THE PROBABILISTIC METHOD, RANDOM GRAPHS AND STEIN'S METHOD

MICHAEL DRMOTA

1. Introduction

The **Probabilistic Method** has been initiated by Paul Erdős [6] in order to prove the existence of certain combinatorial objects. The principle idea is to define a proper probability distribution on a class of (discrete) objects and to show that the probability of a certain property is positive. Of course this also proves that there exists such an object with this property. We will apply this approach to various problems on **random graphs**.

However, the main goal of this course is to give an introduction to **Stein's method** that proves asymptotic normality for sums of (in some sense) weakly dependent random variables. This method has turned out to be very successful, in particular in random graph problems.

There is vast literature on these topics. We just mention few books that are exclusively devoted to them [1, 5, 9, 10].

2. Lower Bound for the Ramsey Number

Let us start with a classical example:

Definition 2.1. *The* **Ramsey number** $R(k,l)$ *is the smallest number* n *such that any 2-coloring of the edges on the conplete graph* K_n *on* n *vertices contains either a monochromatic* K_k *(in* K_n*) of the first color or a monochromatic* K_l *(in* K_n*) of the second color.*

Ramsey's theorem says that $R(k,l)$ exists for all positive integers k and l. For example, it is known that $R(k,k) \le (4+o(1))^k$. However, we are more interested in lower bounds.

Theorem 2.2. *We have* $R(k,k) > 2^{k/2}$ *for all* $k \ge 3$.

Proof. We consider a complete graph K_n with vertex set $\{1,2,\ldots\}$ and a random coloring of the $\binom{n}{2}$ edges with 2 colors. (Each edge is colored independently and with equal probability $\frac{1}{2}$.) Let $R \subseteq \{1,2,\ldots\}$ a set of size k and A_R the event

$$A_R = \{\text{the induced subgraph of } R \text{ is monochromatic}\}.$$

Date: July 24, 2006.

Supported by the Austrian Science Foundation (FWF) FSP-Project S9604.

Then

$$\mathbb{P}(A_R) = 2^{1-\binom{k}{2}} = 2\,\frac{1}{2^{\binom{k}{2}}}.$$

Consequently

$$\mathbb{P}\{\exists\, R \subseteq \{1,2,\ldots\} : |R| = k,\ A_R \text{ occurs}\} \le \binom{n}{k} 2^{1-\binom{k}{2}}.$$

If $n = \lfloor 2^{k/2} \rfloor$ then

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 2\frac{n^k}{k!}\frac{1}{2^{k^2/2-k/2}} \le 2\frac{2^{k/2}}{k!} < 1.$$

Thus

$$\mathbb{P}\{\forall\, R \subseteq \{1,2,\ldots\} : |R| = k,\ R \text{ is not monochromatic}\} > 0$$

and it follows that there exists a 2-coloring of K_n for which each induced subgraph of size k is not monochromatic. □

Note that the proof of Theorem 2.2 also shows that there almost always[1] exists no monochromatic K_k in a random coloring of K_n with $k = \lfloor 2\log_2 n \rfloor$. We just have to observe that $2\,2^{k/2}/k! \to 0$ as $k \to \infty$.

3. First Moment Method

The first moment $\mathbb{E}\,X$ of a random variable X gives only a partial information on the behaviour of X. On the other hand it is (usually) easy to compute. For example, if X can be written as a sum of random variables, $X = \sum Y_i$ then we have

$$\mathbb{E}\,X = \sum_i \mathbb{E}\,Y_i$$

even if there is strong dependence between the Y_i.

Nevertheless, there are at least some usueful properties for X that can be deduced from $\mathbb{E}\,X$.

Theorem 3.1. *Suppose that $\mathbb{E}\,X$ is finite then*

$$\mathbb{P}\{X \le \mathbb{E}\,X\} > 0 \quad \text{and} \quad \mathbb{P}\{X \ge \mathbb{E}\,X\} > 0.$$

Proof. First note that if Y is random variable that is strictly positive, $Y > 0$, then we also have $\mathbb{E}\,Y > 0$.

Now, if $\mathbb{P}\{X \le \mathbb{E}\,X\} = 0$ then $\mathbb{P}\{X > \mathbb{E}\,X\} = 1$ and

$$\mathbb{E}\,X = \mathbb{E}\,\left(\mathbb{I}_{[X>\mathbb{E}\,X]}\cdot X\right) = \mathbb{E}\,X + \mathbb{E}\,\left(\mathbb{I}_{[X>\mathbb{E}\,X]}\cdot (X - \mathbb{E}\,X)\right) > \mathbb{E}\,X$$

leads to a contradiction. □

Another theorem that applies for non-negative integer values random variables is also quite useful.

[1]We use the notion *almost always* as an abbreviation for the property that the probability that a certion condition holds converges to 1 as the *size* of the problem goes to the infinity.

Theorem 3.2. *Suppose that X is a discrete random variable with non-negative integer values. Then*

$$\mathbb{P}\{X > 0\} \leq \mathbb{E}\, X.$$

Proof.

$$\mathbb{E}\, X = \sum_{k\geq 0} k\, \mathbb{P}\{X = k\} \leq \sum_{k\geq 1} \mathbb{P}\{X = k\} = \mathbb{P}\{X > 0\}.$$

□

As an first application we prove Theorem 3.2 a second time.

As above let K_n denote the complete graph of n nodes and adjust K_n with a random edge coloring (with 2 colors). Fix k and let $\mathcal{S}_{n,k}$ denote the set of all subgraphs of K_n with k nodes. Then

$$X_n := \sum_{R\in\mathcal{S}_{n,k}} \mathbb{I}_{[R \text{ is monochromatic}]}$$

is the (random) number of monochromatic subgraphs of K_n that are isomorphic to K_k. Hence

$$\begin{aligned}\mathbb{E}\, X_n &= \sum_{R\in\mathcal{S}_{n,k}} \mathbb{P}\{R \text{ is monochromatic}\}\\ &= \binom{n}{k} 2\; 2^{-\binom{k}{2}}.\end{aligned}$$

Thus, by the first moment method it follows that

$$\mathbb{P}\{X_n > 0\} \leq \binom{n}{k} 2^{1-\binom{k}{2}}$$

and we can proceed as above.

We apply these first moment methods to three other problems.

