The Probabilistic Method

Second Edition

Noga Alon

Joel H. Spencer

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

Contents

Part I

METHODS

This page intentionally left blank

1

The Basic Method

What you need is that your brain is open.
– Paul Erdős

1.1 THE PROBABILISTIC METHOD

The probabilistic method is a powerful tool for tackling many problems in discrete mathematics. Roughly speaking, the method works as follows: Trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this space with positive probability. The method is best illustrated by examples. Here is a simple one. The *Ramsey number* $R(k, \ell)$ is the smallest integer n such that in any two-coloring of the edges of a complete graph on n vertices K_n by red and blue, either there is a red K_k (i.e., a complete subgraph on k vertices all of whose edges are colored red) or there is a blue K_ℓ. Ramsey (1929) showed that $R(k, \ell)$ is finite for any two integers k and ℓ. Let us obtain a lower bound for the diagonal Ramsey numbers $R(k, k)$.

Proposition 1.1.1 *If* $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$ *then* $R(k, k) > n$. *Thus* $R(k, k) > \lfloor 2^{k/2} \rfloor$ *for all* $k \geq 3$.

Proof. Consider a random two-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equally likely. For any fixed set R of k vertices, let A_R be the event that the induced subgraph of K_n on R is *monochromatic* (i.e., that either all its edges are red or they are all blue). Clearly,

$\Pr(A_R) = 2^{1-\binom{k}{2}}$. Since there are $\binom{n}{k}$ possible choices for R, the probability that at least one of the events A_R occurs is at most $\binom{n}{k}2^{1-\binom{k}{2}} < 1$. Thus, with positive probability, no event A_R occurs and there is a two-coloring of K_n without a monochromatic K_k, i.e., $R(k,k) > n$. Note that if $k \geq 3$ and we take $n = \lfloor 2^{k/2} \rfloor$ then $\binom{n}{k}2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{k^2/2}} < 1$ and hence $R(k,k) > \lfloor 2^{k/2} \rfloor$ for all $k \geq 3$. ■

This simple example demonstrates the essence of the probabilistic method. To prove the existence of a good coloring we do not present one explicitly, but rather show, in a nonconstructive way, that it exists. This example appeared in a paper of P. Erdős from 1947. Although Szele had applied the probabilistic method to another combinatorial problem, mentioned in Chapter 2, already in 1943, Erdős was certainly the first one who understood the full power of this method and applied it successfully over the years to numerous problems. One can, of course, claim that the probability is not essential in the proof given above. An equally simple proof can be described by counting; we just check that the total number of two-colorings of K_n is bigger than the number of those containing a monochromatic K_k.

Moreover, since the vast majority of the probability spaces considered in the study of combinatorial problems are finite spaces, this claim applies to most of the applications of the probabilistic method in discrete mathematics. Theoretically, this is, indeed, the case. However, in practice, the probability is essential. It would be hopeless to replace the applications of many of the tools appearing in this book, including, e.g., the second moment method, the Lovász Local Lemma and the concentration via martingales by counting arguments, even when these are applied to finite probability spaces.

The probabilistic method has an interesting algorithmic aspect. Consider, for example, the proof of Proposition 1.1.1 that shows that there is an edge two-coloring of K_n without a monochromatic $K_{2\log_2 n}$. Can we actually find such a coloring? This question, as asked, may sound ridiculous; the total number of possible colorings is finite, so we can try them all until we find the desired one. However, such a procedure may require $2^{\binom{n}{2}}$ steps; an amount of time which is exponential in the size $[= \binom{n}{2}]$ of the problem. Algorithms whose running time is more than polynomial in the size of the problem are usually considered impractical. The class of problems that can be solved in polynomial time, usually denoted by **P** [see, e.g., Aho, Hopcroft and Ullman (1974)], is, in a sense, the class of all solvable problems. In this sense, the exhaustive search approach suggested above for finding a good coloring of K_n is not acceptable, and this is the reason for our remark that the proof of Proposition 1.1.1 is nonconstructive; it does not supply a constructive, efficient and deterministic way of producing a coloring with the desired properties. However, a closer look at the proof shows that, in fact, it can be used to produce, effectively, a coloring which is very likely to be good. This is because for large k, if $n = \lfloor 2^{k/2} \rfloor$ then $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!}\left(\frac{n}{2^{k/2}}\right)^k \leq \frac{2^{1+\frac{k}{2}}}{k!} \ll 1$. Hence, a random coloring of K_n is very likely not to contain a monochromatic $K_{2\log n}$. This means that if, for some reason, we *must* present a two-coloring of the edges of K_{1024} without a monochromatic K_{20} we can simply produce a random two-coloring by flipping a fair coin $\binom{1024}{2}$

times. We can then deliver the resulting coloring safely; the probability that it contains a monochromatic K_{20} is less than $\frac{2^{11}}{20!}$, probably much smaller than our chances of making a mistake in any rigorous proof that a certain coloring is good! Therefore, in some cases the probabilistic, nonconstructive method does supply effective probabilistic algorithms. Moreover, these algorithms can sometimes be converted into deterministic ones. This topic is discussed in some detail in Chapter 15.

The probabilistic method is a powerful tool in Combinatorics and in Graph Theory. It is also extremely useful in Number Theory and in Combinatorial Geometry. More recently it has been applied in the development of efficient algorithmic techniques and in the study of various computational problems. In the rest of this chapter we present several simple examples that demonstrate some of the broad spectrum of topics in which this method is helpful. More complicated examples, involving various more delicate probabilistic arguments, appear in the rest of the book.

1.2 GRAPH THEORY

A *tournament* on a set V of n players is an orientation $T = (V, E)$ of the edges of the complete graph on the set of vertices V. Thus, for every two distinct elements x and y of V either (x, y) or (y, x) is in E, but not both. The name "tournament" is natural, since one can think of the set V as a set of players in which each pair participates in a single match, where (x, y) is in the tournament iff x beats y. We say that T has the property S_k if for every set of k players there is one who beats them all. For example, a directed triangle $T_3 = (V, E)$, where $V = \{1, 2, 3\}$ and $E = \{(1, 2), (2, 3), (3, 1)\}$, has S_1. Is it true that for every finite k there is a tournament T (on more than k vertices) with the property S_k? As shown by Erdős (1963b), this problem, raised by Schütte, can be solved almost trivially by applying probabilistic arguments. Moreover, these arguments even supply a rather sharp estimate for the minimum possible number of vertices in such a tournament. The basic (and natural) idea is that if n is sufficiently large as a function of k, then a *random* tournament on the set $V = \{1, \ldots, n\}$ of n players is very likely to have property S_k. By a random tournament we mean here a tournament T on V obtained by choosing, for each $1 \leq i < j \leq n$, independently, either the edge (i, j) or the edge (j, i), where each of these two choices is equally likely. Observe that in this manner, all the $2^{\binom{n}{2}}$ possible tournaments on V are equally likely, i.e., the probability space considered is symmetric. It is worth noting that we often use in applications symmetric probability spaces. In these cases, we shall sometimes refer to an element of the space as a *random element*, without describing explicitly the probability distribution . Thus, for example, in the proof of Proposition 1.1.1 random two-colorings of K_n were considered, i.e., all possible colorings were equally likely. Similarly, in the proof of the next simple result we study random tournaments on V.

Theorem 1.2.1 *If* $\binom{n}{k}(1-2^{-k})^{n-k} < 1$ *then there is a tournament on* n *vertices that has the property* S_k.

Proof. Consider a random tournament on the set $V = \{1, \ldots, n\}$. For every fixed subset K of size k of V, let A_K be the event that there is no vertex which beats all the members of K. Clearly $\Pr(A_K) = (1-2^{-k})^{n-k}$. This is because for each fixed vertex $v \in V - K$, the probability that v does not beat all the members of K is $1-2^{-k}$, and all these $n-k$ events corresponding to the various possible choices of v are independent. It follows that

$$\Pr\Big(\bigvee_{\substack{K \subset V \\ |K|=k}} A_K\Big) \leq \sum_{\substack{K \subset V \\ |K|=k}} \Pr(A_K) = \binom{n}{k}(1-2^{-k})^{n-k} < 1 .$$

Therefore, with positive probability no event A_K occurs, i.e., there is a tournament on n vertices that has the property S_k. ■

Let $f(k)$ denote the minimum possible number of vertices of a tournament that has the property S_k. Since $\binom{n}{k} < \left(\frac{en}{k}\right)^k$ and $(1-2^{-k})^{n-k} < e^{-(n-k)/2^k}$, Theorem 1.2.1 implies that $f(k) \leq k^2 \cdot 2^k \cdot (\ln 2)\big(1+o(1)\big)$. It is not too difficult to check that $f(1) = 3$ and $f(2) = 7$. As proved by Szekeres [cf. Moon (1968)], $f(k) \geq c_1 \cdot k \cdot 2^k$.

Can one find an explicit construction of tournaments with at most c_2^k vertices having property S_k? Such a construction is known, but is not trivial; it is described in Chapter 9.

A *dominating set* of an undirected graph $G = (V, E)$ is a set $U \subseteq V$ such that every vertex $v \in V - U$ has at least one neighbor in U.

Theorem 1.2.2 *Let* $G = (V, E)$ *be a graph on* n *vertices, with minimum degree* $\delta > 1$. *Then* G *has a dominating set of at most* $n\frac{1+\ln(\delta+1)}{\delta+1}$ *vertices.*

Proof. Let $p \in [0,1]$ be, for the moment, arbitrary. Let us pick, randomly and independently, each vertex of V with probability p. Let X be the (random) set of all vertices picked and let $Y = Y_X$ be the random set of all vertices in $V - X$ that do not have any neighbor in X. The expected value of $|X|$ is clearly np. For each fixed vertex $v \in V$, $\Pr(v \in Y) = \Pr(v \text{ and its neighbors are not in } X) \leq (1-p)^{\delta+1}$. Since the expected value of a sum of random variables is the sum of their expectations (even if they are not independent) and since the random variable $|Y|$ can be written as a sum of n indicator random variables χ_v ($v \in V$), where $\chi_v = 1$ if $v \in Y$ and $\chi_v = 0$ otherwise, we conclude that the expected value of $|X| + |Y|$ is at most $np + n(1-p)^{\delta+1}$. Consequently, there is at least one choice of $X \subseteq V$ such that $|X| + |Y_X| \leq np + n(1-p)^{\delta+1}$. The set $U = X \cup Y_X$ is clearly a dominating set of G whose cardinality is at most this size.

The above argument works for any $p \in [0,1]$. To optimize the result we use elementary calculus. For convenience we bound $1 - p \leq e^{-p}$ (this holds for all nonnegative p and is a fairly close bound when p is small) to give the simpler bound

$$|U| \leq np + ne^{-p(\delta+1)}.$$

Take the derivative of the right-hand side with respect to p and set it equal to zero. The right-hand side is minimized at

$$p = \frac{\ln(\delta + 1)}{\delta + 1}.$$

Formally, we set p equal to this value in the first line of the proof. We now have $|U| \leq n\frac{1+\ln(\delta+1)}{\delta+1}$ as claimed. ■

Three simple but important ideas are incorporated in the last proof. The first is the linearity of expectation; many applications of this simple, yet powerful principle appear in Chapter 2. The second is, maybe, more subtle, and is an example of the "alteration" principle which is discussed in Chapter 3. The random choice did not supply the required dominating set U immediately; it only supplied the set X, which has to be altered a little (by adding to it the set Y_X) to provide the required dominating set. The third involves the optimal choice of p. One often wants to make a random choice but is not certain what probability p should be used. The idea is to carry out the proof with p as a parameter giving a result which is a function of p. At the end that p is selected which gives the optimal result. There is here yet a fourth idea that might be called asymptotic calculus. We wanted the asymptotics of $\min np + n(1-p)^{\delta+1}$ where p ranges over $[0, 1]$. The actual minimum $p = 1 - (\delta + 1)^{-1/\delta}$ is difficult to deal with and in many similar cases precise minima are impossible to find in closed form. Rather, we give away a little bit, bounding $1 - p \leq e^{-p}$, yielding a clean bound. A good part of the *art* of the probabilistic method lies in finding suboptimal but clean bounds. Did we give away too much in this case? The answer depends on the emphasis for the original question. For $\delta = 3$ our rough bound gives $|U| \leq 0.596n$ while the more precise calculation gives $|U| \leq 0.496n$, perhaps a substantial difference. For δ large both methods give asymptotically $n\frac{\ln\delta}{\delta}$.

It can be easily deduced from the results in Alon (1990b) that the bound in Theorem 1.2.2 is nearly optimal. A nonprobabilistic, algorithmic proof of this theorem can be obtained by choosing the vertices for the dominating set one by one, when in each step a vertex that covers the maximum number of yet uncovered vertices is picked. Indeed, for each vertex v denote by $C(v)$ the set consisting of v together with all its neighbours. Suppose that during the process of picking vertices the number of vertices u that do not lie in the union of the sets $C(v)$ of the vertices chosen so far is r. By the assumption, the sum of the cardinalities of the sets $C(u)$ over all such uncovered vertices u is at least $r(\delta + 1)$, and hence, by averaging, there is a vertex v that belongs to at least $r(\delta + 1)/n$ such sets $C(u)$. Adding this v to the set of chosen vertices we observe that the number of uncovered vertices is now at most $r(1 - \frac{\delta+1}{n})$. It follows that in each iteration of the above procedure the number of uncovered vertices decreases by a factor of $1 - (\delta + 1)/n$ and hence after $\frac{n}{\delta+1}\ln(\delta + 1)$ steps there will be at most $n/(\delta + 1)$ yet uncovered vertices which can now be added to the set of chosen vertices to form a dominating set of size at most equal to the one in the conclusion of Theorem 1.2.2.

Combining this with some ideas of Podderyugin and Matula, we can obtain a very efficient algorithm to decide if a given undirected graph on n vertices is, say, $\frac{n}{2}$-edge connected. A *cut* in a graph $G = (V, E)$ is a partition of the set of vertices V into

two nonempty disjoint sets $V = V_1 \cup V_2$. If $v_1 \in V_1$ and $v_2 \in V_2$ we say that the cut *separates* v_1 and v_2. The *size* of the cut is the number of edges of G having one end in V_1 and another end in V_2. In fact, we sometimes identify the cut with the set of these edges. The *edge-connectivity* of G is the minimum size of a cut of G. The following lemma is due to Podderyugin and Matula (independently).

Lemma 1.2.3 *Let $G = (V, E)$ be a graph with minimum degree δ and let $V = V_1 \cup V_2$ be a cut of size smaller than δ in G. Then every dominating set U of G has vertices in V_1 and in V_2.*

Proof. Suppose this is false and $U \subseteq V_1$. Choose, arbitrarily, a vertex $v \in V_2$ and let $v_1, v_2, \ldots, v_\delta$ be δ of its neighbors. For each i, $1 \leq i \leq \delta$, define an edge e_i of the given cut as follows; if $v_i \in V_1$ then $e_i = \{v, v_i\}$, otherwise, $v_i \in V_2$ and since U is dominating there is at least one vertex $u \in U$ such that $\{u, v_i\}$ is an edge; take such a u and put $e_i = \{u, v_i\}$. The δ edges $e_1, \ldots, e_\delta$ are all distinct and all lie in the given cut, contradicting the assumption that its size is less than δ. This completes the proof. ■

Let $G = (V, E)$ be a graph on n vertices, and suppose we wish to decide if G is $n/2$ edge-connected, i.e., if its edge connectivity is at least $n/2$. Matula showed, by applying Lemma 1.2.3, that this can be done in time $O(n^3)$. By the remark following the proof of Theorem 1.2.2, we can slightly improve it and get an $O(n^{8/3} \log n)$ algorithm as follows. We first check if the minimum degree δ of G is at least $n/2$. If not, G is not $n/2$-edge connected, and the algorithm ends. Otherwise, by Theorem 1.2.2 there is a dominating set $U = \{u_1, \ldots, u_k\}$ of G, where $k = O(\log n)$, and it can in fact be found in $O(n^2)$-time. We now find, for each i, $2 \leq i \leq k$, the minimum size s_i of a cut that separates u_1 from u_i. Each of these problems can be solved by solving a standard network flow problem in time $O(n^{8/3})$, [see, e.g., Tarjan (1983).] By Lemma 1.2.3 the edge connectivity of G is simply the minimum between δ and $\min_{2 \leq i \leq k} s_i$. The total time of the algorithm is $O(n^{8/3} \log n)$, as claimed.

1.3 COMBINATORICS

A *hypergraph* is a pair $H = (V, E)$, where V is a finite set whose elements are called *vertices* and E is a family of subsets of V, called *edges*. It is *n-uniform* if each of its edges contains precisely n vertices. We say that H has *property B*, or that it is two-*colorable* if there is a two-coloring of V such that no edge is monochromatic. Let $m(n)$ denote the minimum possible number of edges of an n-uniform hypergraph that does not have property B.

Proposition 1.3.1 [Erdős (1963a)] *Every n-uniform hypergraph with less than 2^{n-1} edges has property B. Therefore $m(n) \geq 2^{n-1}$.*

Proof. Let $H = (V, E)$ be an n-uniform hypergraph with less than 2^{n-1} edges. Color V randomly by two colors. For each edge $e \in E$, let A_e be the event that e is

monochromatic. Clearly $\Pr(A_e) = 2^{1-n}$. Therefore

$$\Pr\left(\bigvee_{e\in E} A_e\right) \leq \sum_{e\in E} \Pr(A_e) < 1$$

and there is a two-coloring without monochromatic edges. ■

In Chapter 3, Section 3.5 we present a more delicate argument, due to Radhakrishnan and Srinivasan, and based on an idea of Beck, that shows that $m(n) \geq \Omega((\frac{n}{\ln n})^{\frac{1}{2}} 2^n)$.

