. For each ordered $x_1,\ldots,x_v$ let $A_{x_1,\ldots,x_v}$ be the event that $x_1,\ldots,x_v$ provides a copy of $H$ in that order. Specifically we define

$$A_{x_1,\ldots,x_v} : \{i,j\} \in E(H) \Rightarrow \{x_i,x_j\} \in E(G).$$

We let $I_{x_1,\ldots,x_v}$ be the corresponding indicator random variable. We define an equivalence class on $v$-tuples by setting $(x_1,\ldots,x_v) \equiv (y_1,\ldots,y_v)$ if there is an automorphism $\sigma$ of $V(H)$ so that $y_{\sigma(i)} = x_i$ for $1 \le i \le v$. Then

$$X = \sum I_{x_1,\ldots,x_v}$$

gives the number of copies of $H$ in $G$ where the sum is taken over one entry from each equivalence class. As there are $(n)_v/a$ terms,

$$\mathrm{E}[X] = \frac{(n)_v}{a}\mathrm{E}[I_{x_1,\ldots,x_v}] = \frac{(n)_v p^e}{a} \sim \frac{n^v p^e}{a}.$$

Our assumption $p \gg n^{-v/e}$ implies $\mathrm{E}[X] \to \infty$. It suffices therefore to show $\Delta^* = o(\mathrm{E}[X])$. Fixing $x_1,\ldots,x_v$,

$$\Delta^* = \sum_{(y_1,\ldots,y_v)\sim(x_1,\ldots,x_v)} \Pr\left[A_{(y_1,\ldots,y_v)} \mid A_{(x_1,\ldots,x_v)}\right].$$

There are $v!/a = O(1)$ terms with $\{y_1,\ldots,y_v\} = \{x_1,\ldots,x_v\}$ and for each the conditional probability is at most 1 (actually, at most $p$), thus contributing $O(1) = o(\mathrm{E}[X])$ to $\Delta^*$. When $\{y_1,\ldots,y_v\} \cap \{x_1,\ldots,x_v\}$ has $i$ elements, $2 \le i \le v-1$ the argument of Theorem 4.4.2 gives that the contribution to $\Delta^*$ is $o(\mathrm{E}[X])$. Altogether $\Delta^* = o(\mathrm{E}[X])$ and we apply Corollary 4.3.5. ■



**Theorem 4.4.5** *Let $H$ be* any *fixed graph. For every subgraph $H'$ of $H$ (including $H$ itself) let $X_{H'}$ denote the number of copies of $H'$ in $G(n,p)$. Assume $p$ is such that $\mathrm{E}[X_{H'}] \to \infty$ for every $H'$. Then*

$$X_H \sim \mathrm{E}[X_H]$$

*almost always.*

**Proof.** Let $H$ have $v$ vertices and $e$ edges. As in Theorem 4.4.4 it suffices to show $\Delta^* = o(\mathrm{E}[X])$. We split $\Delta^*$ into a finite number of terms. For each $H'$ with $w$ vertices and $f$ edges we have those $(y_1,\ldots,y_v)$ that overlap with the fixed $(x_1,\ldots,x_v)$ in a copy of $H'$. These terms contribute, up to constants,

$$n^{v-w}p^{e-f} = \Theta\left(\frac{\mathrm{E}[X_H]}{\mathrm{E}[X_{H'}]}\right) = o(\mathrm{E}[X_H])$$

to $\Delta^*$. Hence Corollary 4.3.5 does apply. ■

## 4.5 CLIQUE NUMBER

Now we fix edge probability $p = \frac{1}{2}$ and consider the clique number $\omega(G)$. We set

$$f(k) = \binom{n}{k}2^{-\binom{k}{2}},$$

the expected number of $k$-cliques. The function $f(k)$ drops under one at $k \sim 2\log_2 n$. [Very roughly, $f(k)$ is like $n^k 2^{-k^2/2}$.]

**Theorem 4.5.1** *Let $k = k(n)$ satisfy $k \sim 2\log_2 n$ and $f(k) \to \infty$. Then almost always $\omega(G) \ge k$.*

**Proof.** For each $k$-set $S$ let $A_S$ be the event "$S$ is a clique" and $X_S$ the corresponding indicator random variable. We set

$$X = \sum_{|S|=k} X_S$$

so that $\omega(G) \ge k$ if and only if $X > 0$. Then $\mathrm{E}[X] = f(k) \to \infty$ and we examine the $\Delta^*$ of Corollary 4.3.5. Fix $S$ and note that $T \sim S$ if and only if $|T \cap S| = i$, where $2 \le i \le k-1$. Hence

$$\Delta^* = \sum_{i=2}^{k-1}\binom{k}{i}\binom{n-k}{k-i}2^{\binom{i}{2}-\binom{k}{2}}$$