Theorem 3.3. *Let $v_1, \dots, v_n \in \mathbb{R}^n$ with unit length $|v_i| = 1$. Then there exist $\varepsilon_1, \dots, \varepsilon_n \in \{-1, +1\}$ with*

$$|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n| \leq \sqrt{n}$$

and also there exits $\varepsilon_1, \dots, \varepsilon_n \in \{-1, +1\}$ with

$$|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n| \geq \sqrt{n}.$$

Proof. Consider the random number

$$X = \left| \sum_{i=1}^{n} \varepsilon_i v_i \right|^2,$$

where $\varepsilon \in \{-1, 1\}$ are independently and randomly chosen with equal probability $\frac{1}{2}$. Since

$$X = \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j v_i \cdot v_j$$

we have

$$\mathbb{E}\, X = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\,(\varepsilon_i \varepsilon_j) v_i \cdot v_j = \sum_{i=1}^{n} v_i \cdot v_i = n.$$

Hence, a direct application of Theorem 3.1 completes the proof. □

Definition 3.4. *A set of nodes I in a graph G is called* **independent** *if no two nodes of I are adjacent.*

The **independence number** $\alpha(G)$ *of G is the maximal size of an independent set of nodes of G.*

Theorem 3.5. *Let G be a graph with n nodes and $m \ge n/2$ edges. Then*

$$\alpha(G) \ge \frac{n^2}{4m}.$$

Proof. We set $p = n/(2m)$. By assumption $0 \le p \le 1$. Let $V = \{v_1, v_2, \ldots, v_n\}$ denote the vertex set of G. We now choose a random subset S of V where each vertex is chosen independently with probability p, that is $\mathbb{P}\{v_i \in S\} = p$. Let $X = |S|$ the (random) size of S and Y the (random) number of edges in $G|_S$. Equivalently,

$$Y = \sum_{e \in E} \mathbb{I}_{[\text{both endpoints of } e \text{ are in } S]},$$

where E denotes the edge set of G. Hence,

$$\mathbb{E}\, Y = \sum_{e \in E} p^2 = mp^2.$$

Further $\mathbb{E}\, X = np$ and consequently

$$\mathbb{E}\,(X - Y) = np - mp^2 = \frac{n^2}{4m}.$$

Thus, there exists some specific S for which the number of vertices of S minus the number of edges of S is at least $n^2/(4m)$. Select one vertex from each edge of S and delete it. This leaves a set S^* with at least $n^2/(4m)$ vertices. Since all edges of S have been destroyed the set S^* is an independent set. Hence $\alpha(G) \ge \frac{n^2}{4m}$. □

Definition 3.6. *The* **girth** $\mathrm{girth}(G)$ *of a graph G is the size of the shortest cycle.*

The **chromatic number** $\chi(G)$ *of a graph G is the smallest number k such that there exists a regular k-coloring of the vertices of G, that is, a coloring of at k colors of the vertices such that adjacent vertices have different colors.*

Theorem 3.7 (Erdős 1959). *For all (positive integers) k and ℓ there exits a graph G with* $\mathrm{girth}(G) > \ell$ *and* $\chi(G) > k$.

Proof. Fix a positive $\theta < 1/\ell$ and set $p = n^{\theta - 1}$, where n will be chosen sufficiently large. Let $\{1, 2, \ldots, n\}$ denote the vertex set of a random graph where we include (undirected) edges between different vertices independently with probability p. [2]

Let X denote the number of cycles of size at most ℓ. Then

$$\mathbb{E}\,X = \sum_{i=3}^{\ell} \frac{(n)_i}{2i} p^i \leq \sum_{i=3}^{\ell} \frac{n^i}{2i} n^{(\theta-1)i} = \sum_{i=3}^{\ell} \frac{n^{\theta i}}{2i} = o(n)$$

as $\theta\ell < 1$. In particular with

$$\mathbb{E}\,X \geq \mathbb{E}\left(X \cdot \mathbb{I}_{[X \geq n/2]}\right) \geq \frac{n}{2}\,\mathbb{P}\{X \geq n/2\}$$

this implies

$$\mathbb{P}\{X \geq n/2\} = o(1).$$

Next we use the fact that

$$\begin{aligned}
\mathbb{P}\{\alpha(G) \geq m\} &= \mathbb{P}\{\exists\, S \subseteq \{1,2,\ldots,n\} : |S| = m,\ S \text{ is independent}\} \\
&\leq \mathbb{E}\left(\sum_{|S|=m} \mathbb{I}_{[S \text{ is independent}]}\right) \\
&= \sum_{|S|=m} \mathbb{P}\{S \text{ is independent}\} \\
&= \binom{n}{m}(1-p)^{\binom{m}{2}} \\
&\leq \frac{n^m}{m!} e^{-p\binom{m}{2}} \\
&\leq (ne^{-p(m-1)/2})^m
\end{aligned}$$

If we use $m = \lceil \frac{3}{p} \log n \rceil$ then $ne^{-p(m-1)/2} \to 0$ as $n \to \infty$ and consequently $\mathbb{P}\{\alpha(G) \geq m\} \to 0$.

Let n be sufficiently large so that both these events have probability less that $\frac{1}{2}$. Then there is a specific G with less than $n/2$ cycles of length at most ℓ and with $\alpha(G) < 3n^{1-\theta} \log n$. Remove from G a vertex from each cycle of length at most ℓ. This gives a graph G^* with at least $n/2$ vertices. G^* has girth greater that ℓ and $\alpha(G^*) \leq \alpha(G)$. Thus,

$$\chi(G^*) \geq \frac{|G^*|}{\alpha(G)} \geq \frac{n/2}{3n^{1-\theta}\log n} = \frac{n^\theta}{6\log n}.$$

Finally we just have to choose n sufficiently large that $n^\theta/(6\log n) > k$. $\square$

4. Second Moment Method

The second moment $\mathbb{E}\,(X^2)$ and the variance $\mathbb{V}X = \mathbb{E}\,(X^2) - (EX)^2 = \mathbb{E}\,((X - \mathbb{E}\,X)^2)$ give much more insight into the behaviour of X than the first moment $\mathbb{E}\,X$ since the variance takes the deviation from the mean into

[2]This is exactly the random graph model $G(n,p)$ that will be discussed in more detail later.

account. This observation is quantified by Chebyshev's Inequality. The use of Chebyshev's Inequality is also called the *Second Moment Method*.

Theorem 4.1 (Chebyshev's Inequality)**.** *Suppose that X has finite second moment. Then*

$$\mathbb{P}\{|X - \mathbb{E}\,X| \ge \lambda\sqrt{\mathbb{V}X}\} \le \frac{1}{\lambda^2}. \tag{4.1}$$

Proof. By definition we have

$$\begin{aligned}\mathbb{V}X &= \mathbb{E}\,((X-\mathbb{E}\,X)^2)\\ &\ge \mathbb{E}\,\left((X-\mathbb{E}\,X)^2\mathbb{I}_{[|X-\mathbb{E}\,X|\ge\kappa]}\right)\\ &\ge \kappa^2\mathbb{P}\{|X-\mathbb{E}\,X|\ge\kappa\}.\end{aligned}$$

Hence, with $\kappa = \lambda\cdot\sqrt{\mathbb{V}X}$ we get (4.1). □

As a consequence we get the following very useful estimate that applies for non-negative integer values random variables.