The best known upper bound to $m(n)$ is found by turning the probabilistic argument "on its head." Basically, the sets become random and each coloring defines an event. Fix V with v points, where we shall later optimize v. Let χ be a coloring of V with a points in one color, $b = v - a$ points in the other. Let $S \subset V$ be a uniformly selected n-set. Then

$$\Pr(S \textit{ is monochromatic under } \chi) = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}}.$$

Let us assume v is even for convenience. As $\binom{y}{n}$ is convex, this expression is minimized when $a = b$. Thus

$$\Pr(S \textit{ is monochromatic under } \chi) \geq p$$

where we set

$$p = \frac{2\binom{v/2}{n}}{\binom{v}{n}}$$

for notational convenience. Now let $S_1, \ldots, S_m$ be uniformly and independently chosen n-sets, m to be determined. For each coloring χ let A_χ be the event that none of the S_i are monochromatic. By the independence of the S_i

$$\Pr(A_\chi) \leq (1-p)^m.$$

There are 2^v colorings so

$$\Pr(\bigvee_{\chi} A_\chi) \leq 2^v (1-p)^m.$$

When this quantity is less than 1 there exist $S_1, \ldots, S_m$ so that no A_χ holds; i.e., $S_1, \ldots, S_m$ is not two-colorable and hence $m(n) \leq m$.

The asymptotics provide a fairly typical example of those encountered when employing the probabilistic method. We first use the inequality $1 - p \leq e^{-p}$. This is valid for all positive p and the terms are quite close when p is small. When

$$m = \left\lceil \frac{v \ln 2}{p} \right\rceil$$

then $2^v(1-p)^m < 2^v e^{-pm} \leq 1$ so $m(n) \leq m$. Now we need to find v to minimize v/p. We may interpret p as twice the probability of picking n white balls from

an urn with $v/2$ white and $v/2$ black balls, sampling without replacement. It is tempting to estimate p by 2^{-n+1}, the probability for sampling with replacement. This approximation would yield $m \sim v2^{n-1}(\ln 2)$. As v gets smaller, however, the approximation becomes less accurate and, as we wish to minimize m, the tradeoff becomes essential. We use a second order approximation

$$p = \frac{2\binom{v/2}{n}}{\binom{v}{n}} = 2^{1-n}\prod_{i=0}^{n-1}\frac{v-2i}{v-i} \sim 2^{1-n}e^{-n^2/2v}$$

as long as $v \gg n^{3/2}$, estimating $\frac{v-2i}{v-i} = 1 - \frac{i}{v} + O(\frac{i^2}{v^2}) = e^{-\frac{i}{v}+O(\frac{i^2}{v^2})}$. Elementary calculus gives $v = n^2/2$ for the optimal value. The evenness of v may require a change of at most 2 which turns out to be asymptotically negligible. This yields the following result of Erdős (1964).

Theorem 1.3.2

$$m(n) < (1+o(1))\frac{e\ln 2}{4}n^2 2^n.$$

Let $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^{h}$ be a family of pairs of subsets of an arbitrary set. We call $\mathcal{F}$ a (k,ℓ)-*system* if $|A_i| = k$ and $|B_i| = \ell$ for all $1 \le i \le h$, $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ for all distinct i, j with $1 \le i, j \le h$. Bollobás (1965) proved the following result, which has many interesting extensions and applications.

Theorem 1.3.3 *If* $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^{h}$ *is a* (k,ℓ)*-system then* $h \le \binom{k+\ell}{k}$.

Proof. Put $X = \bigcup_{i=1}^{h}(A_i \cup B_i)$ and consider a random order π of X. For each i, $1 \le i \le k$, let X_i be the event that all the elements of A_i precede all those of B_i in this order. Clearly $\Pr(X_i) = 1/\binom{k+\ell}{k}$. It is also easy to check that the events X_i are pairwise disjoint. Indeed, assume this is false and let π be an order in which all the elements of A_i precede those of B_i and all the elements of A_j precede those of B_j. Without loss of generality we may assume that the last element of A_i does not appear after the last element of A_j. But in this case, all elements of A_i precede all those of B_j, contradicting the fact that $A_i \cap B_j \neq \emptyset$. Therefore, all the events X_i are pairwise disjoint, as claimed. It follows that $1 \ge \Pr\left(\bigvee_{i=1}^{h} X_i\right) = \sum_{i=1}^{h}\Pr(X_i) = h \cdot 1/\binom{k+\ell}{k}$, completing the proof. ■

Theorem 1.3.3 is sharp, as shown by the family $\mathcal{F} = \{(A, X \setminus A) : A \subset X,\ |A| = k\}$, where $X = \{1, 2, \ldots, k+\ell\}$.

1.4 COMBINATORIAL NUMBER THEORY

A subset A of an abelian group G is called *sum-free* if $(A + A) \cap A = \emptyset$, i.e., if there are no $a_1, a_2, a_3 \in A$ such that $a_1 + a_2 = a_3$.

Theorem 1.4.1 [Erdős (1965a)] *Every set $B = \{b_1, \ldots, b_n\}$ of n nonzero integers contains a sum-free subset A of size $|A| > \frac{1}{3}n$.*

Proof. Let $p = 3k + 2$ be a prime, which satisfies $p > 2 \max_{1 \le i \le n} |b_i|$ and put $C = \{k+1, k+2, \ldots, 2k+1\}$. Observe that C is a sum-free subset of the cyclic group Z_p and that $\frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}$. Let us choose at random an integer x, $1 \le x < p$, according to a uniform distribution on $\{1, 2, \ldots, p-1\}$, and define $d_1, \ldots, d_n$ by $d_i \equiv xb_i (\mathrm{mod}\, p)$, $0 \le d_i < p$. Trivially, for every fixed i, $1 \le i \le n$, as x ranges over all numbers $1, 2, \ldots, p-1$, d_i ranges over all nonzero elements of Z_p and hence $\Pr(d_i \in C) = \frac{|C|}{p-1} > \frac{1}{3}$. Therefore, the expected number of elements b_i such that $d_i \in C$ is more than $\frac{n}{3}$. Consequently, there is an x, $1 \le x < p$ and a subsequence A of B of cardinality $|A| > \frac{n}{3}$, such that $xa (\mathrm{mod}\, p) \in C$ for all $a \in A$. This A is clearly sum-free, since if $a_1 + a_2 = a_3$ for some $a_1, a_2, a_3 \in A$ then $xa_1 + xa_2 \equiv xa_3 (\mathrm{mod}\, p)$, contradicting the fact that C is a sum-free subset of Z_p. This completes the proof. ■

In Alon and Kleitman (1990) it is shown that every set of n nonzero elements of an arbitrary abelian group contains a sum-free subset of more than $2n/7$ elements, and that the constant $2/7$ is best possible. The best possible constant in Theorem 1.4.1 is not known.

1.5 DISJOINT PAIRS

The probabilistic method is most striking when it is applied to prove theorems whose statement does not seem to suggest at all the need for probability. Most of the examples given in the previous sections are simple instances of such statements. In this section we describe a (slightly) more complicated result, due to Alon and Frankl (1985), which solves a conjecture of Daykin and Erdős.

Let $\mathcal{F}$ be a family of m distinct subsets of $X = \{1, 2, \ldots, n\}$. Let $d(\mathcal{F})$ denote the number of disjoint pairs in F, i.e.,

$$d(\mathcal{F}) = \left|\left\{(F, F') : F, F' \in \mathcal{F},\ F \cap F' = \emptyset\right\}\right|.$$

Daykin and Erdős conjectured that if $m = 2^{\left(\frac{1}{2}+\delta\right)n}$, then, for every fixed $\delta > 0$, $d(\mathcal{F}) = o(m^2)$, as n tends to infinity. This result follows from the following theorem, which is a special case of a more general result.

Theorem 1.5.1 *Let $\mathcal{F}$ be a family of $m = 2^{\left(\frac{1}{2}+\delta\right)n}$ subsets of $X = \{1, 2, \ldots, n\}$, where $\delta > 0$. Then*

$$d(\mathcal{F}) < m^{2 - \frac{\delta^2}{2}}. \tag{1.1}$$

Proof. Suppose (1.1) is false and pick independently t members $A_1, A_2, \ldots, A_t$ of $\mathcal{F}$ with repetitions at random, where t is a large positive integer, to be chosen later.

We will show that with positive probability $|A_1 \cup A_2 \cup \ldots \cup A_t| > n/2$ and still this union is disjoint to more than $2^{n/2}$ distinct subsets of X. This contradiction will establish (1.1).

In fact,

$$\begin{aligned}\Pr\left(|A_1 \cup A_2 \cup \ldots \cup A_t| \leq n/2\right) &\leq \textstyle\sum_{S \subset X, |S| \leq n/2} \Pr(A_i \subset S, i = 1, \ldots, t) \\ &\leq 2^n (2^{n/2}/2^{((1/2)+\delta)n})^t = 2^{n(1-\delta t)} . \end{aligned} \tag{1.2}$$

Define

$$v(B) = \left|\{A \in \mathcal{F} : B \cap A = \emptyset\}\right| .$$

Clearly,

$$\sum_{B \in \mathcal{F}} v(B) = 2d(\mathcal{F}) \geq 2m^{2-\delta^2/2} .$$

Let Y be a random variable whose value is the number of members $B \in \mathcal{F}$ which are disjoint to all the A_i $(1 \leq i \leq t)$. By the convexity of z^t the expected value of Y satisfies

$$\begin{aligned} E(Y) &= \sum_{B \in \mathcal{F}} \left(v(B)/m\right)^t = \frac{1}{m^t} \cdot m \left(\frac{\sum v(B)^t}{m}\right) \\ &\geq \frac{1}{m^t} \cdot m \left(\frac{2d(\mathcal{F})}{m}\right)^t \geq 2m^{1-t\delta^2/2} . \end{aligned} \tag{1.3}$$

Since $Y \leq m$ we conclude that

$$\Pr(Y \geq m^{1-t\delta^2/2}) \geq m^{-t\delta^2/2} . \tag{1.4}$$

One can check that for $t = \lceil 1 + 1/\delta \rceil$, $m^{1-t\delta^2/2} > 2^{n/2}$ and the right-hand side of (1.4) is greater than the right-hand side of (1.2). Thus, with positive probability, $|A_1 \cup A_2 \cup \ldots \cup A_t| > n/2$ and still this union is disjoint to more than $2^{n/2}$ members of F. This contradiction implies inequality (1.1). ■

1.6 EXERCISES

1. Prove that if there is a real p, $0 \leq p \leq 1$ such that

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{t} (1-p)^{\binom{t}{2}} < 1,$$

then the Ramsey number $r(k, t)$ satisfies $r(k, t) > n$. Using this, show that

$$r(4, t) \geq \Omega(t^{3/2}/(\ln t)^{3/2}).$$

2. Suppose $n \geq 4$ and let H be an n-uniform hypergraph with at most $\frac{4^{n-1}}{3^n}$ edges. Prove that there is a coloring of the vertices of H by four colors so that in every edge all four colors are represented.

3. (*) Prove that for every two independent, identically distributed real random variables X and Y,

$$\Pr(|X - Y| \leq 2) \leq 3 \Pr(|X - Y| \leq 1).$$

4. (*) Let $G = (V, E)$ be a graph with n vertices and minimum degree $\delta > 10$. Prove that there is a partition of V into two disjoint subsets A and B so that $|A| \leq O(\frac{n \ln \delta}{\delta})$, and each vertex of B has at least one neighbor in A and at least one neighbor in B.

5. (*) Let $G = (V, E)$ be a graph on $n \geq 10$ vertices and suppose that if we add to G any edge not in G then the number of copies of a complete graph on 10 vertices in it increases. Show that the number of edges of G is at least $8n - 36$.

6. (*) Theorem 1.2.1 asserts that for every integer $k > 0$ there is a tournament $T_k = (V, E)$ with $|V| > k$ such that for every set U of at most k vertices of T_k there is a vertex v so that all directed arcs $\{(v, u) : u \in U\}$ are in E.

   Show that each such tournament contains at least $\Omega(k2^k)$ vertices.

7. Let $\{(A_i, B_i), 1 \leq i \leq h\}$ be a family of pairs of subsets of the set of integers such that $|A_i| = k$ for all i and $|B_i| = l$ for all i, $A_i \cap B_i = \emptyset$ and $(A_i \cap B_j) \cup (A_j \cap B_i) \neq \emptyset$ for all $i \neq j$. Prove that $h \leq \frac{(k+l)^{k+l}}{k^k l^l}$.

8. (Prefix-free codes; Kraft Inequality). Let F be a finite collection of binary strings of finite lengths and assume no member of F is a prefix of another one. Let N_i denote the number of strings of length i in F. Prove that

$$\sum_i \frac{N_i}{2^i} \leq 1.$$

9. (*) (Uniquely decipherable codes; Kraft-McMillan Inequality). Let F be a finite collection of binary strings of finite lengths and assume that no two distinct concatenations of two finite sequences of codewords result in the same binary sequence. Let N_i denote the number of strings of length i in F. Prove that

$$\sum_i \frac{N_i}{2^i} \leq 1.$$

10. Prove that there is an absolute constant $c > 0$ with the following property. Let A be an n by n matrix with pairwise distinct entries. Then there is a permutation of the rows of A so that no column in the permuted matrix contains an increasing subsequence of length at least $c\sqrt{n}$.

THE PROBABILISTIC LENS:
The Erdős-Ko-Rado Theorem

A family $\mathcal{F}$ of sets is called intersecting if $A, B \in \mathcal{F}$ implies $A \cap B \neq \emptyset$. Suppose $n \geq 2k$ and let $\mathcal{F}$ be an intersecting family of k-element subsets of an n-set, for definiteness $\{0, \ldots, n-1\}$. The Erdős-Ko-Rado Theorem is that $|\mathcal{F}| \leq \binom{n-1}{k-1}$. This is achievable by taking the family of k-sets containing a particular point. We give a short proof due to Katona (1972).

Lemma 1 *For $0 \leq s \leq n-1$ set $A_s = \{s, s+1, \ldots, s+k-1\}$ where addition is modulo n. Then $\mathcal{F}$ can contain at most k of the sets A_s.*

Proof. Fix some $A_s \in \mathcal{F}$. All other sets A_t that intersect A_s can be partitioned into $k-1$ pairs $\{A_{s-i}, A_{s+k-i}\}$, $(1 \leq i \leq k-1)$, and the members of each such pair are disjoint. The result follows, since $\mathcal{F}$ can contain at most one member of each pair. ■

Now we prove the Erdős-Ko-Rado Theorem. Let a permutation σ of $\{0, \ldots, n-1\}$ and $i \in \{0, \ldots, n-1\}$ be chosen randomly, uniformly and independently and set $A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+k-1)\}$, addition again modulo n. Conditioning on any choice of σ the Lemma gives $\Pr[A \in \mathcal{F}] \leq k/n$. Hence $\Pr[A \in \mathcal{F}] \leq k/n$. But A is uniformly chosen from all k-sets so

$$\frac{k}{n} \geq \Pr[A \in \mathcal{F}] = \frac{|\mathcal{F}|}{\binom{n}{k}}$$

and

$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}.$$

■

2

Linearity of Expectation

The search for truth is more precious than its possession.
– Albert Einstein

2.1 BASICS

Let $X_1, \ldots, X_n$ be random variables, $X = c_1 X_1 + \ldots + c_n X_n$. Linearity of Expectation states that

$$E[X] = c_1 E[X_1] + \ldots + c_n E[X_n].$$

The power of this principle comes from there being no restrictions on the dependence or independence of the X_i. In many instances $E[X]$ can be easily calculated by a judicious decomposition into simple (often indicator) random variables X_i.

Let σ be a random permutation on $\{1, \ldots, n\}$, uniformly chosen. Let $X(\sigma)$ be the number of fixed points of σ. To find $E[X]$ we decompose $X = X_1 + \ldots + X_n$ where X_i is the indicator random variable of the event $\sigma(i) = i$. Then

$$E[X_i] = \Pr[\sigma(i) = i] = \frac{1}{n}$$

so that

$$E[X] = \frac{1}{n} + \ldots + \frac{1}{n} = 1.$$

In applications we often use that there is a point in the probability space for which $X \geq E[X]$ and a point for which $X \leq E[X]$. We have selected results with a

purpose of describing this basic methodology. The following result of Szele (1943), is often-times considered the first use of the probabilistic method.

Theorem 2.1.1 *There is a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.*

Proof. In the random tournament let X be the number of Hamiltonian paths. For each permutation σ let X_σ be the indicator random variable for σ giving a Hamiltonian path – i.e., satisfying $(\sigma(i), \sigma(i+1)) \in T$ for $1 \leq i < n$. Then $X = \sum X_\sigma$ and

$$E[X] = \sum E[X_\sigma] = n!2^{-(n-1)}$$

Thus some tournament has at least $E[X]$ Hamiltonian paths. ■

Szele conjectured that the maximum possible number of Hamiltonian paths in a tournament on n players is at most $\frac{n!}{(2-o(1))^n}$. This was proved in Alon (1990a) and is presented in the Probabilistic Lens: Hamiltonian Paths (following Chapter 4).

2.2 SPLITTING GRAPHS

Theorem 2.2.1 *Let $G = (V, E)$ be a graph with n vertices and e edges. Then G contains a bipartite subgraph with at least $e/2$ edges.*

Proof. Let $T \subseteq V$ be a random subset given by $\Pr[x \in T] = 1/2$, these choices mutually independent. Set $B = V - T$. Call an edge $\{x, y\}$ crossing if exactly one of x, y are in T. Let X be the number of crossing edges. We decompose

$$X = \sum_{\{x,y\} \in E} X_{xy}$$

where X_{xy} is the indicator random variable for $\{x, y\}$ being crossing. Then

$$E[X_{xy}] = 1/2$$

as two fair coin flips have probability $1/2$ of being different. Then

$$E[X] = \sum_{\{x,y\} \in E} E[X_{xy}] = \frac{e}{2}.$$

Thus $X \geq e/2$ for some choice of T and the set of those crossing edges form a bipartite graph. ■

A more subtle probability space gives a small improvement.