Theorem 4.2. *Suppose that X is a discrete random variable with non-negative integer values. Then*

$$\mathbb{P}\{X = 0\} \le \frac{\mathbb{V}X}{(\mathbb{E}\,X)^2}. \tag{4.2}$$

Proof. Set $\lambda = \mathbb{E}\,X/\sqrt{\mathbb{V}X}$ in (4.1). Then

$$\mathbb{P}\{X=0\} \le \mathbb{P}\{|X-\mathbb{E}\,X| \ge \lambda\sqrt{\mathbb{V}X}\} \le \frac{1}{\lambda^2} = \frac{\mathbb{V}X}{(\mathbb{E}\,X)^2}.$$

□

Note that (4.2) can be slightly sharpened. From $\mathbb{E}\,X = \mathbb{E}\,(X\cdot\mathbb{I}_{[X>0]}) \le \sqrt{\mathbb{E}\,X^2}\cdot\sqrt{\mathbb{P}\{X>0\}}$ we get

$$\mathbb{P}\{X>0\} \ge \frac{(\mathbb{E}\,X)^2}{\mathbb{E}\,X^2}$$

which is equivalent to $\mathbb{P}\{X = 0\} \le \mathbb{V}X/\mathbb{E}\,X^2$. This also complements the inequality $\mathbb{P}\{X > 0\} \le \mathbb{E}\,X$.

Another application of Chebychev's inequality is the following property.

Theorem 4.3. *Suppose that X_n is a sequence of random variables with $\mathbb{E}\,X_n \to \infty$ and $\mathbb{E}\,(X_n)^2 \sim (\mathbb{E}\,X_n)^2$ as $n\to\infty$. Then we have almost always $X_n > 0$ and*

$$\frac{X_n}{\mathbb{E}\,X_n} \to 1.$$

Proof. For every $\varepsilon > 0$ (4.1) implies

$$\mathbb{P}\{|X_n - \mathbb{E}\,X_n| \ge \varepsilon\,\mathbb{E}\,X_n\} \le \frac{\mathbb{V}X_n}{\varepsilon^2(\mathbb{E}\,X_n)^2}.$$

Now note that $\mathbb{E}\,(X_n)^2 \sim (\mathbb{E}\,X_n)^2$ is equivalent to $\mathbb{V}X_n = o((\mathbb{E}\,X_n)^2)$. Thus, $\mathbb{P}\{|X_n - \mathbb{E}\,X_n| \ge \varepsilon\,\mathbb{E}\,X_n\} \to 0$ and consequently $X_n \sim \mathbb{E}\,X_n$ almost always. □

The limit relation $\mathbb{P}\{|X_n - \mathbb{E}\,X_n| \geq \varepsilon\,\mathbb{E}\,X_n\} \to 0$ also says that X is concentrated around its mean. Thus, the property $\mathbb{E}\,(X_n)^2 \sim (\mathbb{E}\,X_n)^2$ implies this **concentration property**.

We now apply this procedure to the above example, where $X = X_n$ denotes the number of copies os K_k in a random graph $G(n,p)$ (see below). For simplicity we will only count triangles, that is, $k = 3$.

Let $\mathcal{T}$ denote the set of triangles in $G(n,p)$. Then

$$X = \sum_{1\leq i<j<k\leq n} \mathbb{I}_{[(i,j,k)\in\mathcal{T}]}$$

and

$$\mathbb{E}\,X = \sum_{1\leq i<j<k\leq n} \mathbb{P}\{(i,j,k)\in\mathcal{T}\} = \binom{n}{3}p^3.$$

When we compute the second moment we have to be a little bit more careful. Formally we have

$$\begin{aligned}
\mathbb{E}\,X^2 &= \mathbb{E}\left(\sum_{1\leq i_1<i_2<i_3\leq n}\ \sum_{1\leq j_1<j_2<j_3\leq n} \mathbb{I}_{[(i_1,i_2,i_3)\in\mathcal{T}]}\cdot\mathbb{I}_{[(j_1,j_2,j_3)\in\mathcal{T}]}\right)\\
&= \mathbb{E}\left(\sum_{1\leq i_1<i_2<i_3\leq n}\ \sum_{1\leq j_1<j_2<j_3\leq n} \mathbb{I}_{[(i_1,i_2,i_3),(j_1,j_2,j_3)\in\mathcal{T}]}\right)\\
&= \sum_{1\leq i_1<i_2<i_3\leq n}\ \sum_{1\leq j_1<j_2<j_3\leq n} \mathbb{P}\{(i_1,i_2,i_3),(j_1,j_2,j_3)\in\mathcal{T}\}
\end{aligned}$$

Here we have to distinguish between several cases.

(1) If $|\{i_1,i_2,i_3\}\cap\{j_1,j_2,j_3\}| = 3$, that is, $i_1 = j_1$, $i_2 = j_2$, and $i_3 = j_3$ then

$$\mathbb{P}\{(i_1,i_2,i_3),(j_1,j_2,j_3)\in\mathcal{T}\} = p^3$$

and there are $\binom{n}{3}$ cases of that kind.

(2) If $|\{i_1,i_2,i_3\}\cap\{j_1,j_2,j_3\}| = 2$ then

$$\mathbb{P}\{(i_1,i_2,i_3),(j_1,j_2,j_3)\in\mathcal{T}\} = p^5$$

and there are $12\binom{n}{4}$ cases of that kind.

(3) If $|\{i_1,i_2,i_3\}\cap\{j_1,j_2,j_3\}| \leq 1$ then the events $\{(i_1,j_1,k_1)\in\mathcal{T}\}$ and $\{(i_2,j_2,k_2)\in\mathcal{T}\}$ are independent and consequently

$$\mathbb{P}\{(i_1,j_1,k_1),(i_2,j_2,k_2)\in\mathcal{T}\} = p^6.$$

Thus,

$$\begin{aligned}
\mathbb{E}\,X^2 &= \binom{n}{3}p^3 + 12\binom{n}{4}p^5 + \left(\binom{n}{3}^2 - \binom{n}{3} - 12\binom{n}{4}\right)p^6\\
&= (\mathbb{E}\,X)^2 + \binom{n}{3}p^3(1-p^3) + 12\binom{n}{4}p^5(1-p).
\end{aligned}$$

Here we have $\mathbb{E}\, X^2 \sim (\mathbb{E}\, X)^2$ (and also $\mathbb{E}\, X \to \infty$) if and only if $np \to \infty$. Assuming that we obain that almost always the number of triangles X in $G(n,p)$ is approximated by the expected number of triangles $\mathbb{E}\, X = \binom{n}{3} p^3$.

Obviously the same procecure works for general $k \ge 3$, and it is also possible to cover the even more general case of counting subgraphs that are isomorphic to a given graph H but we will not work out this case.