Theorem 2.2.2 *If G has $2n$ vertices and e edges then it contains a bipartite subgraph with at least $\frac{en}{2n-1}$ edges. If G has $2n+1$ vertices and e edges then it contains a bipartite subgraph with at least $\frac{e(n+1)}{2n+1}$ edges.*

Proof. When G has $2n$ vertices let T be chosen uniformly from among all n-element subsets of V. Any edge $\{x, y\}$ now has probability $\frac{n}{2n-1}$ of being crossing and the proof concludes as before. When G has $2n+1$ vertices choose T uniformly from among all n-element subsets of V and the proof is similar. ■

Here is a more complicated example in which the choice of distribution requires a preliminary lemma. Let $V = V_1 \cup \ldots \cup V_k$ where the V_i are disjoint sets of size n. Let $h : [V]^k \to \{-1, +1\}$ be a two-coloring of the k-sets. A k-set E is crossing if it contains precisely one point from each V_i. For $S \subseteq V$ set $h(S) = \sum h(E)$, the sum over all k-sets $E \subseteq S$.

Theorem 2.2.3 *Suppose $h(E) = +1$ for all crossing k-sets E. Then there is an $S \subseteq V$ for which*

$$|h(S)| \geq c_k n^k.$$

Here c_k is a positive constant, independent of n.

Lemma 2.2.4 *Let P_k denote the set of all homogeneous polynomials $f(p_1, \ldots, p_k)$ of degree k with all coefficients having absolute value at most one and $p_1 p_2 \cdots p_k$ having coefficient one. Then for all $f \in P_k$ there exist $p_1, \ldots, p_k \in [0, 1]$ with*

$$|f(p_1, \ldots, p_k)| \geq c_k.$$

Here c_k is positive and independent of f.

Proof. Set

$$M(f) = \max_{p_1, \ldots, p_k \in [0,1]} |f(p_1, \ldots, p_k)|.$$

For $f \in P_k$, $M(f) > 0$ as f is not the zero polynomial. As P_k is compact and $M : P_k \to R$ is continuous, M must assume its minimum c_k. ■

Proof [Theorem 2.2.3] Define a random $S \subseteq V$ by setting

$$\Pr[x \in S] = p_i, \quad x \in V_i,$$

these choices mutually independent, p_i to be determined. Set $X = h(S)$. For each k-set E set

$$X_E = \begin{cases} h(E) & \text{if } E \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

Say E has type $(a_1, \ldots, a_k)$ if $|E \cap V_i| = a_i$, $1 \leq i \leq k$. For these E,

$$E[X_E] = h(E) \Pr[E \subseteq S] = h(E) p_1^{a_1} \cdots p_k^{a_k}.$$

Combining terms by type

$$E[X] = \sum_{a_1 + \ldots + a_k = k} p_1^{a_1} \cdots p_k^{a_k} \sum_{E \text{ of type } (a_1, \ldots, a_k)} h(E).$$

When $a_1 = \ldots = a_k = 1$ all $h(E) = 1$ by assumption so

$$\sum_{E \text{ of type } (1,\ldots,1)} h(E) = n^k.$$

For any other type there are fewer than n^k terms, each ± 1, so

$$\left| \sum_{E \text{ of type } (a_1,\ldots,a_k)} h(E) \right| \leq n^k.$$

Thus

$$E[X] = n^k f(p_1, \ldots, p_k)$$

where $f \in P_k$, as defined by Lemma 2.2.4.

Now select $p_1, \ldots, p_k \in [0, 1]$ with $|f(p_1, \ldots, p_k)| \geq c_k$. Then

$$E[|X|] \geq E[X] \geq c_k n^k.$$

Some particular value of $|X|$ must exceed or equal its expectation. Hence there is a particular set $S \subseteq V$ with

$$|X| = |h(S)| \geq c_k n^k.$$

■

Theorem 2.2.3 has an interesting application to Ramsey Theory. It is known (see Erdős (1965b)) that given any coloring with two colors of the k-sets of an n-set there exist k disjoint m-sets, $m = \Theta((\ln n)^{1/(k-1)})$, so that all crossing k-sets are the same color. From Theorem 2.2.3 there then exists a set of size $\Theta((\ln n)^{1/(k-1)})$, at least $\frac{1}{2} + \epsilon_k$ of whose k-sets are the same color. This is somewhat surprising since it is known that there are colorings in which the largest monochromatic set has size at most the $k - 2$-fold logarithm of n.

2.3 TWO QUICKIES

Linearity of Expectation sometimes gives very quick results.

Theorem 2.3.1 *There is a two-coloring of K_n with at most*

$$\binom{n}{a} 2^{1-\binom{a}{2}}$$

monochromatic K_a.

Proof [outline] Take a random coloring . Let X be the number of monochromatic K_a and find $E[X]$. For some coloring the value of X is at most this expectation. ■

In Chapter 15 it is shown how such a coloring can be found deterministically and efficiently.

Theorem 2.3.2 *There is a two-coloring of $K_{m,n}$ with at most*

$$\binom{m}{a}\binom{n}{b}2^{1-ab}$$

monochromatic $K_{a,b}$.

Proof [outline] Take a random coloring. Let X be the number of monochromatic $K_{a,b}$ and find $E[X]$. For some coloring the value of X is at most this expectation. ■

2.4 BALANCING VECTORS

The next result has an elegant *non*probabilistic proof, which we defer to the end of this chapter. Here $|v|$ is the usual Euclidean norm.

Theorem 2.4.1 *Let $v_1, \ldots, v_n \in R^n$, all $|v_i| = 1$. Then there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that*

$$|\epsilon_1 v_1 + \ldots + \epsilon_n v_n| \leq \sqrt{n},$$

and also there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that

$$|\epsilon_1 v_1 + \ldots + \epsilon_n v_n| \geq \sqrt{n}.$$

Proof. Let $\epsilon_1, \ldots, \epsilon_n$ be selected uniformly and independently from $\{-1, +1\}$. Set

$$X = |\epsilon_1 v_1 + \ldots + \epsilon_n v_n|^2.$$

Then

$$X = \sum_{i=1}^{n}\sum_{j=1}^{n} \epsilon_i \epsilon_j v_i \cdot v_j.$$

Thus

$$E[X] = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i \cdot v_j E[\epsilon_i \epsilon_j].$$

When $i \neq j$, $E[\epsilon_i \epsilon_j] = E[\epsilon_i]E[\epsilon_j] = 0$. When $i = j$, $\epsilon_i^2 = 1$ so $E[\epsilon_i^2] = 1$. Thus

$$E[X] = \sum_{i=1}^{n} v_i \cdot v_i = n.$$

Hence there exist specific $\epsilon_1, \ldots, \epsilon_n = \pm 1$ with $X \geq n$ and with $X \leq n$. Taking square roots gives the theorem. ■

The next result includes part of Theorem 2.4.1 as a linear translate of the $p_1 = \ldots = p_n = 1/2$ case.

Theorem 2.4.2 *Let $v_1, \ldots, v_n \in R^n$, all $|v_i| \leq 1$. Let $p_1, \ldots, p_n \in [0,1]$ be arbitrary and set $w = p_1 v_1 + \ldots + p_n v_n$. Then there exist $\epsilon_1, \ldots, \epsilon_n \in \{0,1\}$ so that, setting $v = \epsilon_1 v_1 + \ldots + \epsilon_n v_n$,*

$$|w - v| \leq \frac{\sqrt{n}}{2}.$$

Proof. Pick ϵ_i independently with

$$\Pr[\epsilon_i = 1] = p_i, \Pr[\epsilon_i = 0] = 1 - p_i.$$

The random choice of ϵ_i gives a random v and a random variable

$$X = |w - v|^2.$$

We expand

$$X = \left| \sum_{i=1}^{n} (p_i - \epsilon_i) v_i \right|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j (p_i - \epsilon_i)(p_j - \epsilon_j)$$

so that

$$E[X] = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j E[(p_i - \epsilon_i)(p_j - \epsilon_j)].$$

For $i \neq j$,

$$E[(p_i - \epsilon_i)(p_j - \epsilon_j)] = E[p_i - \epsilon_i] E[p_j - \epsilon_j] = 0.$$

For $i = j$,

$$E[(p_i - \epsilon_i)^2] = p_i (p_i - 1)^2 + (1 - p_i) p_i^2 = p_i (1 - p_i) \leq \frac{1}{4},$$

($E[(p_i - \epsilon_i)^2] = \mathrm{Var}[\epsilon_i]$, the *variance* to be discussed in Chapter 4.) Thus

$$E[X] = \sum_{i=1}^{n} p_i (1 - p_i) |v_i|^2 \leq \frac{1}{4} \sum_{i=1}^{n} |v_i|^2 \leq \frac{n}{4}$$

and the proof concludes as in that of Theorem 2.4.1. ■

2.5 UNBALANCING LIGHTS

Theorem 2.5.1 *Let $a_{ij} = \pm 1$ for $1 \leq i, j \leq n$. Then there exist $x_i, y_j = \pm 1$, $1 \leq i, j \leq n$ so that*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j \geq \left(\sqrt{\frac{2}{\pi}} + o(1) \right) n^{3/2}.$$

This result has an amusing interpretation. Let an $n \times n$ array of lights be given, each either on ($a_{ij} = +1$) or off ($a_{ij} = -1$). Suppose for each row and each column there is a switch so that if the switch is pulled ($x_i = -1$ for row i and $y_j = -1$ for column j) all of the lights in that line are "switched": on to off or off to on. Then for any initial configuration it is possible to perform switches so that the number of lights on minus the number of lights off is at least $(\sqrt{\frac{2}{\pi}} + o(1))n^{3/2}$.

Proof [Theorem 2.5.1] Forget the x's. Let $y_1, \ldots, y_n = \pm 1$ be selected independently and uniformly and set

$$R_i = \sum_{j=1}^{n} a_{ij} y_j,$$

$$R = \sum_{i=1}^{n} |R_i|.$$

Fix i. Regardless of a_{ij}, $a_{ij}y_j$ is $+1$ or -1 with probability $1/2$ and their values (over j) are independent. (I.e., whatever the i-th row is initially after random switching it becomes a uniformly distributed row, all 2^n possibilities equally likely.) Thus R_i has distribution S_n – the distribution of the sum of n independent uniform $\{-1, 1\}$ random variables – and so

$$E[|R_i|] = E[|S_n|] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right)\sqrt{n}.$$

These asymptotics may be found by estimating S_n by $\sqrt{n}N$ where N is standard normal and using elementary calculus. Alternatively, a closed form

$$E[|S_n|] = n2^{1-n}\binom{n-1}{\lfloor (n-1)/2 \rfloor}$$

may be derived combinatorially (a problem in the 1974 Putnam competition!) and the asymptotics follows from Stirling's formula.

Now apply Linearity of Expectation to R:

$$E[R] = \sum_{i=1}^{n} E[|R_i|] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}.$$

There exist $y_1, \ldots, y_n = \pm 1$ with R at least this value. Finally, pick x_i with the same sign as R_i so that

$$\sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij} y_j = \sum_{i=1}^{n} x_i R_i = \sum_{i=1}^{n} |R_i| = R \geq \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}.$$

■

Another result on unbalancing lights appears in the Probabilistic Lens: Unbalancing Lights (following Chapter 12.)

2.6 WITHOUT COIN FLIPS

A nonprobabilistic proof of Theorem 2.2.1 may be given by placing each vertex in either T or B sequentially. At each stage place x in either T or B so that at least half of the edges from x to previous vertices are crossing. With this effective algorithm at least half the edges will be crossing.

There is also a simple sequential algorithm for choosing signs in Theorem 2.4.1 When the sign for v_i is to be chosen a partial sum $w = \epsilon_1 v_1 + \ldots + \epsilon_{i-1} v_{i-1}$ has been calculated. Now if it is desired that the sum be small select $\epsilon_i = \pm 1$ so that $\epsilon_i v_i$ makes an acute (or right) angle with w. If the sum need be big make the angle obtuse or right. In the extreme case when all angles are right angles Pythagoras and induction give that the final w has norm $\sqrt{n}$, otherwise it is either less than $\sqrt{n}$ or greater than $\sqrt{n}$ as desired.

For Theorem 2.4.2 a greedy algorithm produces the desired ϵ_i. Given $v_1, \ldots, v_n \in R^n, p_1, \ldots, p_n \in [0,1]$ suppose $\epsilon_1, \ldots, \epsilon_{s-1} \in \{0,1\}$ have already been chosen. Set $w_{s-1} = \sum_{i=1}^{s-1} (p_i - \epsilon_i) v_i$, the partial sum. Select ϵ_s so that

$$w_s = w_{s-1} + (p_s - \epsilon_s) v_s = \sum_{i=1}^{s} (p_i - \epsilon_i) v_i$$

has minimal norm. A random $\epsilon_s \in \{0,1\}$ chosen with $\Pr[\epsilon_s = 1] = p_s$ gives

$$\begin{aligned} E[|w_s|^2] &= |w_{s-1}|^2 + 2 w_{s-1} \cdot v_s E[p_s - \epsilon_s] + |v_s|^2 E(p_s - \epsilon_s)^2 \\ &= |w_{s-1}|^2 + p_s(1-p_s)|v_s|^2 \end{aligned} \tag{2.1}$$

so for some choice of $\epsilon_s \in \{0,1\}$,

$$|w_s|^2 \leq |w_{s-1}|^2 + p_s(1-p_s)|v_s|^2.$$

As this holds for all $1 \leq s \leq n$ (taking $w_0 = 0$), the final

$$|w_n|^2 \leq \sum_{i=1}^{n} p_i(1-p_i)|v_i|^2.$$

While the proofs appear similar, a direct implementation of the proof of Theorem 2.4.2 to find $\epsilon_1, \ldots, \epsilon_n$ might take an exhaustive search with exponential time. In applying the greedy algorithm at the s-th stage one makes two calculations of $|w_s|^2$, depending on whether $\epsilon_s = 0$ or 1, and picks that ϵ_s giving the smaller value. Hence there are only a linear number of calculations of norms to be made and the entire algorithm takes only quadratic time. In Chapter 15 we discuss several similar examples in a more general setting.

2.7 EXERCISES

1. Suppose $n \geq 2$ and let $H = (V, E)$ be an n-uniform hypergraph with $|E| = 4^{n-1}$ edges. Show that there is a coloring of V by four colors so that no edge is monochromatic.

2. Prove that there is a positive constant c so that every set A of n nonzero reals contains a subset $B \subset A$ of size $|B| \geq cn$ so that there are no $b_1, b_2, b_3, b_4 \in B$ satisfying
$$b_1 + 2b_2 = 2b_3 + 2b_4.$$

3. Prove that every set of n non-zero **real** numbers contains a subset A of **strictly** more than $n/3$ numbers such that there are no $a_1, a_2, a_3 \in A$ satisfying $a_1 + a_2 = a_3$.

4. Suppose $p > n > 10m^2$, with p prime, and let $0 < a_1 < a_2, < \ldots < a_m < p$ be integers. Prove that there is an integer x, $0 < x < p$ for which the m numbers
$$((xa_i) \pmod p) \bmod n, \quad (1 \leq i \leq m)$$
are pairwise distinct.

5. Let H be a graph, and let $n > |V(H)|$ be an integer. Suppose there is a graph on n vertices and t edges containing no copy of H, and suppose that $tk > n^2 \log_e n$. Show that there is a coloring of the edges of the complete graph on n vertices by k colors with no monochromatic copy of H.

6. (*) Prove, using the technique in the probabilistic lens on Hamiltonian paths, that there is a constant $c > 0$ such that for every even $n \geq 4$ the following holds: For every undirected complete graph K on n vertices whose edges are colored red and blue, the number of alternating Hamilton cycles in K (that is, properly edge-colored cycles of length n) is at most
$$n^c \frac{n!}{2^n}.$$

7. Let $\mathcal{F}$ be a family of subsets of $N = \{1, 2, \ldots, n\}$, and suppose there are no $A, B \in \mathcal{F}$ satisfying $A \subset B$. Let $\sigma \in S_n$ be a random permutation of the elements of N and consider the random variable
$$X = |\{i : \{\sigma(1), \sigma(2), \ldots, \sigma(i)\} \in \mathcal{F}\}|.$$
By considering the expectation of X prove that $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

8. (*) Let X be a collection of pairwise orthogonal unit vectors in R^n and suppose the projection of each of these vectors on the first k coordinates is of Euclidean norm at least ϵ. Show that $|X| \leq k/\epsilon^2$, and this is tight for all $\epsilon = k/2^r < 1$.

9. Let $G = (V, E)$ be a bipartite graph with n vertices and a list $S(v)$ of more than $\log_2 n$ colors associated with each vertex $v \in V$. Prove that there is a proper coloring of G assigning to each vertex v a color from its list $S(v)$.

THE PROBABILISTIC LENS:
Brégman's Theorem

Let $A = [a_{ij}]$ be an $n \times n$ matrix with all $a_{ij} \in \{0, 1\}$. Let $r_i = \sum_{1 \leq j \leq n} a_{ij}$ be the number of ones in the i-th row. Let S be the set of permutations $\sigma \in S_n$ with $a_{i,\sigma i} = 1$ for $1 \leq i \leq n$. Then the permanent $\mathrm{per}(A)$ is simply $|S|$. The following result was conjectured by Minc and proved by Brégman (1973). The proof presented here is similar to that of Schrijver (1978).

Theorem 1 [Brégman's Theorem]

$$\mathrm{per}(A) \leq \prod_{1 \leq i \leq n} (r_i!)^{1/r_i}.$$

Pick $\sigma \in S$ and $\tau \in S_n$ independently and uniformly. Set $A^1 = A$. Let $R_{\tau 1}$ be the number of ones in row $\tau 1$ in A^1. Delete row $\tau 1$ and column $\sigma\tau 1$ from A^1 to give A^2. In general, let A^i denote A with rows $\tau 1, \ldots, \tau(i-1)$ and columns $\sigma 1, \ldots, \sigma\tau(i-1)$ deleted and let $R_{\tau i}$ denote the number of ones of row τi in A^i. (This is nonzero as the $\sigma\tau i$-th column has a one.) Set

$$L = L(\sigma, \tau) = \prod_{1 \leq i \leq n} R_{\tau i}.$$

We think, roughly, of L as Lazyman's permanent calculation. There are $R_{\tau 1}$ choices for a one in row $\tau 1$, each of which leads to a diferent subpermanent calculation. Instead, Lazyman takes the factor $R_{\tau 1}$, takes the one from permutation σ, and examines A^2. As $\sigma \in S$ is chosen uniformly Lazyman tends toward the high subpermanents and so it should not be surprising that he tends to overestimate the permanent. To make this precise we define the geometric mean $G[Y]$. If $Y > 0$ takes values $a_1, \ldots, a_s$ with probabilities $p_1, \ldots, p_s$ respectively, then $G[Y] = \prod a_i^{p_i}$.