5. Random Graphs

In the previous examples we have usesd several times a random graph (or a random coloring on the complete graph). We will now introduce a notion that makes these things precise and has also become standard in the literature.

Definition 5.1. *Let n be a positive integer and p a real number with $0 \le p \le 1$. The* **random graph** *$G(n,p)$ is a probability space over the set of graphs on the vertex set $\{1, 2, \ldots, n\}$ determined by*

$$\mathbb{P}\{(i,j) \in G\} = p$$

for all possible (undirected) edges (i,j) with $1 \le i, j \le$ and $i \ne j$ with these events mutually independent.

Similarly one also considers random graphs $G(n,m)$, where m is also a given integer with $0 \le m \le \binom{n}{2}$. Here one considers the set of all graphs on the set of vertices $\{1, 2, \ldots, n\}$ with exactly m (undirected) edges where each of these graphs is equally likely. Due to the law of large numbers $G(n,m)$ will have very similar properties as $G(n,p)$ with $p = m/\binom{n}{2}$. However, in this course we will only work the the $G(n,p)$-model.

We have implicitly used this kind of notion in several previous examples. For example a random coloring is modelled by $G(n, \frac{1}{2})$. In this context we have counted (more or less) the number X of subgraphs that are isomorphic to as complete graph K_k and have determined its expected value. For general p we get (in completely the same way)

$$\mathbb{E}\, X = \binom{n}{k} p^{\binom{k}{2}}.$$

We now come back to coloring problems. The next property for the chromatic number in $G(n, \frac{1}{2})$ is due to Bollobas.

Theorem 5.2. *We have, almost always in $G(n, \frac{1}{2})$,*

$$\chi(G) \sim \frac{n}{2 \log_2 n}.$$

Sketch of the proof. We will give a full proof of the upper bound. The (complicated) proof of the lower bound is just indicated.

Let $\alpha(G)$ denote the independece number of G. The proof of Theorem 2.2 also shows (and that has been also noted after its proof) that almost always there exists no complete subgraph $K_{\lfloor 2 \log_2 n \rfloor}$ in $G(n, \frac{1}{2})$. This also holds for the complement. Consequently almost always there is no independent set of size $\lfloor 2\log_2 n \rfloor$. Hence,

$$\chi(G) \geq \frac{n}{\alpha(G)} \geq \frac{n}{2\log_2 n}.$$

For the proof of the lower bound we use the abbreviation $m = \lfloor n/(\log n)^4 \rfloor$. For any set S of m vertices let $G|_S$ denote the restriction of to G to S. Obviously, $G|_S$ has the distribution $G(m, \frac{1}{2})$. Let $k = k(m) = k_0(m) - 4$, where $k_0 = k_0(n)$ is defined by

$$\binom{n}{k_0 - 1} 2^{-\binom{k_0-1}{2}} > 1 > \binom{n}{k_0} 2^{-\binom{k_0}{2}}$$

and note that $k \sim 2\log_2 m \sim 2\log_2 n$.

The crutial step is to show an inequality of the form

$$\mathbb{P}\{\alpha(G|_S) < k\} < e^{-m^{2+o(1)}}. \tag{5.1}$$

The proof of (5.1) goes beyond the scope of this course (it can be either proved via Azuma's inequality on martingales or with help of correlation estimates).

There are now $\binom{n}{m} < 2^n = 2^{m^{1+o(1)}}$ such sets S. Hence

$$\mathbb{P}\{\alpha(G|_S) < k \text{ for some } m\text{-set } S\} < 2^{m^{1+o(1)}} e^{-m^{2+o(1)}} = o(1).$$

That is, always always *every* m vertices contain a k-element independent set.

Now suppose that G has this property. We pull out k-element independent sets and give each a distint color until there are less than m vertices left. Then we give each point a distinct color. By this procedure

$$\begin{aligned}
\chi(G) &\leq \left\lceil \frac{n-m}{k} \right\rceil + m \leq \frac{n}{k} + m \\
&= \frac{n}{2\log_2 n}(1 + o(1)) + o\left(\frac{n}{\log_2 n}\right) \\
&= \frac{n}{2\log_2 n}(1 + o(1)),
\end{aligned}$$

and this occurs for almost all G. □

6. Central Limit Theorem

Definition 6.1. *A random variable Z is said to be* **normally distributed** *(or* **Gaussian***) with mean μ and variance σ^2 if its distribution function $F_Z(x) = \mathbb{P}\{Z \leq x\}$ is given by*

$$F_Z(x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

We will write $\mathcal{L}(Z) = N(\mu, \sigma^2)$.

Note that the density of Z is given by

$$f_Z(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

and the characteristic function by

$$\varphi_Z(t) = \mathbb{E}\, e^{itZ} = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

Definition 6.2. *We say that a sequence of random variables X_n converges weakly to a random variable X:*

$$X_n \xrightarrow{\mathrm{d}} X$$

if we have

$$\mathbb{E}\, h(X_n) \to \mathbb{E}\, h(X)$$

for all continuous and bounded functions $h : \mathbb{R} \to \mathbb{R}$.

It is well known that $X_n \xrightarrow{\mathrm{d}} X$ is equivalent to

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{6.1}$$

for all points of continuity of $F_X(x)$. If X is a continuous random variable, that is, $F_X(x)$ is continuous, then convergence in (6.1) is uniform. This means that

$$\|F_{X_n} - F_X\|_\infty = \sup_{x\in\mathbb{R}} |F_{X_n}(x) - F_X(x)| \to 0.$$

Another criterion is that

$$\mathbb{E}\, e^{itX_n} \to \mathbb{E}\, e^{itX} \tag{6.2}$$

for all $t \in \mathbb{R}$ (Levy's criterion). This criterion is even more powerful. Suppose that X_n is a sequence of random variables such that for all $t \in \mathbb{R}$ the limit

$$\psi(t) := \lim_{n\to\infty} \mathbb{E}\, e^{itX_n}$$

exists and $\psi(t)$ is a function that is continous at $t = 0$. Then $\psi(t)$ is the characteristic function of a random variable X for which we have $X_n \xrightarrow{\mathrm{d}} X$.

With help of these preliminaries we formulate the most easy variant of a central limit theorem.

Theorem 6.3. *Suppose that $Y_1, Y_2, \ldots$ are iid*[3] *random variables with finite second moment $\mathbb{E}\, Y_i^2 < \infty$. Then $S_n = Y_1 + Y_2 + \ldots + Y_n$ satisfies a central limit theorem, that is,*

$$\tilde{S}_n = \frac{S_n - \mathbb{E}\, S_n}{\sqrt{\mathbb{V} S_n}} \xrightarrow{\mathrm{d}} N(0,1). \tag{6.3}$$

[3]The abbreviation *iid* denotes *independently and identically distributed.*

Proof. The proof usues Levy's criterion (6.2). Set $\mu = \mathbb{E}\, Y_i$, $\sigma^2 = \mathbb{V} Y_i = \mathbb{E}\,(Y_i^2) - (\mathbb{E}\, Y_i)^2$ and note that $\mathbb{E}\, S_n = n\mu$ and $\mathbb{V} S_n = n\sigma^2$.

By definition we have

$$\varphi_{Y_i}(t) = \mathbb{E}\, e^{itY_i} = e^{it\mu - \frac{1}{2}\sigma^2 t^2\,(1+o(1))}$$

as $t \to 0$. Hence, we get, as $n \to \infty$,

$$\begin{aligned}
\varphi_{\tilde{S}_n}(t) &= \mathbb{E}\, e^{it\tilde{S}_n} \\
&= e^{-it\sqrt{n}\mu/\sigma}\mathbb{E}\, e^{(it/(\sqrt{n}\sigma))(Y_1+\cdots+Y_n)} \\
&= e^{-it\sqrt{n}\mu/\sigma}\left(\mathbb{E}\, e^{(it/(\sqrt{n}\sigma))Y_1}\right)^n \\
&= e^{-it\sqrt{n}\mu/\sigma} e^{it\sqrt{n}\mu/\sigma - \frac{1}{2}t^2\,(1+o(1))} \\
&= e^{-\frac{1}{2}t^2\,(1+o(1))} \to e^{-\frac{1}{2}t^2}.
\end{aligned}$$

Thus, by Levy's criterion we have proved (6.3). (Note that the $o(1)$-term is first used for $t \to 0$ and then for $n \to \infty$ since we use apply the first limit relation for $t/\sqrt{n}$.) $\square$

We can be much more precise. For example, if the third moments $\mathbb{E}\,|Y_i|^3$ are bounded then also get a uniform approximation of the form

$$\mathbb{P}\{S_n \le n\mu + x\sqrt{n}\,\sigma\} = \Phi(x) + O\left(\frac{\mathbb{E}\,|Y_i - \mu|^3}{\sigma^3\sqrt{n}}\right). \tag{6.4}$$

7. Stein's Method

In contract to the preceding method that makes use of Levi's theorem, Stein's method is based on a completely different approach.

Lemma 7.1. *A random variable Z is normally distributed with mean μ and variance σ^2 (that is, $\mathcal{L}(Z) = N(\mu, \sigma^2)^4$) if and only if*

$$\mathbb{E}\,(Z - \mu) f(Z) = \sigma^2 \mathbb{E}\, f'(Z) \tag{7.1}$$

holds for all smooth functions f with $f(x)e^{-\frac{1}{2}x^2} \to 0$ as $|x| \to \infty$ and finite integral $\int_{-\infty}^{\infty} |xf(x)|e^{-\frac{1}{2}x^2}\,dx$.

Proof. We just give a proof for $\mu = 0$ and $\sigma^2 = 1$.

First suppose that $\mathcal{L}(Z) = N(0,1)$. Hence,

$$\begin{aligned}
\mathbb{E}\, f'(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(x) e^{-\frac{1}{2}x^2}\,dx \\
&= \frac{1}{\sqrt{2\pi}} f(x) e^{-\frac{1}{2}x^2}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x f(x) e^{-\frac{1}{2}x^2}\,dx \\
&= 0 + \mathbb{E}\, Z f(Z).
\end{aligned}$$

[4]We denote by $\mathcal{L}(X)$ the law of a random variable X, that is, the probability measure on $\mathbb{R}$ that is induced by X.

Conversely, suppose that (7.1) holds. Then for every bounded (and absolutely integrable) function $g(x)$ with $\int_{-\infty}^{\infty} g(x)e^{-\frac{1}{2}x^2}\,dx = 0$ there exists $f(x)$ with

$$f'(x) - xf(x) = g(x)$$

that satisfies $f(x)e^{-\frac{1}{2}x^2} \to 0$ as $|x| \to \infty$ and $\int_{-\infty}^{\infty} |xf(x)|e^{-\frac{1}{2}x^2}\,dx < \infty$, too. We just have to set

$$f(x) = e^{\frac{1}{2}x^2} \int_{-\infty}^{x} g(y)e^{-\frac{1}{2}y^2}\,dy$$

or equivalently

$$f(x) = -e^{\frac{1}{2}x^2} \int_{x}^{\infty} g(y)e^{-\frac{1}{2}y^2}\,dy$$

Thus, if we use

$$g(x) = \mathbb{I}_{[x\le x_0]} - \Phi(x_0)$$

then

$$0 = \mathbb{E}\, f'(Z) - \mathbb{E}\, Zf(Z) = \mathbb{P}\{Z \le x_0\} - \Phi(x_0)$$

which says that $\mathcal{L}(Z) = N(0,1)$. □

For every bounded absolutely integrable function h we set

$$Nh = \mathbb{E}\, h(Z/\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(x/\sigma)e^{-\frac{1}{2}x^2}\,dx.$$

Lemma 7.2. *For every bounded function h with bounded derivative there exists a function f with bouded second derivative with*

$$\sigma^2 f'(w) - wf(w) = h(w/\sigma) - Nh. \tag{7.2}$$

f is explicitly given by

$$f(x) = e^{\frac{1}{2}x^2} \int_{-\infty}^{x} \left(h(y/\sigma) - Nh\right) e^{-\frac{1}{2}y^2}\,dy \tag{7.3}$$

and satisfies

$$\|f''\|_\infty \le K_{\mathrm{univ}} \cdot (\|h\|_\infty + \|h'\|_\infty) \tag{7.4}$$

for a universal constant $K_{\mathrm{univ}} > 0$.

The equation (7.2) is also called **Stein's equation**.

Proof. Formula (7.3) is obvious, compare to the proof of the preceding Lemma 7.1. The non-trivial part is the proof of (7.4).

Again we restrict ourselves to the case $\sigma^2 = 1$. We set $\overline{h}(x) = h(x) - Nh$. (Observe that $N\overline{h} = 0$ and that $\|\overline{h}\|_\infty \le 2\|h\|$.) We use the abbreviations

$$H_0 = \|\overline{h}\|_\infty, \quad H_1 = \|\overline{h}'\|_\infty = \|h'\|_\infty$$

and

$$F_0 = \|f\|_\infty, \quad F_1\|f'\|_\infty, \quad F_{11} = \|(xf)'\|_\infty, \quad F_2 = \|f''\|_\infty.$$

Further we set

$$c_1 = \sup_{x\geq 0} \left| x\left(1 - xe^{\frac{1}{2}x^2}\int_x^\infty e^{-\frac{1}{2}u^2}\,du\right)\right|,$$
$$c_2 = \sup_{x\geq 0} e^{\frac{1}{2}x^2}\int_x^\infty e^{-\frac{1}{2}u^2}\,du,$$
$$c_3 = \sup_{x\geq 0} x\,e^{\frac{1}{2}x^2}\int_x^\infty e^{-\frac{1}{2}u^2}\,du.$$

(The values of these constants is of no importance, however, note that $c_3 = 1$.)