Equivalently, $G[Y] = e^{E[\ln Y]}$. Linearity of Expectation translates into the geometric mean of a product being the product of the geometric means.

Claim 2.7.1 $\mathrm{per}(A) \leq G[L]$.

Proof. We show this for any fixed τ. Set $\tau 1 = 1$ for convenience of notation. We use induction on the size of the matrix. Reorder, for convenience, so that the first row has ones in the first r columns where $r = r_1$. For $1 \leq j \leq r$ let t_j be the permanent of A with the first row and j-th column removed or, equivalently, the number of $\sigma \in S$ with $\sigma 1 = j$. Set

$$t = \frac{t_1 + \ldots + t_r}{r}$$

so that $\mathrm{per}(A) = rt$. Conditioning on $\sigma 1 = j$, $R_2 \cdots R_n$ is Lazyman's calculation of $\mathrm{per}(A^2)$, where A^2 is A with the first row and j-th column removed. By induction

$$G[R_2 \cdots R_n | \sigma 1 = j] \geq t_j$$

and so

$$G[L] \geq \prod_{j=1}^{r} (rt_j)^{t_j/\mathrm{per}(A)} = r \prod_{j=1}^{r} t_j^{t_j/rt}.$$

Lemma 2

$$\left(\prod_{j=1}^{r} t_j^{t_j} \right)^{1/r} \geq t^t.$$

Proof. Taking logarithms, this is equivalent to

$$\frac{1}{r} \sum_{j=1}^{r} t_j \ln t_j \geq t \ln t$$

which follows from the convexity of the function $f(x) = x \ln x$. ■

Applying the Lemma,

$$G[L] \geq r \prod_{j=1}^{r} t_j^{t_j/rt} \geq r(t^t)^{1/t} = rt = \mathrm{per}(A).$$

■

Now we calculate $G[L]$ conditional on a fixed σ. For convenience of notation reorder so that $\sigma i = i$, all i, and assume that the first row has ones in precisely the first r_1 columns. With τ selected uniformly the columns $1, \ldots, r_1$ are deleted in order uniform over all $r_1!$ possibilities. R_1 is the number of those columns remaining when the first column is to be deleted. As the first column is equally likely to be in

any position among those r_1 columns R_1 is uniformly distributed from 1 to r_1 and $G[R_1] = (r_1!)^{1/r_1}$. "Linearity" then gives

$$G[L] = G\left[\prod_{i=1}^{n} R_i\right] = \prod_{i=1}^{n} G[R_i] = \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

The overall $G[L]$ is the geometric mean of the conditional $G[L]$ and hence has the same value. That is,

$$\text{per}(A) \leq G[L] = \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

3

Alterations

> Beauty is the first test: there is no permanent place in the world for ugly mathematics.
> – G.H. Hardy

The basic probabilistic method was described in Chapter 1 as follows: Trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this space with positive probability. In this chapter we consider situations where the "random" structure does not have all the desired properties but may have a few "blemishes." With a small alteration we remove the blemishes, giving the desired structure.

3.1 RAMSEY NUMBERS

Recall from Section 1.1 in Chapter 1 that $R(k,l) > n$ means there exists a two-coloring of the edges of K_n by red and blue so that there is neither a red K_k nor a blue K_l.

Theorem 3.1.1 *For any integer n,*

$$R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Proof. Consider a random two-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equally likely. For

any set R of k vertices let X_R be the indicator random variable for the event that the induced subgraph of K_n on R is monochromatic. Set $X = \sum X_R$, the sum over all such R. From Linearity of Expectation,

$$E[X] = \sum E[X_R] = m \text{ with } m = \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Thus there exists a two-coloring for which $X \leq m$. Fix such a coloring. Remove from K_n one vertex from each monochromatic k-set. At most m vertices have been removed (we may have "removed" the same vertex more than once but this only helps) so s vertices remain with $s \geq n - m$. This coloring on these s points has no monochromatic k-set. ■

We are left with the "calculus" problem of finding that n which will optimize the inequality. Some analysis shows that we should take $n \sim e^{-1}k2^{k/2}(1 - o(1))$ giving

$$R(k,k) > \frac{1}{e}(1 + o(1))k2^{k/2}.$$

A careful examination of Proposition 1.1.1 gives the lower bound

$$R(k,k) > \frac{1}{e\sqrt{2}}(1 + o(1))k2^{k/2}.$$

The more powerful Lovász Local Lemma – see Chapter 5 – gives

$$R(k,k) > \frac{\sqrt{2}}{e}(1 + o(1))k2^{k/2}.$$

The distinctions between these bounds may be considered inconsequential since the best known upper bound for $R(k,k)$ is $(4 + o(1))^k$. The upper bounds do not involve probabilistic methods and may be found, for example, in Graham, Rothschild and Spencer (1990). We give all three lower bounds in following our philosophy of emphasizing *methodologies* rather than results.

In dealing with the off-diagonal Ramsey numbers the distinction between the basic method and the alteration is given in the following two results.

Theorem 3.1.2 *If there exists $p \in [0,1]$ with*

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}} < 1$$

then $R(k,l) > n$.

Theorem 3.1.3 *For all integers n and $p \in [0,1]$,*

$$R(k,l) > n - \binom{n}{k} p^{\binom{k}{2}} - \binom{n}{l}(1-p)^{\binom{l}{2}}.$$

Proof. In both cases we consider a random two-coloring of K_n obtained by coloring each edge independently either red or blue, where each edge is red with probability

p. Let X be the number of red k-sets plus the number of blue l-sets. Linearity of Expectation gives

$$E[X] = \binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}}.$$

For Theorem 3.1.2, $E[X] < 1$ so there exists a two-coloring with $X = 0$. For Theorem 3.1.3 there exists a two-coloring with s "bad" sets (either red k-sets or blue l-sets), $s \leq E[X]$. Removing one point from each bad set gives a coloring of at least $n - s$ points with no bad sets. ■

The asymptotics of Theorems 3.1.2, 3.1.3 can get fairly complex. Oftentimes Theorem 3.1.3 gives a substantial improvement on Theorem 3.1.2. Even further improvements may be found using the Lovász Local Lemma. These bounds have been analyzed in Spencer (1977).

3.2 INDEPENDENT SETS

Here is a short and sweet argument that gives roughly half of the celebrated Turán's Theorem. $\alpha(G)$ is the independence number of a graph G; $\alpha(G) \geq t$ means there exist t vertices with no edges between them.

Theorem 3.2.1 *Let $G = (V, E)$ have n vertices and $nd/2$ edges, $d \geq 1$. Then $\alpha(G) \geq n/2d$.*

Proof. Let $S \subseteq V$ be a random subset defined by

$$\Pr[v \in S] = p,$$

p to be determined, the events $v \in S$ being mutually independent. Let $X = |S|$ and let Y be the number of edges in $G|_S$. For each $e = \{i, j\} \in E$ let Y_e be the indicator random variable for the event $i, j \in S$ so that $Y = \sum_{e \in E} Y_e$. For any such e,

$$E[Y_e] = \Pr[i, j \in S] = p^2,$$

so by Linearity of Expectation,

$$E[Y] = \sum_{e \in E} E[Y_e] = \frac{nd}{2} p^2.$$

Clearly $E[X] = np$, so, again by Linearity of Expectation,

$$E[X - Y] = np - \frac{nd}{2} p^2.$$

We set $p = 1/d$ (here using $d \geq 1$) to maximize this quantity, giving

$$E[X - Y] = \frac{n}{2d}.$$

Thus there exists a specific S for whom the number of vertices of S minus the number of edges in S is at least $n/2d$. Select one vertex from each edge of S and delete it. This leaves a set S^* with at least $n/2d$ vertices. All edges having been destroyed, S^* is an independent set. ■

The full result of Turán is given in The Probabilistic Lens: Turán's Theorem (following Chapter 6).

3.3 COMBINATORIAL GEOMETRY

For a set S of n points in the unit square U, let $T(S)$ be the minimum area of a triangle whose vertices are three distinct points of S. Put $T(n) = \max T(S)$, where S ranges over all sets of n points in U. Heilbronn conjectured that $T(n) = O(1/n^2)$. This conjecture was disproved by Komlós, Pintz and Szemerédi (1982) who showed, by a rather involved probabilistic construction, that there is a set S of n points in U such that $T(S) = \Omega(\log n/n^2)$. As this argument is rather complicated, we only present here a simpler one showing that $T(n) = \Omega(1/n^2)$.

Theorem 3.3.1 *There is a set S of n points in the unit square U such that $T(S) \geq 1/(100n^2)$.*

Proof. We first make a calculation. Let P, Q, R be independently and uniformly selected from U and let $\mu = \mu(PQR)$ denote the area of the triangle PQR. We bound $\Pr[\mu \leq \epsilon]$ as follows. Let x be the distance from P to Q so that $\Pr[b \leq x \leq b + \Delta b] \leq \pi(b + \Delta b)^2 - \pi b^2$ and in the limit $\Pr[b \leq x \leq b + db] \leq 2\pi b db$. Given P, Q at distance b, the altitude from R to the line PQ must have height $h \leq 2\epsilon/b$ and so R must lie in a strip of width $4\epsilon/b$ and length at most $\sqrt{2}$. This occurs with probability at most $4\sqrt{2}\epsilon/b$. As $0 \leq b \leq \sqrt{2}$ the total probability is bounded by

$$\int_0^{\sqrt{2}} (2\pi b)(4\sqrt{2}\epsilon/b)db = 16\pi\epsilon.$$

Now let $P_1, \ldots, P_{2n}$ be selected uniformly and independently in U and let X denote the number of triangles $P_i P_j P_k$ with area less than $1/(100n^2)$. For each particular i, j, k the probability of this occuring is less than $0.6n^{-2}$ and so

$$E[X] \leq \binom{2n}{3}(0.6n^{-2}) < n.$$

Thus there exists a specific set of $2n$ vertices with fewer than n triangles of area less than $1/(100n^2)$. Delete one vertex from the set from each such triangle. This leaves at least n vertices and now no triangle has area less than $1/(100n^2)$. ■

We note the following construction of Erdős showing $T(n) \geq 1/(2(n-1)^2)$ with n prime. On $[0, n-1] \times [0, n-1]$ consider the n points (x, x^2) where x^2 is reduced mod$n$. (More formally, (x, y) where $y \equiv x^2 \bmod n$ and $0 \leq y < n$.) If some three points of this set were collinear they would line on a line $y = mx + b$ and m would be

a rational number with denominator less than n. But then in Z_n^2 the parabola $y = x^2$ would intersect the line $y = mx + b$ in three points, so that the quadratic $x^2 - mx - b$ would have three distinct roots, an impossibility. Triangles between lattice points in the plane have as their areas either half-integers or integers, hence the areas must be at least $1/2$. Contracting the plane by an $n - 1$ factor in both coordinates gives the desired set. While this gem does better than Theorem 3.3.1 it does not lead to the improvements of Komlós, Pintz and Szemerédi.

3.4 PACKING

Let C be a bounded measurable subset of R^d and let $B(x)$ denote the cube $[0, x]^d$ of side x. A *packing* of C into $B(x)$ is a family of mutually disjoint copies of C, all lying inside $B(x)$. Let $f(x)$ denote the largest size of such a family. The packing constant $\delta = \delta(C)$ is defined by

$$\delta(C) = \mu(C) \lim_{x \to \infty} f(x) x^{-d},$$

where $\mu(C)$ is the measure of C. This is the maximal proportion of space that may be packed by copies of C. (This limit can be proven always to exist but even without that result the following result holds with lim replaced by lim inf.)

Theorem 3.4.1 *Let C be bounded, convex, and centrally symmetric around the origin. Then*

$$\delta(C) \geq 2^{-d-1}.$$

Proof. Let P, Q be selected independently and uniformly from $B(x)$ and consider the event $(C + P) \cap (C + Q) \neq \emptyset$. For this to occur we must have, for some $c_1, c_2 \in C$,

$$P - Q = c_1 - c_2 = 2\frac{c_1 - c_2}{2} \in 2C$$

by central symmetry and convexity. The event $P \in Q + 2C$ has probability at most $\mu(2C)x^{-d}$ for each given Q, hence

$$\Pr[(C + P) \cap (C + Q) \neq \emptyset] \leq \mu(2C)x^{-d} = 2^d x^{-d} \mu(C).$$

Now let $P_1, \ldots, P_n$ be selected independently and uniformly from $B(x)$ and let X be the number of $i < j$ with $(C + P_i) \cap (C + P_j) \neq \emptyset$. From linearity of expectation,

$$E[X] \leq \frac{n^2}{2} 2^d x^{-d} \mu(C).$$

Hence there exists a specific choice of n points with fewer than that many intersecting copies of C. For each P_i, P_j with $(C + P_i) \cap (C + P_j) \neq \emptyset$ remove either P_i or P_j from the set. This leaves at least $n - \frac{n^2}{2} 2^d x^{-d} \mu(C)$ nonintersecting copies of C. Set $n = x^d 2^{-d} / \mu(C)$ to maximize this quantity, so that there are at least $x^d 2^{-d-1} / \mu(C)$

nonintersecting copies of C. These do not all lie inside $B(x)$ but, letting w denote an upper bound on the absolute values of the coordinates of the points of C, they do all lie inside a cube of side $x + 2w$. Hence

$$f(x + 2w) \geq x^d 2^{-d-1}/\mu(C)$$

and so

$$\delta(C) \geq \lim_{x\to\infty} \mu(C) f(x + 2w)(x + 2w)^{-d} \geq 2^{-d-1}.$$

■

A simple greedy algorithm does somewhat better. Let $P_1, \ldots, P_m$ be *any* maximal subset of $[0, x]^d$ with the property that the sets $C + P_i$ are disjoint. We have seen that $C + P_i$ overlaps $C + P$ if and only if $P \in 2C + P_i$. Hence the sets $2C + P_i$ must cover $[0, x]^d$. As each such set has measure $\mu(2C) = 2^d\mu(C)$ we must have $m \geq x^d 2^{-d}/\mu(C)$. As before, all sets $C + P_i$ lie in a cube of side $x + 2w$, w a constant, so that

$$f(x + 2w) \geq m \geq x^d 2^{-d}/\mu(C)$$

and so

$$\delta(C) \geq 2^{-d}.$$

A still further improvement appears in the Probabilistic Lens: Efficient Packing (following Chapter 13).

3.5 RECOLORING

Suppose that a random coloring leaves a set of blemishes. Here we apply a random recoloring to the blemishes to remove them. If the recoloring is too weak then not all the blemishes are removed. If the recoloring is too strong then new blemishes are created. The recoloring is given a parameter p and these too possibilities are decreasing and increasing functions of p. Calculus then points us to the optimal p.

We use the notation of §1.3 on Property B: $m(n) > m$ means that given any n-uniform hypergraph $H = (V, E)$ with m edges there exists a two-coloring of V so that no edge is monochromatic. Beck (1978) improved Erdős' 1963 bound to $m(n) = \Omega(2^n n^{1/3})$. Building on his methods, Radhakrishnan and Srinivasan (2000) proved $m(n) = \Omega(2^n(n/\ln n)^{1/2})$ and it is that proof we shall give. While this proof is neither long nor technically complex it has a number of subtle and beautiful steps and it is not surprising that it took more than thirty-five years to find it. That said, the upper and lower bounds on $m(n)$ remain quite far apart!

Theorem 3.5.1 *If there exists* $p \in [0, 1]$ *with*

$$k(1 - p)^n + k^2 p < 1$$

then $m(n) > 2^{n-1}k$.

Corollary 3.5.2 $m(n) = \Omega(2^n(n/\ln n)^{1/2})$.

Proof. Bound $1 - p \leq e^{-p}$. The function $ke^{-pn} + k^2p$ is minimized at $p = \ln(n/k)/n$. Substituting back in, if

$$\frac{k^2}{n}[1 + \ln(n/k)] < 1$$

then the condition of Theorem 3.5.1 holds. This inequality is true when $k = c(n/\ln n)^{1/2}$ for any $c < \sqrt{2}$ with n sufficiently large. ■

The condition of Theorem 3.5.1 is somewhat typical; one wants the total failure probability to be less than 1 and there are two types of failure. Oftentimes one finds reasonable bounds by requiring the stronger condition that each failure type has probability less than one-half. Here $k^2p \leq \frac{1}{2}$ gives $p \leq \frac{1}{2}k^{-2}$. Plugging the maximal possible p into the second inequality $k(1-p)^n \leq \frac{1}{2}$ gives $2k^2 \ln(2k) \leq n$. This again holds when $k = c(n/\ln n)^{1/2}$ though now we have the weaker condition $c < 1$. We recommend this rougher approach as a first attempt at a problem, when the approximate range of the parameters is still in doubt. The refinements of calculus can be placed in the published work!

Proof [Theorem 3.5.1] Fix $H = (V, E)$ with $m = 2^{n-1}k$ edges and p satisfying the condition. We describe a randomized algorithm that yields a coloring of V. It is best to preprocess the randomness: Each $v \in V$ flips a first coin, which comes up heads with probability $\frac{1}{2}$ and a second coin, which comes up heads (representing potential recoloration) with probability p. In addition (and importantly), the vertices of V are ordered randomly.

Step 1. Color each $v \in V$ red if its first coin was heads, otherwise blue. Call this the first coloring. Let D (for dangerous) denote the set of $v \in V$ that lie in some (possibly many) monochromatic $e \in E$.

Step 2. Consider the elements of D sequentially in the (random) order of V. When d is being considered call it still dangerous if there is some (possibly many) $e \in H$ containing d that was monochromatic in the first coloring and for which no vertices have yet changed color. If d is not still dangerous then do nothing. But if it is still dangerous then check its second coin. If it is heads then change the color of d, otherwise do nothing. We call the coloring at the time of termination the final coloring.