The essence of the proof is to show relations between H_0, H_1 and F_0, F_1, F_{11}, F_2. In particular we show that

(1) $F_0 \leq c_2 H_0$,
(2) $F_1 \leq 2H_0$,
(3) $F_{11} \leq (c_1 + c_2)H_0 + H_1$,
(4) $F_2 \leq (c_1 + c_2)H_0 + 2H_1$.

Of course, (4) implies (7.4) and proves the lemma.

First, (1) follows directly from (7.3). We just note that we also have

$$f(x) = -e^{\frac{1}{2}x^2}\int_x^\infty \left(\overline{h}(y)\right) e^{-\frac{1}{2}y^2}\,dy.$$

For the proof of (2) we just have to use the relation $f'(x) = xf(x) + \overline{h}(x)$ which gives $F_1 \leq \|xf(x)\|_\infty + H_0$. Further (7.3) directly implies $\|xf(x)\|_\infty \leq c_3 H_0 = H_0$ and consequently (2).

Next observe that $(xf(x))' = f(x) + x^2 f(x) + x\overline{h}(x)$. We already know that $F_0 \leq c_2 H_0$. Thus, we can concentrate on $x^2 f(x) + x\overline{h}(x)$. Here we use the fact that

$$\begin{aligned}
x^2 f(x) + x\overline{h}(x) &= -x^2 e^{\frac{1}{2}x^2}\int_x^\infty \overline{h}(y)e^{-\frac{1}{2}y^2}\,dy + x\overline{h}(x)\\
&= -x^2 e^{\frac{1}{2}x^2}\int_x^\infty \overline{h}(y)e^{-\frac{1}{2}y^2}\,dy + xe^{\frac{1}{2}x^2}\int_x^\infty y\overline{h}(y)e^{-\frac{1}{2}y^2}\,dy\\
&- xe^{\frac{1}{2}x^2}\int_x^\infty y\overline{h}(y)e^{-\frac{1}{2}y^2}\,dy + xh(x)\\
&= x^2 e^{\frac{1}{2}x^2}\int_x^\infty \overline{h}(y)\left(\frac{y}{x} - 1\right)e^{-\frac{1}{2}y^2}\,dy\\
&- xe^{\frac{1}{2}x^2}\int_x^\infty \overline{h}'(y)e^{-\frac{1}{2}y^2}\,dy.
\end{aligned}$$

From this representation we obtain

$$\|x^2 f(x) + x\overline{h}(x)\|_\infty \leq c_1 H_0 + H_1$$

and, thus, (3).

Finally, by applying (3) we have

$$\begin{aligned}|f'(x+t)-f'(x)| &= |\overline{h}(x+t)-\overline{h}(x)+(x+t)f(x+t)-xf(x)| \\ &\le |t|H_1+|t|F_{11} \\ &\le |t|\left((c_1+c_2)H_0+2H_1\right).\end{aligned}$$

which implies $F_2 \le (c_1+c_2)H_0+2H_1$ as proposed. $\square$

In what follows we will use the following *norm* $\|h\|$ of a function h that is defined by

$$\|h\| := K_{\mathrm{univ}} \cdot \left(\|h\|_\infty + \|h'\|_\infty\right).$$

Further, if P and Q are two probability measures then we introduce the distance

$$d_1(P,Q) := \sup_{\|h\|\le 1} |\mathbb{E}\, h(X) - \mathbb{E}\, h(Y)|$$

in which X and Y are random variables with $\mathcal{L}(X)=P$ and $\mathcal{L}(Y)=Q$. This norm is maybe a little unusual but it perfectly fits to Stein's method.

We note that $d_1(\mathcal{L}(X_n),\mathcal{L}(X)) \to 0$ is equivalent to weak convergence $X_n \xrightarrow{\mathrm{d}} X$. There are also inequalities between this norm and several other kinds of norms but we do not stress this problem here.

We now formulate the general situation. Suppose that a random variable W can be composed in the following way: There is a finite (index) set I and for every $i \in I$ there is $K_i \subseteq I$, a subset of I with $i \in K_i$. Further there are $X_i, W_i, Z_i, Z_{ik}, W_{ik}, V_{ik}$ square integrable random variables for $i \in I$ and $k \in K_i$ with the following conditions:

(1) $W = \sum_{i\in I} X_i$,
(2) $\mathbb{E}\, X_i = 0$ for $i \in I$,
(3) $\mathbb{V} W = 1$,
(4) $W = W_i + Z_i$ for all $i \in I$ and W_i is independent of X_i,
(5) $Z_i = \sum_{k\in K_i} Z_{ik}$ for $i \in I$,
(6) $W_i = W_{ik} + V_{ik}$ for $i \in I$ and $k \in K_i$,
(7) W_{ik} is independent of the pair (X_i, Z_{ik}) for $i \in I$ and $k \in K_i$.

Theorem 7.3. *Suppose that a random variable W decomposes as introduced above. Then*

$$\begin{aligned}d_1\left(\mathcal{L}(W), N(0,1)\right) &\le \frac{1}{2}\sum_{i\in I} \mathbb{E}\, \left(|X_i| Z_i^2\right) \\ &\quad + \sum_{i\in I}\sum_{k\in K_i}\Big(\mathbb{E}\,|X_i Z_{ik} V_{ik}| + \mathbb{E}\,|X_i Z_{ik}|\cdot \mathbb{E}\,|Z_i+V_{ik}|\Big).\end{aligned} \tag{7.5}$$

Proof. The core of the proof is to show that

$$
\begin{aligned}
&|\mathbb{E}\, W f(W) - \mathbb{E}\, f'(W)| \\
&\le \|f''\|_\infty \cdot \Bigg(\frac{1}{2} \sum_{i\in I} \mathbb{E}\, \left(|X_i| Z_i^2\right) \\
&\qquad + \sum_{i\in I} \sum_{k\in K_i} \Big(\mathbb{E}\, |X_i Z_{ik} V_{ik}| + \mathbb{E}\, |X_i Z_{ik}| \cdot \mathbb{E}\, |Z_i + V_{ik}| \Big) \Bigg).
\end{aligned} \tag{7.6}
$$

Now, if h is a smooth bounded function with bounded derivative then we can find $f(x)$ with $f'(x) - x f(x) = h(x) - Nh = \overline{h}(x)$. Hence, if $\mathcal{L}(Z) = N(0,1)$ then

$$\mathbb{E}\, h(W) - \mathbb{E}\, h(Z) = \mathbb{E}\, f'(W) - \mathbb{E}\, W f(W)$$

and (7.5) follows from Lemma 7.2.