We say the algorithm fails if some $e \in H$ is monochromatic in the final coloring. We shall bound the failure probability by $k(1-p)^n + k^2p$. The assumption of Theorem 3.5.1 then assures us that with positive probability the algorithm succeeds. This, by our usual magic, means that there is some running of the algorithm which yields a final coloring with no monochromatic e, that is, there exists a two-coloring of V with no monochromatic edge. For convenience, we bound the probability that some $e \in H$ is red in the final coloring, the failure probability for the algorithm is at most twice that.

An $e \in E$ can be red in the final coloring in two ways. Either e was red in the first coloring and remained red through to the final coloring or e was not red in the first coloring but was red in the final coloring. (The structure of the algorithm assures us that points cannot change color more than once.) Let A_e be the first event and C_e the

second. Then

$$\Pr[A_e] = 2^{-n}(1-p)^n.$$

The first factor is the probability e is red in the first coloring, that all first coins of e came up heads. The second factor is the probability that all second coins came up tails. If they all did, then no $v \in e$ would be recolored in Step 2. Inversely, if any second coins of $v \in e$ came up heads there would be a *first* v (in the ordering) that came up heads. When it did v was still dangerous as e was still monochromatic and so v does look at its second coin and change its color. We have

$$2\sum_{e \in H} \Pr[A_e] = k(1-p)^n$$

giving the first addend of our failure probability.

In Beck's 1978 proof, given in our first edition, there was no notion of "still dangerous" – every $d \in D$ changed its color if and only if its second coin was heads. The values $\Pr[A_e] = 2^{-n}(1-p)^n$ are the same in both arguments. Beck's had bounded $\Pr[C_e] \leq k^2 p e^{pn}$. The new argument avoids excessive recoloration and leads to a better bound on $\Pr[C_e]$. We turn to the ingenious bounding of $\Pr[C_e]$.

For distinct $e, f \in E$ we say e *blames* f if:

- e, f overlap in precisely one element. Call it v.
- In the first coloring f was blue and in the final coloring e was red.
- In Step 2 v was the *last* vertex of e that changed color from blue to red.
- When v changed its color f was still entirely blue.

Suppose C_e holds. Some points of e changed color from blue to red so there is a *last* point v that did so. But why did v flip its coin? It must have been still dangerous. That is, v must be in some (perhaps many) set f that was blue in the first coloring and was still blue when v was considered. Can e, f overlap in another vertex v' ? No! For such a v' would necessarily have been blue in the first coloring (as $v' \in f$) and red in the final coloring (as $v' \in e$), but then v' changed color before v. Hence f was no longer entirely blue when v was considered and so e could not blame f. Therefore, when C_e holds, e blames some f. Let B_{ef} be the event that e blames f. Then $\sum_e \Pr[C_e] \leq \sum_{e \neq f} \Pr[B_{ef}]$. As there are less than $(2^{n-1}k)^2$ pairs $e \neq f$ it now suffices to bound $\Pr[B_{ef}] \leq 2^{1-2n}p$.

Let e, f with $e \cap f = \{v\}$ (otherwise B_{ef} cannot occur) be fixed. The random ordering of V induces a random ordering σ of $e \cup f$. Let $i = i(\sigma)$ denote the number of $v' \in e$ coming before v in the ordering and let $j = j(\sigma)$ denote the number of $v' \in f$ coming before v in the ordering. Fixing σ we claim

$$\Pr[B_{ef}|\sigma] \leq \frac{p}{2} 2^{-n+1}(1-p)^j 2^{-n+1+i} \left(\frac{1+p}{2}\right)^i.$$

Let's take the factors one at a time. First, v itself must start blue and turn red. Second, all other $v' \in f$ must start blue. Third, all $v' \in f$ coming before v must have second coin tails. Fourth, all $v' \in e$ coming after v must start red (since v is the last point of e to change color). Finally, all $v' \in e$ coming before v must either start red or start

blue and turn red. [The final factor may well be a substantial overestimate. Those $v' \in e$ coming before v which start blue must not only have second coin heads but must themselves lie in an $e' \in H$ monochromatic under the first coloring. Attempts to further improve bounds on $m(n)$ have often centered on this overestimate but (thus far!) to no avail.]

We can then write

$$\Pr[B_{ef}] \leq 2^{1-2n} pE\left[(1+p)^i(1-p)^j\right]$$

where the expectation is over the uniform choice of σ. The following gem therefore completes the argument.

Lemma 3.5.3 $E\left[(1+p)^i(1-p)^j\right] \leq 1$.

Proof. Fix a matching between $e - \{v\}$ and $f - \{v\}$; think of Mr. & Mrs. Jones; Mr. & Mrs. Smith, etc. Condition on how many of each pair (two Joneses, one Smith, no Taylors, etc.) come before v. The conditional expectation of $(1+p)^i(1-p)^j$ splits into factors for each pair. When there is no Taylor there is no factor. When there are two Joneses there is a factor $(1+p)(1-p) \leq 1$. When there is one Smith the factor is equally likely to be $1+p$ or $1-p$ and so the conditional expectation gets a factor of one. All factors are at most one so their product is at most 1. ■
The desired result follows. ■

3.6 CONTINUOUS TIME

Discrete random processes can sometimes be analyzed by placing them in a continuous time framework. This allows the powerful methods of analysis (such as integration!) to be applied. The approach seems most effective when dealing with random orderings. We give two examples.

Property B. We modify the proof that $m(n) = \Omega(2^n n^{1/2} \ln^{-1/2} n)$ of the previous section. We assign to each vertex $v \in V$ a "birth time" x_v. The x_v are independent real variables, each uniform in $[0,1]$. The ordering of V is then the ordering (under less than) of the x_v. We now claim

$$\Pr[B_{ef}] \leq \sum_{l=0}^{n-1} \binom{n-1}{l} 2^{1-2n} \int_0^1 x^l p^{l+1} (1-xp)^{n-1}\, dx.$$

For $T \subseteq e - \{v\}$ let B_{efT} be the event that B_{ef} and in the first coloring e had precisely $T \cup \{v\}$ blue. There are $\binom{n-1}{l}$ choices for an l-set T, with l ranging from 0 to $n-1$. The first coloring on $e \cup f$ is then determined and has probability 2^{1-2n} of occuring. Suppose v has birth time $x_v = x$. All $w \in T \cup \{v\}$ must have second coin flip heads – probability p^{l+1}. All $w \in T$ must be born before v – so that $x_w < x$ which has probability x^l. No $w \in f - \{v\}$ can be born before v and have coin flip heads. Each such w has probability xp of doing that so there is probability

$(1-xp)^{n-1}$ that no w does. As $x_v = x$ was uniform in $[0,1]$ we integrate over x. Recombining terms,

$$\Pr[B_{ef}] \leq 2^{1-2n}p \int_0^1 (1+xp)^{n-1}(1-xp)^{n-1}\,dx.$$

The integrand is always at most one so $\Pr[B_{ef}] \leq 2^{1-2n}p$. The remainder of the proof is unchanged.

Random Greedy Packing. Let H be a $(k+1)$-uniform hypergraph on a vertex set V of size N. The $e \in H$, which we call edges, are simply subsets of V of size $k+1$. We assume:
Degree Condition: Every $v \in V$ is in precisely D edges.
Codegree Condition: Every distinct pair $v, v' \in V$ have only $o(D)$ edges in common.

We think of k fixed ($k = 2$ being an illustrative example) and the asymptotics as $N, D \to \infty$, with no set relationship between N and D.

A packing is a family P of vertex disjoint edges $e \in H$. Clearly $|P| \leq N/(k+1)$. We define a randomized algorithm to produce a (not necessarily optimal) packing. Assign to each $e \in H$ uniformly and independently a birth time $x_e \in [0, D)$. (The choice of $[0, D)$ rather than $[0, 1]$ proves to be a technical convenience. Note that as the x_e are real variables with probability one there are no ties.) At time zero $P \leftarrow \emptyset$. As time progresses from 0 to D when an edge e is born it is added to P if possible – that is, unless there is already some $e' \in P$ which overlaps e. Let P_c denote the value of P just before time c – when all e with birthtimes $t_e < c$ have been examined. Set $P^{\text{FINAL}} = P_D$. Note that by time D all edges have been born and their births were in random order. Thus P^{FINAL} is identical to the discrete process – often called the random greedy algorithm – in which H is first randomly ordered and then the $e \in H$ are considered sequentially.

Theorem 3.6.1 [Spencer (1995)] *The expected value of* $|P^{\text{FINAL}}|$ *is asymptotic to* $N/(k+1)$.

We say $v \in V$ survives at time c if no $e \in P_c$ contains v and we let S_c denote the set of $v \in V$ so surviving. Rather than looking at P^{FINAL} we shall examine P_c where c is an arbitrary fixed nonnegative real. Let

$$f(c) = \lim^{*} \Pr[v \in S_c]$$

where, formally, we mean here that for all $\epsilon > 0$ there exist D_0, N_0 and $\delta > 0$ so that if H is (k+1)-uniform on $N > N_0$ vertices with each v in $D > D_0$ edges and every distinct pair $v, v' \in V$ has less than δD common edges then $|f(c) - \Pr[v \in S_c]| < \epsilon$ for all $v \in V$.

The heart of the argument lies in showing that $f(c)$ exists by defining a continuous time birth process yielding that value. We now describe the birth process, omitting some of the epsilondeltamanship needed to formally show the limit.

Our birth process starts at time c and time goes backwards to 0. It begins with root Eve, our anthropomorphized v. Eve has births in time interval $[0, c)$. The number

of births is given by a Poisson distribution with mean c and given their number their times are uniformly and independently distributed. [This is a standard Poisson process with intensity one. Equivalently, on any infinitesmal time interval $[x, x+dx)$ Eve has probability dx of giving birth and these events are independent over disjoint intervals.] Our fertile Eve always gives birth to k-tuplets. Each child is born fertile under the same rules, so if Alice in born at time x she (in our unisexual model) has a Poisson distribution with mean x of births, uniformly distributed in $[0, x)$.

The resulting random tree $T = T_c$ can be shown to be finite (note the time interval is finite) with probability 1. Given a finite T we say for each vertex Alice that Alice survives or dies according to the following scheme.

Menendez Rule: If Alice has given birth to a set (or possibly several sets) of k-tuplets all of whom survived then she dies; otherwise she survives.

In particular, if Alice is childless she survives. We can then work our way up the tree to determine of each vertex whether she survives or dies.

Example. $c = 10, k = 2$. Eve gives birth to Alice, Barbara at time 8.3 and then to Rachel, Sienna at time 4.3. Alice gives birth to Nancy, Olive at time 5.7 and Rachel gives birth to Linda, Mayavati at time 0.4. There are no other births. Leaves Nancy, Olive, Linda , Mayavati, Barbara and Sienna then survive. Working up the tree Alice and Rachel die. In neither of Eve's births did both children survive and therefore Eve survives.

We define $f(c)$ to be the probability that the root Eve survives in the random birthtree $T = T_c$.

We outline the equivalence by defining a tree $T = T_c(v)$ for $v \in H$. For each edge e containing v with birthtime $t = t_e < c$ we say that $e - \{v\}$ is a set of k-tuplets born to v at time t. We work recursively; if w is born at time t then for each e' containing w with birthtime $t' = t_{e'} < t$ we say that $e' - \{w\}$ is a set of k-tuplets born to w at time t'. Possibly this process does not give a tree since the same vertex w may be reached in more than one way – the simplest example is if $v \in e, e'$ where both have birthtimes less than c and e, e' share another common vertex w. Then the process is stillborn and $T_c(v)$ is not defined. We'll argue that for any particular tree T,

$$\lim^{*} \Pr[T_c(v) \cong T] = \Pr[T_c = T]. \tag{3.1}$$

As $\sum_T \Pr[T_c = T] = 1$ this gives a rather roundabout argument that the process defining $T_c(v)$ is almost never stillborn.

We find $T_c(v)$ in stages. First consider the D edges e containing v. The number of them with birthtime $t_e < c$ has Binomial Distribution BIN$[D, \frac{c}{D}]$ which approaches (critically) the Poisson Distribution with mean c. Given that there are l such e their birthtimes t_e are uniformly distributed. There are (by the codegree condition) $o(D^2)$ pairs e, e' containing v and also some other vertex so there is probability $o(1)$ that two such e, e' have birthtime less than c. Now suppose $T_c(v)$ has been built out to a certain level and a vertex w has been born at time t. There are only $o(D)$ common edges between w and any of the finite number of w' already born, so there are still $\sim D$ edges e containing w and no other such w'. We now examine their

birthtimes, the number with $t_e < x$ has Binomial Distribution BIN$[D - o(D), \frac{x}{D}]$ which approaches the Poisson Distribution with mean x. As above, almost surely no two such e, e' will have a common vertex other than w itself. For any fixed T the calculation of $\Pr[T_c(v) \cong T]$ involves a finite number of these limits, which allows us to conclude (3.1).

With $c < d$ the random tree T_d includes T_c as a subtree by considering only those births of Eve occuring in $[0, c)$. If Eve survives in T_d she must survive in T_c. Hence $f(d) \leq f(c)$. We now claim

$$\lim_{c \to \infty} f(c) = 0.$$

If not, the nondecreasing f would have a limit $L > 0$ and all $f(x) \geq L$. Suppose in T_c Eve had i births. In each birth there would be probability at least L^k that all k children survived. The probability that Eve survived would then be at most $(1 - L^k)^i$. Since the number of Eve's births is Poisson with mean c,

$$f(c) \leq \sum_{i=0}^{\infty} e^{-c} \frac{c^i}{i!} (1 - L^k)^i = e^{-L^k c}$$

but then $\lim_{c \to \infty} f(c) = 0$, a contradiction.

By linearity of expectation $E[|S_c|] \to f(c)n$. As $(k+1)|P_c| + |S_c| = n$, $E[|P_c|] \to (1 - f(c))n/(k+1)$. But $E[|P^{\text{FINAL}}|] \geq E[|P_c|]$. We make $f(c)$ arbitrarily small by taking c appropriately big, so that $E[|P^{\text{FINAL}}|] \geq (1 - o(1))n/(k+1)$. As $|P^{\text{FINAL}}| \leq n/(k+1)$ always, the theorem follows.

Remark. We can actually say more about $f(c)$. For Δc small, $f(c + \Delta c) - f(c) \sim -(\Delta c) f(c)^{k+1}$ as, roughly, an Eve starting at time $c + \Delta c$ might have a birth in time interval $[c, c + \Delta c)$ all of whose children survive while Eve has no births in $[0, c)$ all of whose children survive. Letting $\Delta c \to 0$ yields the differential equation $f'(c) = -f(c)^{k+1}$. The initial value $f(0) = 1$ gives a unique solution $f(c) = (1 + ck)^{-1/k}$. It is intriguing to plug in $c = D$. This is not justified as our limit arguments were for c fixed and $N, D \to \infty$. Nonetheless, that *would* yield $E[|S_D|] = O(ND^{-1/k})$, that the random greedy algorithm would leave $O(ND^{-1/k})$ vertices uncovered. Suppose we replace the codegree condition by the stronger condition that every distinct pair $v, v' \in V$ have at most one edge in common. There is computer simulation data that in those cases the random greedy algorithm does leave $O(ND^{-1/k})$ vertices uncovered. This remains an open question, though it is shown in Alon, Kim and Spencer (1997) that this is the case for a modified version of the greedy algorithm.

Corollary 3.6.2 *Under the assumptions of the theorem there exists a packing P of size $\sim N/(k+1)$.*

Proof. We have defined a random process which gives a packing with expected size $\sim N/(k+1)$ and our usual magic implies such a P must exist. ■

In particular, this gives an alternate proof to the Erdős-Hanani conjecture, first proved by Rödl as given in §4.7. We use the notation of that section and define the packing number $m(n,k,l)$ as the maximal size of a family F of k-element subsets of $[n] = \{1,\ldots,n\}$ such that no l-set is contained in more than one k-set. Define a hypergraph $H = H(n,k,l)$ as follows: The vertices of H are the l-element subsets of $[n]$. For each k-element $A \subset [n]$ we define an edge e_A as the set of l-element subsets of A. A family F satisfying the above conditions then corresponds to a packing $P = \{e_A : A \in F\}$ in H. H has $N = \binom{n}{l}$ vertices. Each edge e_A has size $K+1 = \binom{k}{l}$. Each vertex is in $D = \binom{n-l}{k-l}$ edges. The number of edges containing two vertices v, v' depends on their intersection. It is largest (given $v \neq v'$) when v, v' (considered as l-sets) overlap in $l-1$ points and then it is $\binom{n-l-1}{k-l-1}$. We assume (as in §4.7) that k, l are fixed and $n \to \infty$ so this number of common edges is $o(D)$. The assumptions of §4.7 give $K+1$ fixed, $N, D \to \infty$ so that there exists P with

$$m(n,k,l) = |P| \sim N/(K+1) \sim \frac{\binom{n}{l}}{\binom{k}{l}}.$$

3.7 EXERCISES

1. Prove that the Ramsey number $r(k,k)$ satisfies, for every integer n,

$$r(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}},$$

and conclude that

$$r(k,k) \geq (1-o(1))\frac{k}{e}2^{k/2}.$$

2. Prove that the Ramsey number $r(4,k)$ satisfies

$$r(4,k) \geq \Omega((k/\ln k)^2).$$

3. Prove that every three-uniform hypergraph with n vertices and $m \geq n/3$ edges contains an independent set of size at least $\frac{2n^{3/2}}{3\sqrt{3}\sqrt{m}}$.

4. (*) Show that there is a finite n_0 such that any directed graph on $n > n_0$ vertices in which each outdegree is at least $\log_2 n - \frac{1}{10}\log_2\log_2 n$ contains an even simple directed cycle.

THE PROBABILISTIC LENS:
High Girth and High Chromatic Number

Many consider this one of the most pleasing uses of the probabilistic method, as the result is surprising and does not appear to call for nonconstructive techniques. The *girth* of a graph G is the size of its shortest cycle, $\alpha(G)$ is the size of the largest independent set in G and $\chi(G)$ denotes its chromatic number.