The proof of (7.6) relies on Taylor's expansion. First write

$$
\begin{aligned}
\mathbb{E}\, W f(W) - \mathbb{E}\, f'(W) &= \mathbb{E}\, W f(W) - \sum_{i\in I} \mathbb{E}\, (X_i Z_i f'(W_i)) \\
&\quad + \sum_{i\in I} \mathbb{E}\, (X_i Z_i f'(W_i) - \sum_{i\in I} \sum_{k\in K_i} \mathbb{E}\, (X_i Z_{ik})\, \mathbb{E}\, f'(W_{ik}) \\
&\quad + \sum_{i\in I} \sum_{k\in K_i} \mathbb{E}\, (X_i Z_{ik}) \left(\mathbb{E}\, f'(W_{ik}) - \mathbb{E}\, f'(W) \right),
\end{aligned} \tag{7.7}
$$

which is possible since

$$
\begin{aligned}
1 = \mathbb{E}\, W^2 &= \sum_{i\in I} \mathbb{E}\, (X_i W) \\
&= \sum_{i\in I} \mathbb{E}\, (X_i) \mathbb{E}\, (W_i) + \sum_{i\in I} \mathbb{E}\, (X_i Z_i) \\
&= \sum_{i\in I} \mathbb{E}\, (X_i Z_i) \\
&= \sum_{i\in I} \sum_{k\in K_i} \mathbb{E}\, (X_i Z_{ik}).
\end{aligned}
$$

First we have

$$
\begin{aligned}
W f(W) &= \sum_{i\in I} X_i f(W) \\
&= \sum_{i\in I} X_i \left(f(W_i) + Z_i f'(W_i) + \frac{1}{2} Z_i^2 f''(W_i + \theta_i Z_i) \right)
\end{aligned}
$$

for some $\theta_i \in [0,1]$. Since X_i and W_i are independent we have $\mathbb{E}\, (X_i f(W_i)) = \mathbb{E}\, X_i \cdot \mathbb{E}\, f(W_i)$ and consequently

$$\left| \mathbb{E}\, W f(W) - \sum_{i\in I} \mathbb{E}\, (X_i Z_i f'(W_i)) \right| \le \frac{\|f''\|}{2} \cdot \sum_{i\in I} \mathbb{E}\, (|X_i| Z_i^2). \tag{7.8}$$

Moreover,

$$\begin{aligned} X_i Z_i f'(W_i) &= \sum_{k\in K_i} X_i Z_{ik} f'(W_i) \\ &= \sum_{k\in K_i} X_i Z_{ik} \left(f'(W_{ik} + V_{ik} f''(W_{ik} + \theta_{ik} V_{ik}) \right) \end{aligned}$$

and, thus,

$$\begin{aligned} &\left| \sum_{i\in I} \mathbb{E}\,(X_i Z_i f'(W_i) - \sum_{i\in I} \sum_{k\in K_i} \mathbb{E}\,(X_i Z_{ik})\, \mathbb{E}\, f'(W_{ik}) \right| \\ &\qquad \le \|f''\| \cdot \sum_{i\in I} \sum_{k\in K_i} \mathbb{E}\,|X_i Z_{ik} V_{ik}| \end{aligned} \tag{7.9}$$

Finally, we use the decomposition $W_{ik} = W_i - V_{ik} = W - Z_i - V_{ik}$ so that

$$f'(W_{ik}) = f'(W) - (Z_i + V_{ik}) f''(W - \theta(Z_i + V_{ik}))$$

which implies

$$|\mathbb{E}\, f'(W_{ik}) - \mathbb{E}\, f'(W)| \le \|f''\| \cdot \mathbb{E}\,|(Z_i + V_{ik}|. \tag{7.10}$$

Putting the three estimates (7.8), (7.9), and (7.10) together and inserting them into the decomposition (7.7) we directly obtain (7.6). This completes the proof of the theorem. □

The assumptions of the theorem look a little bit confusing and notationally overloaded at a first sight. Therefore we will discuss it (and several modifications) in detail.

Suppose that we have iid random variables $Y_1, Y_2, \ldots$ with finite third moment. As above we set $\mu = \mathbb{E}\, Y_i$ and $\sigma^2 = \mathbb{V} Y_i$. If we set

$$X_i = \frac{Y_i - \mu}{\sqrt{n}\, \sigma}$$

then $W = X_1 + \cdots + X_n$ is exactly

$$W = \frac{Y_1 + \cdots + Y_n - \mu n}{\sqrt{n}\, \sigma}.$$

Note that the X_i are iid, too. Further we can use the above decomposition with $K_i = \{i\}$ and

$$\begin{aligned} Z_i &= X_i, \\ W_{ik} &= X_k, \\ V_{ik} &= 0. \end{aligned}$$

Then all assumptions are satisfied and we get the bound

$$d_1\left(\mathcal{L}(W), N(0,1)\right) \le \frac{1}{\sigma^3 \sqrt{n}} \left(\frac{1}{2} \mathbb{E}\, \left(|Y_i - \mu|^3 \right) + \mathbb{E}\, |Y_i - \mu| \right).$$

This estimate is of the same form as (6.4).

Next we make a slight simplification. We call the decomposition of W **dissociated** if we have $Z_{ik} = X_k$ and (also a simple representation of V_{ik}). More precisely, this means that there are square integrabel random variables $X_i, W_i, Z_i, W_{ik}, V_{ik}$ for $i \in I$ and $k \in K_i$ with the following conditions:

(1) $W = \sum_{i \in I} X_i$,
(2) $\mathbb{E}\, X_i = 0$ for $i \in I$,
(3) $\mathbb{V}W = 1$,
(4) $W = W_i + Z_i$ with $Z_i = \sum_{k \in K_i} X_k$ for $i \in I$,
(5) $W_i = W - \sum_{k \in K_i} X_k$ is independent of X_i for $i \in I$,
(6) $W_i = W_{ik} + V_{ik}$ with $V_{ik} = \sum_{j \in K_k \setminus K_i} X_j$ for $i \in I$ and $k \in K_i$,
(7) $W_{ik} = W - \sum_{j \in K_i \cup K_k} X_j$ is independent of (X_i, X_k) for $i \in I$ and $k \in K_i$.

In this case, Theorem 7.3 simplifies in the following way.

Theorem 7.4. *Suppose that a random variable W decomposes in an dissociated way with the property that $k \in K_i$ iff $k \in K_i$. Then*

$$d_1\left(\mathcal{L}(W), N(0,1)\right) \leq 2 \sum_{i \in I} \sum_{j,k \in K_i} \Bigg(\mathbb{E}\,\left(|X_i X_j X_k|\right) + \mathbb{E}\,\left(|X_i X_j|\right) \mathbb{E}\,|X_k| \Bigg). \tag{7.11}$$

Sketch of the Proof. One just has to show that

$$\begin{aligned}\frac{1}{2} \sum_{i \in I} \mathbb{E}\,\left(|X_i| Z_i^2\right) &+ \sum_{i \in I} \sum_{k \in K_i} \Bigg(\mathbb{E}\,|X_i Z_{ik} V_{ik}| + \mathbb{E}\,|X_i Z_{ik}| \cdot \mathbb{E}\,|Z_i + V_{ik}| \Bigg) \\ &\leq 2 \sum_{i \in I} \sum_{j,k \in K_i} \Bigg(\mathbb{E}\,\left(|X_i X_j X_k|\right) + \mathbb{E}\,\left(|X_i X_j|\right) \mathbb{E}\,|X_k| \Bigg),\end{aligned}$$

where $Z_{ik} = X_k$, $Z_i = \sum_{k \in K_i} X_k$, and $V_{ik} = \sum_{j \in K_k \setminus K_i} X_j$. We leave the details to the reader. $\square$

The situation gets even more transparent if we use the notion of a **dependency graph**. Suppose that we have an index set I and as system of random variables X_i for $i \in I$. A dependency graph L for this system is a graph with vertex set I that has the property that whenever A, B are subsets of I that are not inter connected by an edge then two subsystems $(X_i : i \in A)$ and $(X_j : j \in B)$ are independent.

For every $i \in I$ let K_i consists of i and of the neighbors of i in L. Then the above decomposition (with $Z_i = \sum_{k \in K_i} X_k$, and $V_{ik} = \sum_{j \in K_k \setminus K_i} X_j$) is dissociated. Thus, we can directly apply Theorem 7.4 if we have a proper dependency graph.

8. Application to Random Graphs

We will finally apply Stein's method to an example on random graphs. Let

$$I = \{i = (i_1, i_2, i_3) : 1 \le i_1 < i_2 < i_3 \le n\}$$

be the set of all subsets of $\{1, 2, \ldots, n\}$ of size 3. and et $\mathcal{T}$ denote the set of triangles in $G(n, p)$. Then

$$X = \sum_{i\in I} \mathbb{I}_{[i=(i_1,i_2,i_3)\in\mathcal{T}]}$$

counts the number of triangles. We already know that

$$\mathbb{E}\, X = \binom{n}{3} p^3$$

and

$$\sigma^2 := \mathbb{V} X = \binom{n}{3} p^3(1 - p^3) + 12\binom{n}{4} p^5(1 - p).$$

We are only interested in the case where $p \le \frac{1}{2}$ and $\mathbb{E}\, X \to \infty$, that is,

$$np \to \infty \quad \text{and} \quad 0 \le p \le \frac{1}{2}.$$

In this case we also have $\sigma^2 = \mathbb{V} X \to \infty$ We then set

$$X_i := \frac{1}{\sigma} \left(\mathbb{I}_{[i=(i_1,i_2,i_3)\in\mathcal{T}]} - p^3\right)$$

and

$$W = \sum_{i\in I} X_i = \frac{X - \mathbb{E}\, X}{\sqrt{\mathbb{V} X}}.$$

It is now possible to define a dependency graph L. The vertex set is (of course) $V(L) = I$, and the edge set is given by

$$E(L) = \{(i, j) : |\{i_1, i_2, i_3\} \cap \{j_1, j_2, j_3\}| \ge 2\}.$$

It is immediately clear that L satisfies the proposed properties of a dependency graph.

Now for every $i = (i_1, i_2, i_3) \in I$ we define

$$K_i = \{k = (k_1, k_2, k_3) \in I : |\{i_1, i_2, i_3\} \cap \{k_1, k_2, k_3\}| \ge 2\}.$$

Note that $|K_i| = 1 + 3(n - 3)$ for all $i \in I$. In particular we can apply Theorem 7.4.

For this purpose we have to estimate

$$\sum_{i\in I} \sum_{j,k\in K_i} \Big(\mathbb{E}\, (|X_i X_j X_k|) + \mathbb{E}\, (|X_i X_j|)\, \mathbb{E}\, |X_k|\Big).$$

For example, if $i = j = k$ then we have

$$\mathbb{E}\, (|X_i X_j X_k|) = \mathbb{E}\, (|X_i|^3) = \frac{1}{\sigma^3} \left(p^3(1 - p^3)^3 + (1 - p^3)p^9\right) \le \frac{2p^3}{\sigma^3}$$

which contributes to

$$\frac{1}{\sigma^3}\binom{n}{3}p^3 = O\left(\frac{n^3p^3}{\sigma^3}\right).$$

In a similar way we can deal with the other cases and end up with an upper bound of the form

$$d_1\left(\mathcal{L}(W), N(0,1)\right) = O\left(\frac{n^3p^3(1+np^2)^2}{n^{9/2}p^{9/2}(1+np^2)^{3/2}}\right) = O\left((np)^{-3/2}(1+np)^{1/2}\right).$$

Thus, we have proved the following theorem.

Theorem 8.1. *Suppose that $0 < p \le \frac{1}{2}$ and $np \to \infty$. Then the number of triangles in a random graph $G(n,p)$ satisfies a central limit theorem.*

References

[1] Noga Alon and Joel H. Spencer. *The probabilistic method.* Second edition. With an appendix on the life and work of Paul Erdős. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000

[2] Andrew D. Barbour and G. K. Eagleson, Multiple comparisons and sums of dissociated random variables. *Adv. in Appl. Probab.* 17 (1985), no. 1, 147–162.

[3] Andrew D. Barbour, Lars Holst, and Svante Janson, Svante *Poisson approximation*, Oxford Studies in Probability, 2. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1992.

[4] Andrew D. Barbour, MichałKaroński, and Andrzej Ruciński, A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* 47 (1989), no. 2, 125–145.

[5] Béla Bollobás, *Random graphs.* Second edition. Cambridge Studies in Advanced Mathematics, 73. Cambridge University Press, Cambridge, 2001.

[6] Paul Erdős, Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.* 53, (1947). 292–294

[7] Paul Erdős, Graph theory and probability. *Canad. J. Math.* 11 1959 34–38.

[8] Paul Erdős and A. Rényi, A. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 5 1960 17–61.

[9] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000

[10] Valentin F. Kolchin, *Random graphs.* Encyclopedia of Mathematics and its Applications, 53. Cambridge University Press, Cambridge, 1999.

[11] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables.* Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 583–602. Univ. California Press, Berkeley, Calif., 1972.

[12] Charles Stein, *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

Michael Drmota, Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstrasse 8–10/104, A-1040 Wien, Austria

E-mail address: `michael.drmota@tuwien.ac.at`