Theorem 1 [Erdős (1959)] *For all k, l there exists a graph G with $girth(G) > l$ and $\chi(G) > k$.*

Proof. Fix $\theta < 1/l$ and let $G \sim G(n,p)$ with $p = n^{\theta-1}$. (I.e., G is a random graph on n vertices chosen by picking each pair of vertices as an edge randomly and independently with probability p). Let X be the number of cycles of size at most l. Then

$$E[X] = \sum_{i=3}^{l} \frac{(n)_i}{2i} p^i \leq \sum_{i=3}^{l} \frac{n^{\theta i}}{2i} = o(n)$$

as $\theta l < 1$. In particular,

$$\Pr[X \geq n/2] = o(1).$$

Set $x = \lceil \frac{3}{p} \ln n \rceil$ so that

$$\Pr[\alpha(G) \geq x] \leq \binom{n}{x}(1-p)^{\binom{x}{2}} < \left[ne^{-p(x-1)/2}\right]^x = o(1).$$

Let n be sufficiently large so that both these events have probability less than 0.5. Then there is a specific G with less than $n/2$ cycles of length at most l and with

$\alpha(G) < 3n^{1-\theta} \ln n$. 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 than l and $\alpha(G^*) \leq \alpha(G)$. Thus

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

To complete the proof, let n be sufficiently large so that this is greater than k. ■

This page intentionally left blank

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

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

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

Theorem 4.1.1 [Chebyschev's Inequality] *For any positive* λ,

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

Proof.

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

■

The use of Chebyschev's Inequality is called the *Second Moment Method.*

Chebyschev's Inequality is best 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 μ. Note, however, that when X is a normal distribution with mean μ and standard deviation σ then

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

and for λ 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

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

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

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

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 – i.e., that $X_i = 1$ 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

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

and so

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

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 folllowing 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 ν. [Here 10 could be any (large) constant.] Now

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

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

$$E[X_p] = 1/p + O(1/n).$$

By linearity of expectation,

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

where here we used the well-known fact that $\sum_{p \leq x} \frac{1}{p} = \ln\ln x + O(1)$, 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 $\mathrm{Var}[X_p] = \frac{1}{p}(1 - \frac{1}{p}) + O(\frac{1}{n})$,

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

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

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

Thus

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

Thus

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

and similarly

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

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

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

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

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

Theorem 4.2.2 *Let λ be fixed, positive, negative or zero. Then*

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

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

$$\mu=E[Y]=\sum_{p\le M}p^{-1}=\ln\ln n+o((\ln\ln n)^{1/2})$$

and

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

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

For any distinct primes $p_1,\dots,p_s\le M$,

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

We let k be an arbitrary fixed positive integer and compare $E[\tilde{X}^k]$ and $E[\tilde{Y}^k]$. 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 $X_{p_1} \ldots X_{p_s}$. The same expansion applies to $\tilde{Y}$. As the corresponding terms have expectations within $O(n^{-1})$ the total difference

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

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.

4.3 MORE BASICS

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

$$\Pr[X > 0] \leq E[X]$$

so that if $E[X] \to 0$ 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 $E[X] \to \infty$. 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 $E[X]$ can make for lively debate but few would deny that it is quite large. Yet we may believe – or hope – that $\Pr[X \neq 0]$ is very close to zero. We can sometimes deduce $X > 0$ almost always if we have further information about $\mathrm{Var}[X]$.

Theorem 4.3.1

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{E[X]^2}.$$

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

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

■

We generally apply this result in asymptotic terms.

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

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

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

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

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

Suppose again $X = X_1 + \ldots + X_m$ 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 over ordered pairs)

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

Note that when $i \sim j$,

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

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

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

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

Let us say $X_1, \ldots, X_m$ are *symmetric* if for every $i \neq j$ there is an automorphism 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

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

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

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

where i is any fixed index. Then

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

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

4.4 RANDOM GRAPHS

The definition of the random graph $G(n,p)$ and of "threshold function" are given in Chapter 10, 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 $\omega(G) \geq 4$ 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

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

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 $\omega(G) \geq 4$ if and only if $X > 0$. Linearity of Expectation gives

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

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

Now suppose $p(n) \gg n^{-2/3}$ so that $E[X] \to \infty$ and consider the Δ^* 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 – i.e., 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 $\Pr[A_T | A_S] = p^5$. There are $O(n)$ sets T with $|S \cap T| = 3$ and for each of these $\Pr[A_T | A_S] = p^3$. Thus

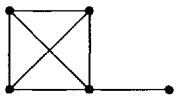
$$\Delta^* = O(n^2 p^5) + O(np^3) = o(n^4 p^6) = o(E[X])$$

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

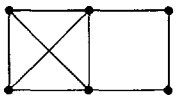
The proof of Theorem 4.4.1 appears to require a fortuitous calculation of Δ^*. The following definitions pave the way for the more general Theorem 4.4.2.

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 occuring 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

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

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

Now assume $p \gg n^{-v/e}$ so that $E[X] \to \infty$ and consider the Δ^* 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 – i.e., if and only if $|S \cap T| = i$ with $2 \leq i \leq v-1$. Let S be fixed. We split

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

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

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

and

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

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 α satisfy $v/e < \alpha < v_1/e_1$ 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 σ 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,

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

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

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

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 $O(1) = o(E[X])$ to Δ^*. 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 Δ^* is $o(E[X])$. Altogether $\Delta^* = o(E[X])$ 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 $E[X_{H'}] \to \infty$ for every H'. Then*

$$X_H \sim E[X_H]$$

almost always.

Proof. Let H have v vertices and e edges. As in Theorem 4.4.4 it suffices to show $\Delta^* = o(E[X])$. We split Δ^* 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,

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

to Δ^*. 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 $\omega(G) \geq k$.*

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

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

and so

$$\frac{\Delta^*}{E[X]} = \sum_{i=2}^{k-1} g(i)$$

where we set

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}.$$

Observe that $g(i)$ may be thought of as the probability that a randomly chosen T will intersect a fixed S in i points times the factor increase in $\Pr[A_T]$ when it does. Setting $i = 2$,

$$g(2) = 2\frac{\binom{k}{2}\binom{n-k}{k-2}}{\binom{n}{k}} \sim \frac{k^4}{n^2} \leq o(n^{-1}).$$

At the other extreme $i = k-1$,

$$g(k-1) = \frac{k(n-k)2^{-(k-1)}}{\binom{n}{k}2^{-\binom{k}{2}}} \sim \frac{2kn2^{-k}}{E[X]}.$$

As $k \sim 2\log_2 n$, the numerator is $n^{-1+o(1)}$. The denominator approaches infinity and so $g(k-1) \leq o(n^{-1})$. Some detailed calculation (which we omit) gives that the remaining $g(i)$ and their sum are also negligible so that Corollary 4.3.5 applies. ■

Theorem 4.5.1 leads to a strong concentration result for $\omega(G)$. For $k \sim 2\log_2 n$,

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1}2^{-k} = n^{-1+o(1)} = o(1).$$

Let $k_0 = k_0(n)$ be that value with $f(k_0) \geq 1 > f(k_0+1)$. For "most" n the function $f(k)$ will jump from a large $f(k_0)$ to a small $f(k_0+1)$. The probability that G contains a clique of size k_0+1 is at most $f(k_0+1)$ which will be very small.

When $f(k_0)$ is large Theorem 4.5.1 implies that G contains a clique of size k_0 with probability nearly 1. Together, with very high probability $\omega(G) = k_0$. For some n one of the values $f(k_0), f(k_0 + 1)$ may be of moderate size so this argument does not apply. Still one may show a strong concentration result found independently by Bollobás and Erdős (1976) and Matula (1976).

Corollary 4.5.2 *There exists $k = k(n)$ so that*

$$\Pr[\omega(G) = k \text{ or } k+1] \to 1.$$

We give yet stronger results on the distribution of $\omega(G)$ in Section 10.2.

4.6 DISTINCT SUMS

A set $x_1, \ldots, x_k$ of positive integers is said to have distinct sums if all sums

$$\sum_{i \in S} x_i, S \subseteq \{1, \ldots, k\}$$

are distinct. Let $f(n)$ denote the maximal k for which there exists a set

$$\{x_1, \ldots, x_k\} \subset \{1, \ldots, n\}$$

with distinct sums. The simplest example of a set with distinct sums is $\{2^i : i \leq \log_2 n\}$. This example shows

$$f(n) \geq 1 + \lfloor \log_2 n \rfloor.$$

Erdős offered \$300 for a proof or disproof that

$$f(n) \leq \log_2 n + C$$

for some constant C. From above, as all $2^{f(n)}$ sums are distinct and less than nk,

$$2^{f(n)} < nk = nf(n),$$

and so

$$f(n) < \log_2 n + \log_2 \log_2 n + O(1).$$

Examination of the second moment gives a modest improvement. Fix $\{x_1, \ldots, x_k\} \subset \{1, \ldots, n\}$ with distinct sums. Let $\epsilon_1, \ldots, \epsilon_k$ be independent with

$$\Pr[\epsilon_i = 1] = \Pr[\epsilon_i = 0] = \frac{1}{2}$$

and set

$$X = \epsilon_1 x_1 + \ldots + \epsilon_k x_k.$$

(We may think of X as a random sum.) Set

$$\mu = E[X] = \frac{x_1 + \ldots + x_k}{2}$$

and $\sigma^2 = \mathrm{Var}[X]$. We bound

$$\sigma^2 = \frac{x_1^2 + \ldots + x_k^2}{4} \leq \frac{n^2 k}{4}$$

so that $\sigma \leq n\sqrt{k}/2$. By Chebyschev's Inequality for any $\lambda > 1$,

$$\Pr[|X - \mu| \geq \lambda n\sqrt{k}/2] \leq \lambda^{-2}.$$

Reversing,

$$1 - \frac{1}{\lambda^2} \leq \Pr[|X - \mu| < \lambda n\sqrt{k}/2].$$

But X has any particular value with probability either zero or 2^{-k} since, critically, a sum can be achieved in at most one way. Thus

$$\Pr[|X - \mu| < \lambda n\sqrt{k}/2] \leq 2^{-k}(\lambda n\sqrt{k} + 1)$$

and

$$n \geq \frac{2^k(1 - \lambda^{-2}) - 1}{\sqrt{k}\lambda}.$$

While $\lambda = \sqrt{3}$ gives optimal results any choice of $\lambda > 1$ gives:

Theorem 4.6.1

$$f(n) \leq \log_2 n + \frac{1}{2}\log_2 \log_2 n + O(1).$$

4.7 THE RÖDL NIBBLE

For $2 \leq l < k < n$ let $M(n, k, l)$, the covering number, denote the minimal size of a family $\mathcal{K}$ of k-element subsets of $\{1, \ldots, n\}$ having the property that every l-element set is contained in at least one $A \in \mathcal{K}$. Clearly $M(n, k, l) \geq \binom{n}{l}/\binom{k}{l}$ since each k-set covers $\binom{k}{l}$ l-sets and every l-set must be covered. Equality holds if and only if the family $\mathcal{K}$ has the property that every l-set is contained in exactly one $A \in \mathcal{K}$. This is called an (n, k, l) tactical configuration (or block design). For example, $(n, 3, 2)$ tactical configurations are better known as Steiner Triple Systems. The question of the existence of tactical configurations is a central one for combinatorics but one for which probabilistic methods (at least so far!) play little role. In 1963 Paul Erdős and Haim Hanani conjectured that for fixed $2 \leq l < k$,

$$\lim_{n\to\infty} \frac{M(n, k, l)}{\binom{n}{l}/\binom{k}{l}} = 1.$$

Their conjecture was, roughly, that one can get asymptotically close to a tactical configuration. While this conjecture seemed ideal for a probabilistic analysis it was a full generation before Rödl (1985) found the proof, which we describe in this section. [One may similarly define the packing number $m(n,k,l)$ as the maximal size of a family $\mathcal{K}$ of k-element subsets of $\{1,\ldots,n\}$ having the property that every l-element set is contained in at most one $A \in \mathcal{K}$. Erdős and Hanani noticed from elementary arguments that

$$\lim_{n\to\infty} \frac{M(n,k,l)}{\binom{n}{l}/\binom{k}{l}} = 1 \Longleftrightarrow \lim_{n\to\infty} \frac{m(n,k,l)}{\binom{n}{l}/\binom{k}{l}} = 1.$$

While the Rödl result may be formulated in terms of either packing or covering here we deal only with the covering problem.]

Several researchers realized that the Rödl method applies in a much more general setting, dealing with covers in uniform hypergraphs. This was first observed by Frankl and Rödl, and has been simplified and extended by Pippenger and Spencer (1989) as well as by Kahn (1996). Our treatment here follows the one in Pippenger and Spencer (1989), and is based on the description of Füredi (1988), where the main tool is the second moment method.

For an r-uniform hypergraph $H = (V,E)$ and for a vertex $x \in V$, we let $d_H(x)$ [or simply $d(x)$, when there is no danger of confusion] denote the degree of x in H, that is, the number of edges containing x. Similarly, for $x,y \in V$, $d(x,y) = d_H(x,y)$ is the number of edges of H containing both x and y. A covering of H is a set of edges whose union contains all vertices. In what follows, whenever we write $\pm\delta$ we mean a quantity between $-\delta$ and δ. The following theorem is due to Pippenger, following Frankl and Rödl.

Theorem 4.7.1 *For every integer $r \geq 2$ and reals $k \geq 1$ and $a > 0$, there are $\gamma = \gamma(r,k,a) > 0$ and $d_0 = d_0(r,k,a)$ such that for every $n \geq D \geq d_0$ the following holds.*

Every r-uniform hypergraph $H = (V,E)$ on a set V of n vertices in which all vertices have positive degrees and which satisfies the following conditions:
(1) For all vertices $x \in V$ but at most γn of them, $d(x) = (1 \pm \gamma)D$,
(2) For all $x \in V$, $d(x) < kD$,
(3) For any two distinct $x,y \in V$, $d(x,y) < \gamma D$
contains a cover of at most $(1+a)\frac{n}{r}$ edges.

The basic idea in the proof is simple. Fixing a small $\epsilon > 0$ one shows that a random set of roughly $\epsilon n/r$ edges has, with high probability, only some $O(\epsilon^2 n)$ vertices covered more than once, and hence covers at least $\epsilon n - O(\epsilon^2 n)$ vertices. Moreover, after deleting the vertices covered, the induced hypergraph on the remaining vertices still satisfies the properties described in (1),(2) and (3) above (for some other values of n, γ, k and D). Therefore, one can choose again a random set of edges of this hypergraph, covering roughly an ϵ-fraction of its vertices with nearly no overlaps. Proceeding in this way for a large number of times we are finally left with at most ϵn uncovered vertices, and we then cover them trivially, by taking for each of them an

arbitrarily chosen edge containing it. Since ϵ is sufficiently small, although this last step is very inefficient, it can be tolerated.

The technical details require a careful application of the second moment method, used several times in the proof of the following lemma.

Lemma 4.7.2 *For every integer $r \geq 2$ and reals $K \geq 1$ and $\epsilon > 0$, and for every real $\delta' > 0$, there are $\delta = \delta(r, K, \epsilon, \delta') > 0$ and $D_0 = D_0(r, K, \epsilon, \delta')$ such that for every $n \geq D \geq D_0$ the following holds.*

Every r-uniform hypergraph $H = (V, E)$ on a set V of n vertices which satisfies the following conditions:
(i) For all vertices $x \in V$ but at most δn of them, $d(x) = (1 \pm \delta)D$,
(ii) For all $x \in V$, $d(x) < KD$,
(iii) For any two distinct $x, y \in V$, $d(x, y) < \delta D$
contains a set E' of edges with the following properties:
(iv) $|E'| = \frac{\epsilon n}{r}(1 \pm \delta')$,
(v) The set $V' = V - \cup_{e \in E'} e$ is of cardinality $|V'| = ne^{-\epsilon}(1 \pm \delta')$,
(vi) For all vertices $x \in V'$ but at most $\delta'|V'|$ of them, the degree $d'(x)$ of x in the induced hypergraph of H on V' satisfies $d'(x) = De^{-\epsilon(r-1)}(1 \pm \delta')$.

Proof. Throughout the proof we assume, whenever this is needed, that D (and hence n) are sufficiently large. We denote by $\delta_1, \delta_2, \ldots$ positive constants (that can be explicitly estimated) that tend to 0 when δ tends to 0 and D tends to infinity (for fixed r, K, ϵ). Therefore, by choosing δ and D_0 appropriately we can ensure that each of those will be smaller than δ'.

Let E' be a random subset of E obtained by picking, randomly and independently, each edge in E to be a member of E' with probability $p = \frac{\epsilon}{D}$. We have to show that with positive probability, the properties (iv), (v) and (vi) hold.

The proof that (iv) holds is easy. Note that by the assumptions H has at least $(1 - \delta)n$ vertices of degree at least $(1 - \delta)D$, showing that its number of edges is at least $\frac{(1-\delta)^2 nD}{r}$. Similarly, the number of edges of H does not exceed $\frac{(1+\delta)Dn + \delta nKD}{r}$. Therefore, $|E| = (1 \pm \delta_1)\frac{Dn}{r}$. It follows that the expected value of the size of E' satisfies $\mathbf{E}(|E'|) = |E|p = (1 \pm \delta_1)\frac{\epsilon n}{r}$ and its variance is $\mathrm{Var}(|E'|) = |E|p(1-p) \leq (1 \pm \delta_1)\frac{\epsilon n}{r}$. Therefore, by Chebyschev's Inequality, for an appropriately chosen $\delta_2 > 0$,

$$\Pr(|E'| = (1 \pm \delta_2)\frac{\epsilon n}{r}) > 0.99,$$

say, giving (iv).

To prove (v), define, for each vertex $x \in V$ an indicator random variable I_x, where $I_x = 1$ if $x \notin \cup_{e \in E'} e$ and $I_x = 0$ otherwise. Note that $|V'| = \sum_{x \in V} I_x$. Call a vertex $x \in V$ *good* if $d(x) = (1 \pm \delta)D$; otherwise call it *bad*. If x is good, then

$$\mathbf{E}(I_x) = \Pr(I_x = 1) = (1-p)^{d(x)} = \left(1 - \frac{\epsilon}{D}\right)^{(1 \pm \delta)D} = e^{-\epsilon}(1 \pm \delta_3).$$

If x is bad then, clearly, $0 \leq \mathbf{E}(I_x) \leq 1$. Since there are at most δn bad vertices it follows, by linearity of expectation, that the expected value of $|V'|$ is $ne^{-\epsilon}(1 \pm \delta_4)$.

To compute the variance of $|V'| = \sum_{x \in V} I_x$, note that

$$\begin{aligned} \mathrm{Var}(|V'|) &= \sum_{x \in V} \mathrm{Var}(I_x) + \sum_{x,y \in V, x \neq y} \mathrm{Cov}(I_x, I_y) \\ &\leq \mathbf{E}(|V'|) + \sum_{x,y \in V, x \neq y} \mathrm{Cov}(I_x, I_y). \end{aligned} \tag{4.1}$$

However,

$$\begin{aligned} \mathrm{Cov}(I_x, I_y) &= \mathbf{E}(I_x I_y) - \mathbf{E}(I_x)\mathbf{E}(I_y) \\ &= (1-p)^{d(x)+d(y)-d(x,y)} - (1-p)^{d(x)+d(y)} \\ &\leq (1-p)^{-d(x,y)} - 1 \leq (1 - \tfrac{\epsilon}{D})^{-\delta D} \leq \delta_5. \end{aligned}$$

It follows that

$$\mathrm{Var}(|V'|) \leq \mathbf{E}(|V'|) + \delta_5 n^2 \leq \delta_6 (\mathbf{E}(|V'|))^2,$$

which, by Chebyschev, implies that with probability at least 0.99

$$|V'| = (1 \pm \delta_7)\mathbf{E}(|V'|) = (1 \pm \delta_8) n e^{-\epsilon},$$

as claimed in (v).

It remains to prove (vi). To do so note, first, that all but at most $\delta_9 n$ vertices x satisfy the following two conditions:
(A) $d(x) = (1 \pm \delta)D$, and
(B) all but at most $\delta_{10} D$ edges $e \in E$ with $x \in e$ satisfy

$$|\{f \in E : x \notin f, f \cap e \neq \emptyset\}| = (1 \pm \delta_{11})(r-1)D. \tag{4.2}$$

Indeed, (A) holds for all but $\delta n < \delta_9 n/2$ vertices, by assumption. Moreover, the total number of edges containing vertices whose degrees are not $(1 \pm \delta)D$ is at most $\delta n K D$ and hence the number of vertices contained in more than $\delta_{10} D$ such edges is at most $\delta n K D r/(\delta_{10} D) \leq \delta_9 n/2$ for an appropriate choice of δ_9, δ_{10}. Note, next, that if $x \in e$ and e contains no vertex of degree which is not $(1 \pm \delta)D$ then, since $d(y,z) < \delta D$ for all y, z, the number of edges f not containing x that intersect e is at most $(r-1)(1 \pm \delta)D$ and at least $(r-1)(1 \pm \delta)D - \binom{r-1}{2}\delta D$,
and hence e satisfies (4.2).

It thus suffices to show that for most of the vertices x satisfying (A) and (B), $d'(x)$ satisfies (vi). Fix such a vertex x. Call an edge e with $x \in e$ *good* if it satisfies (4.2). Conditioning on $x \in V'$, the probability that a good edge containing x stays in the hypergraph on V' is $(1-p)^{(1 \pm \delta_{11})(r-1)D}$. Therefore, the expected value of $d'(x)$ is

$$\mathbf{E}(d'(x)) = (1 \pm \delta_{10} \pm \delta) D (1-p)^{(1 \pm \delta_{11})(r-1)D} \pm \delta_{10} D = e^{-\epsilon(r-1)} D (1 \pm \delta_{12}).$$

For each edge e containing x, let I_e denote the indicator random variable whose value is 1 iff e is contained in V'. Then, the degree $d'(x)$ is simply the sum of these indicator random variables, conditioned on $x \in V'$. It follows that

$$\begin{aligned}\mathrm{Var}(d'(x)) &\leq \mathbf{E}(d'(x)) + \sum_{x\in e, x\in f} \mathrm{Cov}(I_e, I_f) \\ &\leq \mathbf{E}(d'(x)) + 2\delta_{10}D^2(1\pm\delta) + \sum_{x\in e, x\in f, x, f\,good} \mathrm{Cov}(I_e, I_f) \quad (4.3)\end{aligned}$$

It remains to bound the sum $\sum_{x\in e, x\in f, x, f\text{good}} \mathrm{Cov}(I_e, I_f)$. For each fixed good e this sum is a sum of the form $\sum_{x\in f, f\text{good}} \mathrm{Cov}(I_e, I_f)$. There are at most $(r-1)\delta D$ edges f in the last sum for which $|e\cap f| > 1$, and their contribution to the sum cannot exceed $(r-1)\delta D$. If $e\cap f = \{x\}$ then let $t(e,f)$ denote the number of edges of H that intersect both e and f and do not contain x. Clearly, in this case, $t(e,f) \leq (r-1)^2\delta D$. It follows that for such e and f, $\mathrm{Cov}(I_e, I_f) \leq (1-p)^{-t(e,f)} - 1 \leq \delta_{13}$, implying that for each fixed good edge e,

$$\sum_{x\in f, f\text{good}} \mathrm{Cov}(I_e, I_f) \leq (r-1)\delta D + D(1+\delta)\delta_{13} \leq \delta_{14}D.$$

As the sum $\sum_{x\in e, x\in f, e, f\,good} \mathrm{Cov}(I_e, I_f)$ is the sum of at most $D(1+\delta)$ such quantities, we conclude that

$$\mathrm{Var}(d'(x)) \leq \mathbf{E}(d'(x)) + \delta_{15}D^2 \leq \delta_{16}(\mathbf{E}(d'(x))^2.$$

It thus follows, by Chebyschev, that with probability at most δ_{17}, $d'(x)$ is not $(1\pm\delta_{18})De^{-\epsilon(r-1)}$, and therefore, by Markov, that with probability at least, say, 0.99, for all but at most $\delta_{19}n$ vertices, $d'(x) = (1\pm\delta_{18})De^{-\epsilon(r-1)}$. This completes the proof of the lemma. ■

Proof [Theorem 4.7.1] Fix $\epsilon > 0$ such that

$$\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon < 1 + a,$$

and fix $1/10 > \delta > 0$ such that

$$(1+4\delta)\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon < 1 + a.$$

Fix an integer t so that $e^{-\epsilon t} < \epsilon$. The theorem is proved by applying the lemma t times. Put $\delta = \delta_t$ and then define, by reverse induction $\delta_t > \delta_{t-1} > \cdots > \delta_0$ such that $\delta_i \leq \delta_{i+1}e^{-\epsilon(r-1)}$, $\prod_{i=0}^{t}(1+\delta_i) < 1 + 2\delta$, and for $n \geq D \geq R_i$ one can apply the lemma with r, $K = ke^{\epsilon i(r-1)}$, ϵ, $\delta' = \delta_{i+1}$ and $\delta = \delta_i$. This will give the assertion of the theorem with $\gamma = \delta_0$, $d_0 = \max R_i$. Indeed, by applying the lemma repeatedly we obtain a decreasing sequence of sets of vertices $V = V_0, V_1, \ldots, V_t$, each contained in the previous one, and a sequence of sets of edges $E_1, E_2, \ldots, E_t$, where E_i is the set of edges E' obtained in the application of the lemma to the hypergraph induced on V_{i-1}. Here

$$|V_i| = |V_{i-1}|e^{-\epsilon}(1 \pm \delta_i) \; (\; = |V_0|e^{-i\epsilon}(1 \pm 2\delta) \;),$$
$$|E_i| = \frac{\epsilon|V_{i-1}|}{r}(1 \pm \delta_i) \leq (1 + 4\delta)\frac{\epsilon n}{r}e^{-(i-1)\epsilon},$$

and

$$D_i = D_{i-1}e^{-\epsilon(r-1)} = De^{-\epsilon i(r-1)}.$$

By covering each vertex of V_t separately by an edge containing it we conclude that the total number of edges in the cover obtained is at most

$$\begin{aligned}(1 + 4\delta)\sum_{i=0}^{t-1}\frac{\epsilon n}{r}e^{-i\epsilon} + |V_t| \; &\leq (1 + 4\delta)\frac{\epsilon n}{r}\frac{1}{1 - e^{-\epsilon}} + (1 + 2\delta)ne^{-\epsilon t} \\ &\leq \frac{n}{r}[(1 + 4\delta)(\frac{\epsilon}{1 - e^{-\epsilon}} + r\epsilon)] < (1 + a)\frac{n}{r}.\end{aligned}$$

This completes the proof. ■

We conclude the section by showing how the theorem quickly implies Rödl solution of the Erdős-Hanani problem mentioned in the beginning of the section.

Theorem 4.7.3 (Rödl) *For k, l fixed,*

$$M(n, k, l) \leq (1 + o(1))\binom{n}{l} \Big/ \binom{k}{l}$$

where the $o(1)$ term tends to zero as n tends to infinity.

Proof. Put $r = \binom{k}{l}$ and let H be the r-uniform hypergraph whose vertices are all l-subsets of $\{1, 2, \ldots, n\}$, and whose edges are all collections of $\binom{k}{l}$ l-tuples that lie in a k-set. H has $\binom{n}{l}$ vertices, each of its vertices has degree $D = \binom{n-l}{k-l}$, and every two distinct vertices lie in at most $\binom{n-l-1}{k-l-1} = o(D)$ common edges. Therefore, by Theorem 4.7.1, H has a cover of size at most $(1 + o(1))\binom{n}{l} \Big/ \binom{k}{l}$, as needed. ■

4.8 EXERCISES

1. Let X be a random variable taking integral nonnegative values, let $E(X^2)$ denote the expectation of its square, and let $\mathrm{Var}(X)$ denote its variance. Prove that
$$\Pr(X = 0) \leq \frac{\mathrm{Var}(X)}{E(X^2)}.$$

2. (*) Show that there is a positive constant c such that the following holds. For any n reals $a_1, a_2, \ldots, a_n$ satisfying $\sum_{i=1}^{n} a_i^2 = 1$, if $(\epsilon_1, \ldots, \epsilon_n)$ is a $\{-1, 1\}$-random vector obtained by choosing each ϵ_i randomly and independently with

uniform distribution to be either -1 or 1, then

$$\Pr(|\sum_{i=1}^{n} \epsilon_i a_i| \leq 1) \geq c.$$

3. (*) Show that there is a positive constant c such that the following holds. For any n vectors $a_1, a_2, \ldots, a_n \in R^2$ satisfying $\sum_{i=1}^{n} ||a_i||^2 = 1$ and $||a_i|| \leq 1/10$, where $||\cdot||$ denotes the usual Euclidean norm, if $(\epsilon_1, \ldots, \epsilon_n)$ is a $\{-1, 1\}$-random vector obtained by choosing each ϵ_i randomly and independently with uniform distribution to be either -1 or 1, then

$$\Pr\left(\left|\left|\sum_{i=1}^{n} \epsilon_i a_i\right|\right| \leq 1/3\right) \geq c.$$

4. Let X be a random variable with expectation $E(X) = 0$ and variance σ^2. Prove that for all $\lambda > 0$,

$$\Pr[X \geq \lambda] \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

5. Let $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n)$ be n two-dimensional vectors, where each x_i and each y_i is an integer whose absolute value does not exceed $\frac{2^{n/2}}{100\sqrt{n}}$. Show that there are two disjoint sets $I, J \subset \{1, 2, \ldots, n\}$ such that

$$\sum_{i \in I} v_i = \sum_{j \in J} v_j.$$

6. (*) Prove that for every set X of at least $4k^2$ distinct residue classes modulo a prime p, there is an integer a such that the set $\{ax \pmod p : x \in X\}$ intersects every interval in $\{0, 1, \ldots, p-1\}$ of length at least p/k.

THE PROBABILISTIC LENS:
Hamiltonian Paths

What is the maximum possible number of directed Hamilton paths in a tournament on n vertices? Denote this number by $P(n)$. The first application of the probabilistic method in Combinatorics is the result of Szele (1943) described in Chapter 2 which states that $P(n) \geq n!/2^{n-1}$. This bound follows immediately from the observation that the right-hand side is the expected number of such paths in a random tournament on n vertices. In the same paper Szele shows that

$$\frac{\frac{1}{2} \leq}{\lim, n \to \infty \left(\frac{P(n)}{n!}\right)^{1/n} \leq \frac{1}{2^{3/4}},}$$

proves that this limit does exist, and conjectures that its correct value is $1/2$.

This conjecture is proved in Alon (1990a). The proof is given below. The main tool is the Brégman proof of the Minc Conjecture for the permanent of a $(0,1)$-matrix, described in the Probabilistic Lens; Brégman Theorem (following Chapter 2).

Theorem 1 *There exists a positive constant c such that for every n,*

$$P(n) \leq cn^{3/2} \frac{n!}{2^{n-1}}.$$

Proof. For a tournament T, denote by $P(T)$ the number of directed Hamilton paths of T. Similarly, $C(T)$ denotes the number of directed Hamilton cycles of T, and $F(T)$ denotes the number of spanning subgraphs of T in which the indegree and the outdegree of every vertex is exactly 1. Clearly,

$$C(T) \leq F(T). \tag{1}$$

If $T = (V, E)$ is a tournament on a set $V = \{1, 2, \ldots, n\}$ of n vertices, the *adjacency matrix* of T is the n by n $(0,1)$-matrix $A_T = (a_{ij})$ defined by $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. Let r_i denote the number of ones in row i. Clearly,

$$\sum_{i}^{n} r_i = \binom{n}{2}. \tag{2}$$

By interpreting combinatorially the terms in the expansion of the permanent $\mathrm{per}(A_T)$, it follows that

$$\mathrm{per}(A_T) = F(T). \tag{3}$$

We need the following technical lemma.

Lemma 2 *For every two integers a, b satisfying $b \geq a + 2 > a \geq 1$ the inequality*

$$(a!)^{1/a} \cdot (b!)^{1/b} < ((a+1)!)^{1/(a+1)} \cdot ((b-1)!)^{1/(b-1)}$$

holds.

Proof. The assertion is simply that $f(a) < f(b-1)$, for the function f defined by $f(a) = (a!)^{1/a}/((a+1)!)^{1/(a+1)}$. Thus, it suffices to show that for every integer $x \geq 2$, $f(x-1) < f(x)$. Substituting the expression for f and raising both sides to the power $x(x-1)(x+1)$ it follows that it suffices to show that for all $x \geq 2$,

$$((x-1)!)^{x(x+1)} \cdot ((x+1)!)^{x(x-1)} < (x!)^{2(x^2-1)},$$

i.e.,

$$\left(\frac{x^x}{x!}\right)^2 > \left(\frac{x+1}{x}\right)^{x(x-1)}.$$

This is certainly true for $x = 2$. For $x \geq 3$ it follows from the facts that $4^x > e^{x+1}$, that $x! < (\frac{x+1}{2})^x$ and that $e^{x-1} > (\frac{x+1}{x})^{x(x-1)}$. ■

Corollary 3 *Define $g(x) = (x!)^{1/x}$. For every integer $S \geq n$ the maximum of the function $\prod_{i=1}^{n} g(x_i)$ subject to the constraints $\sum_{i=1}^{n} x_i = S$ and $x_i \geq 1$ are integers, is obtained iff the variables x_i are as equal as possible (i.e., iff each x_i is either $\lfloor S/n \rfloor$ or $\lceil S/n \rceil$.)*

Proof. If there are two indices i and j such that $x_i \geq x_j + 2$ then, by Lemma 2, the value of the product would increase once we add one to x_j and subtract one from x_i. ■

Returning to our tournament T we observe that the numbers r_i defined above are precisely the outdegrees of the vertices of T. If at least one of these is 0, then clearly $C(T) = F(T) = 0$. Otherwise, by Brégman's Theorem, by Corollary 3 and by (2) and (3), $F(T)$ is at most the value of the function $\prod_{i=1}^{n} (r_i!)^{1/r_i}$, where the integral

variables r_i satisfy (2) and are as equal as possible. By a straightforward (though somewhat tedious) derivation of the asymptotics using Stirling's formula this gives:

Proposition 4 *For every tournament T on n vertices,*

$$C(T) \leq F(T) \leq (1+o(1))\frac{\sqrt{\pi}}{\sqrt{2}e} n^{3/2} \frac{(n-1)!}{2^n}.$$

To complete the proof of the theorem, we have to derive a bound for the number of Hamiltonian paths in a tournament from the above result. Given a tournament S on n vertices, let T be the random tournament obtained from S by adding to it a new vertex y and by orienting each edge connecting y with one of the vertices of S, randomly and independently. For every fixed Hamiltonian path in S, the probability that it can be extended to a Hamiltonian cycle in T is precisely $1/4$. Thus, the expected number of Hamiltonian cycles in T is $\frac{1}{4}P(S)$ and hence there is a specific T for which $C(T) \geq \frac{1}{4}P(S)$. However, by Proposition 4, $C(T) \leq (1+o(1))\frac{\sqrt{\pi}}{\sqrt{2}e}(n+1)^{3/2}\frac{n!}{2^{n+1}}$, and thus $P(S) \leq O(n^{3/2}\frac{n!}{2^{n-1}})$, completing the proof of Theorem 1. ■

5

The Local Lemma

It's a thing that non-mathematicians don't realize. Mathematics is actually an esthetic subject almost entirely.
– John Conway

5.1 THE LEMMA

In a typical probabilistic proof of a combinatorial result, one usually has to show that the probability of a certain event is positive. However, many of these proofs actually give more and show that the probability of the event considered is not only positive but is large. In fact, most probabilistic proofs deal with events that hold with high probability, i.e., a probability that tends to 1 as the dimensions of the problem grow. For example, consider the proof given in Chapter 1 that for each $k \geq 1$ there are tournaments in which for every set of k players there is one who beats them all. The proof actually shows that for every fixed k if the number n of players is sufficiently large then almost all tournaments with n players satisfy this property; i.e., the probability that a random tournament with n players has the desired property tends to 1 as n tends to infinity.

On the other hand, there is a trivial case in which one can show that a certain event holds with positive, though very small, probability. Indeed, if we have n mutually independent events and each of them holds with probability at least $p > 0$, then the probability that all events hold simultaneously is at least p^n, which is positive, although it may be exponentially small in n.

It is natural to expect that the case of mutual independence can be generalized to that of rare dependencies, and provide a more general way of proving that certain

events hold with positive, though small, proability. Such a generalization is, indeed, possible, and is stated in the following lemma, known as the Lovász Local Lemma. This simple lemma, first proved in Erdős and Lovász (1975) is an extremely powerful tool, as it supplies a way for dealing with rare events.

Lemma 5.1.1 [The Local Lemma; General Case] *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. A directed graph $D = (V, E)$ on the set of vertices $V = \{1, 2, \ldots, n\}$ is called a dependency digraph for the events $A_1, \ldots, A_n$ if for each i, $1 \leq i \leq n$, the event A_i is mutually independent of all the events $\{A_j : (i,j) \notin E\}$. Suppose that $D = (V, E)$ is a dependency digraph for the above events and suppose there are real numbers $x_1, \ldots, x_n$ such that $0 \leq x_i < 1$ and $\Pr(A_i) \leq x_i \prod_{(i,j)\in E}(1-x_j)$ for all $1 \leq i \leq n$. Then $\Pr\left(\bigwedge_{i=1}^{n} \overline{A}_i\right) \geq \prod_{i=1}^{n}(1-x_i)$. In particular, with positive probability no event A_i holds.*

Proof. We first prove, by induction on s, that for any $S \subset \{1, \ldots, n\}$, $|S| = s < n$ and any $i \notin S$,

$$\Pr\left(A_i \Big| \bigwedge_{j\in S} \overline{A}_j\right) \leq x_i \,. \tag{5.1}$$

This is certainly true for $s = 0$. Assuming it holds for all $s' < s$, we prove it for S. Put $S_1 = \{j \in S; (i,j) \in E\}, S_2 = S \backslash S_1$. Then

$$\Pr\left(A_i \Big| \bigwedge_{j\in S} \overline{A}_j\right) = \frac{\Pr\left(A_i \wedge (\bigwedge_{j\in S_1} \overline{A}_j) \Big| \bigwedge_{\ell\in S_2} \overline{A}_\ell\right)}{\Pr\left(\bigwedge_{j\in S_1} \overline{A}_j \Big| \bigwedge_{\ell\in S_2} \overline{A}_\ell\right)} \,. \tag{5.2}$$

To bound the numerator observe that since A_i is mutually independent of the events $\{A_\ell : \ell \in S_2\}$,

$$\begin{aligned} \Pr\left(A_i \wedge \left(\bigwedge_{j\in S_1} \overline{A}_j\right) \Big| \bigwedge_{\ell\in S_2} \overline{A}_\ell\right) &\leq \Pr\left(A_i \Big| \bigwedge_{\ell\in S_2} \overline{A}_\ell\right) \\ &= \Pr(A_i) \leq x_i \prod_{(i,j)\in E}(1-x_j) \,. \end{aligned} \tag{5.3}$$

The denominator, on the other hand, can be bounded by the induction hypothesis. Indeed, suppose $S_1 = \{j_1, j_2, \ldots, j_r\}$. If $r = 0$ then the denominator is 1, and (5.1) follows. Otherwise

$$\begin{aligned} &\Pr\left(\overline{A}_{j_1} \wedge \overline{A}_{j_2} \wedge \ldots \wedge \overline{A}_{j_r} \Big| \bigwedge_{\ell\in S_2} \overline{A}_\ell\right) \\ &= \left(1 - \Pr(A_{j_1} \Big| \bigwedge_{\ell\in S_2} \overline{A}_\ell)\right) \cdot \left(1 - \Pr\left(A_{j_2} \Big| \overline{A}_{j_1} \wedge \bigwedge_{\ell\in S_2} \overline{A}_\ell\right)\right) \cdot \ldots \\ &\quad \ldots \cdot \left(1 - \Pr\left(A_{j_r} \Big| \overline{A}_{j_1} \wedge \ldots \wedge \overline{A}_{j_{r-1}} \wedge \bigwedge_{\ell\in S_2} \overline{A}_\ell\right)\right) \\ &\geq (1-x_{j_1})(1-x_{j_2}) \ldots (1-x_{j_r}) \geq \prod_{(i,j)\in E}(1-x_j). \end{aligned} \tag{5.4}$$

Substituting (5.3) and (5.4) into (5.2) we conclude that $\Pr\left(A_i \Big| \bigwedge_{j\in S} \overline{A}_j\right) \leq x_i$, completing the proof of the induction.

The assertion of Lemma 5.1.1 now follows easily, as

$$\Pr\left(\bigwedge_{i=1}^{n}\overline{A}_i\right) = (1-\Pr(A_1))\cdot(1-\Pr(A_2|\overline{A}_1))\cdot\ldots$$
$$\ldots\cdot(1-\Pr(A_n|\bigwedge_{i=1}^{n-1}\overline{A}_i) \geq \prod_{i=1}^{n}(1-x_i) \quad , \qquad (5.5)$$

completing the proof. ■

Corollary 5.1.2 [The Local Lemma; Symmetric Case] *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all the other events A_j but at most d, and that $\Pr(A_i) \leq p$ for all $1 \leq i \leq n$. If*

$$ep(d+1) \leq 1 \qquad (5.6)$$

then $\Pr\left(\bigwedge_{i=1}^{n}\overline{A}_i\right) > 0$.

Proof. If $d = 0$ the result is trivial. Otherwise, by the assumption there is a dependency digraph $D = (V, E)$ for the events $A_1, \ldots, A_n$ in which for each i, $\left|\{j : (i,j) \in E\}\right| \leq d$. The result now follows from Lemma 5.1.1 by taking $x_i = 1/(d+1)(< 1)$ for all i and using the fact that for any $d \geq 1$, $\left(1-\frac{1}{d+1}\right)^d > 1/e$. ■

It is worth noting that as shown by Shearer in 1985, the constant "e" is the best possible constant in inequality (5.6). Note also that the proof of Lemma 5.1.1 indicates that the conclusion remains true even when we replace the two assumptions that each A_i is mutually independent of $\{A_j : (i,j) \notin E\}$ and that $\Pr(A_i) \leq x_i \prod_{(ij)\in E}(1-x_j)$ by the weaker assumption that for each i and each $S_2 \subset \{1,\ldots,n\}\backslash\{j : (i,j) \in E\}$, $\Pr\left(A_i|\bigwedge_{j\in S_2}\overline{A}_j\right) \leq x_i \prod_{(i,j)\in E}(1-x_j)$. This turns out to be useful in certain applications.

In the next few sections we present various applications of the Local Lemma for obtaining combinatorial results. There is no known proof of any of these results, which do not use the Local Lemma.

5.2 PROPERTY B AND MULTICOLORED SETS OF REAL NUMBERS

Recall that a hypergraph $H = (V, E)$ has property B, (i.e. is two-colorable), if there is a coloring of V by two colors so that no edge $f \in E$ is monochromatic.

Theorem 5.2.1 *Let $H = (V, E)$ be a hypergraph in which every edge has at least k elements, and suppose that each edge of H intersects at most d other edges. If $e(d+1) \leq 2^{k-1}$ then H has property B.*

Proof. Color each vertex v of H, randomly and independently, either blue or red (with equal probability). For each edge $f \in E$, let A_f be the event that f is monochromatic.

Clearly $\Pr(A_f) = 2/2^{|f|} \leq 1/2^{k-1}$. Moreover, each event A_f is clearly mutually independent of all the other events $A_{f'}$ for all edges f' that do not intersect f. The result now follows from Corollary 5.1.2. ■

A special case of Theorem 5.2.1 is that for any $k \geq 9$, any k-uniform k-regular hypergraph H has property B. Indeed, since any edge f of such an H contains k vertices, each of which is incident with k edges (including f), it follows that f intersects at most $d = k(k-1)$ other edges. The desired result follows, since $e(k(k-1)+1) < 2^{k-1}$ for each $k \geq 9$.

The next result we consider, which appeared in the original paper of Erdős and Lovász, deals with k-colorings of the real numbers. For a k-coloring $c : \mathbb{R} \to \{1, 2, \ldots, k\}$ of the real numbers by the k colors $1, 2, \ldots, k$, and for a subset $T \subset \mathbb{R}$, we say that T is *multicolored* (with respect to c) if $c(T) = \{1, 2, \ldots, k\}$, i.e., if T contains elements of all colors.

Theorem 5.2.2 *Let m and k be two positive integers satisfying*

$$e\,(m(m-1)+1)\,k\left(1-\frac{1}{k}\right)^m \leq 1\,. \tag{5.7}$$

Then, for any set S of m real numbers there is a k-coloring so that each translation $x + S$ (for $x \in \mathbb{R}$) is multicolored.

Notice that (5.7) holds whenever $m > (3 + o(1))k \log k$.

Proof. We first fix a *finite* subset $X \subseteq \mathbb{R}$ and show the existence of a k-coloring so that each translation $x + S$ (for $x \in X$) is multicolored. This is an easy consequence of the Local Lemma. Indeed, put $Y = \bigcup_{x \in X}(x + S)$ and let $c : Y \to \{1, 2, \ldots, k\}$ be a random k-coloring of Y obtained by choosing, for each $y \in Y$, randomly and independently, $c(y) \in \{1, 2, \ldots, k\}$ according to a uniform distribution on $\{1, 2, \ldots, k\}$. For each $x \in X$, let A_x be the event that $x + S$ is not multicolored (with respect to c). Clearly $\Pr(A_x) \leq k\left(1 - \frac{1}{k}\right)^m$. Moreover, each event A_x is mutually independent of all the other events $A_{x'}$ but those for which $(x + S) \cap (x' + S) \neq \emptyset$. As there are at most $m(m-1)$ such events, the desired result follows from Corollary 5.1.2.

We can now prove the existence of a coloring of the set of all reals with the desired properties, by a standard compactness argument. Since the discrete space with k points is (trivially) compact, Tikhonov's Theorem (which is equivalent to the axiom of choice) implies that an arbitrary product of such spaces is compact. In particular, the space of all functions from $\mathbb{R}$ to $\{1, 2, \ldots, k\}$, with the usual product topology, is compact. In this space for every fixed $x \in \mathbb{R}$, the set C_x of all colorings c, such that $x + S$ is multicolored, is closed. (In fact, it is both open and closed, since a basis to the open sets is the set of all colorings whose values are prescribed in a finite number of places). As we proved above, the intersection of any finite number of sets C_x is nonempty. It thus follows, by compactness, that the intersection of all sets C_x is nonempty. Any coloring in this intersection has the properties in the conclusion of Theorem 5.2.2. ■

Note that it is impossible, in general, to apply the Local Lemma to an infinite number of events and conclude that in some point of the probability space none of them holds. In fact, there are trivial examples of countably many mutually independent events A_i, satisfying $\Pr(A_i) = 1/2$ and $\bigwedge_{i\geq 1} \overline{A}_i = \emptyset$. Thus the compactness argument is essential in the above proof.

5.3 LOWER BOUNDS FOR RAMSEY NUMBERS

The deriviation of lower bounds for Ramsey numbers by Erdős in 1947 was one of the first applications of the probabilistic method. The Local Lemma provides a simple way of improving these bounds. Let us obtain, first, a lower bound for the diagonal Ramsey number $R(k,k)$. Consider a random two-coloring of the edges of K_n. For each set S of k vertices of K_n, let A_S be the event that the complete graph on S is monochromatic. Clearly $\Pr(A_S) = 2^{1-\binom{k}{2}}$. It is obvious that each event A_s is mutually independent of all the events A_T, but those which satisfy $|S \cap T| \geq 2$, since this is the only case in which the corresponding complete graphs share an edge. We can therefore apply Corollary 5.1.2 with $p = 2^{1-\binom{k}{2}}$ and $d = \binom{k}{2}\binom{n}{k-2}$ to conclude:

Proposition 5.3.1 *If* $e\left(\binom{k}{2}\binom{n}{k-2} + 1\right) \cdot 2^{1-\binom{k}{2}} < 1$ *then* $R(k,k) > n$.

A short computation shows that this gives $R(k,k) > \frac{\sqrt{2}}{e}(1+o(1))\, k2^{k/2}$, only a factor 2 improvement on the bound obtained by the straightforward probabilistic method. Although this minor improvement is somewhat disappointing it is certainly not surprising; the Local Lemma is most powerful when the dependencies between events are rare, and this is not the case here. Indeed, there is a total number of $K = \binom{n}{k}$ events considered, and the maximum outdegree d in the dependency digraph is roughly $\binom{k}{2}\binom{n}{k-2}$. For large k and much larger n (which is the case of interest for us) we have $d > K^{1-O(1/k)}$, i.e., quite a lot of dependencies. On the other hand, if we consider small sets S, e.g., sets of size 3, we observe that out of the total $K = \binom{n}{3}$ of them each shares an edge with only $3(n-3) \approx K^{1/3}$. This suggests that the Local Lemma may be much more significant in improving the off-diagonal Ramsey numbers $R(k,\ell)$, especially if one of the parameters, say ℓ, is small. Let us consider, for example, following Spencer (1977), the Ramsey number $R(k,3)$. Here, of course, we have to apply the nonsymmetric form of the Local Lemma. Let us two-color the edges of K_n randomly and independently, where each edge is colored blue with probability p. For each set of three vertices T, let A_T be the event that the triangle on T is blue. Similarly, for each set of k vertices S, let B_S be the event that the complete graph on S is red. Clearly $\Pr(A_T) = p^3$ and $\Pr(B_S) = (1-p)^{\binom{k}{2}}$. Construct a dependency digraph for the events A_T and B_S by joining two vertices by edges (in both directions) iff the corresponding complete graphs share an edge. Clearly, each A_T-node of the dependency graph is adjacent to $3(n-3) < 3n$ $A_{T'}$-nodes and to at most $\binom{n}{k}$ $B_{S'}$-nodes. Similarly, each B_S-node is adjacent to $\binom{k}{2}(n-k) < k^2 n/2$ A_T nodes and to at most $\binom{n}{k}$ $B_{S'}$-nodes. It

follows from the general case of the Local Lemma (Lemma 5.1.1) that if we can find a $0 < p < 1$ and two real numbers $0 \leq x < 1$ and $0 \leq y < 1$ such that

$$p^3 \leq x(1-x)^{3n}(1-y)^{\binom{n}{k}}$$

and

$$(1-p)^{\binom{k}{2}} \leq y(1-x)^{k^2n/2}(1-y)^{\binom{n}{k}}$$

then $R(k,3) > n$.

Our objective is to find the largest possible $k = k(n)$ for which there is such a choice of p, x and y. An elementary (but tedious) computation shows that the best choice is when $p = c_1 n^{-1/2}$, $k = c_2 n^{1/2} \log n, x = c_3/n^{3/2}$ and $y = \frac{c_4}{e^{n^{1/2}\log^2 n}}$. This gives that $R(k,3) > c_5 k^2/\log^2 k$. A similar argument gives that $R(k,4) > k^{5/2+o(1)}$. In both cases the amount of computation required is considerable. However, the hard work does pay; the bound $R(k,3) > c_5 k^2/\log^2 k$ matches a lower bound of Erdős proved in 1961 by a highly complicated probabilistic argument. This was improved to $R(k,3) > c_6 k^2/\log k$ by Kim (1995). The bound above for $R(k,4)$ is better than any bound for $R(k,4)$ known to be proven without the Local Lemma.

5.4 A GEOMETRIC RESULT

A family of open unit balls F in the three-dimensional Euclidean space $\mathbb{R}^3$ is called a *k-fold covering* of $\mathbb{R}^3$ if any point $x \in \mathbb{R}^3$ belongs to at least k balls. In particular, a 1-fold covering is simply called a *covering*. A k-fold covering $\mathcal{F}$ is called *decomposable* if there is a partition of $\mathcal{F}$ into two pairwise disjoint families $\mathcal{F}_1$ and $\mathcal{F}_2$, each being a covering of $\mathbb{R}^3$. Mani-Levitska and Pach (1988) constructed, for any integer $k \geq 1$, a nondecomposable k-fold covering of $\mathbb{R}^3$ by open unit balls. On the other hand they proved that any k-fold covering of $\mathbb{R}^3$ in which no point is covered by more than $c2^{k/3}$ balls is decomposable. This reveals a somewhat surprising phenomenon: that it is more difficult to decompose coverings that cover some of the points of $\mathbb{R}^3$ too often than to decompose coverings that cover every point about the same number of times. The exact statement of the Mani--Pach Theorem is the following.

Theorem 5.4.1 *Let $\mathcal{F} = \{B_i\}_{i\in I}$ be a k-fold covering of the three-dimensional Euclidean space by open unit balls. Suppose, further, that no point of $\mathbb{R}^3$ is contained in more than t members of $\mathcal{F}$. If*

$$e \cdot t^3 2^{18}/2^{k-1} \leq 1$$

then $\mathcal{F}$ is decomposable.

Proof. Define an infinite hypergraph $H = (V(H), E(H))$ as follows. The set of vertices of H, $V(H)$, is simply $\mathcal{F} = \{B_i\}_{i\in I}$. For each $x \in \mathbb{R}^3$ let E_x be the set of

balls $B_i \in \mathcal{F}$ which contain x. The set of edges of H, $E(H)$, is simply the set of E_x, with the understanding that when $E_x = E_y$ the edge is taken only once. We claim each edge E_x intersects less than $t^3 2^{18}$ other edges E_y of H. If $x \in B_i$ the center of B_i is within distance 1 of x. If now $B_j \cap B_i \neq \emptyset$ the center of B_j is within distance three of x and so B_j lies entirely inside the ball of radius four centered at x. Such a B_j covers precisely $4^{-3} = 2^{-6}$ of the volume of that ball. As no vertex is covered more than t times there can be at most $2^6 t$ such balls. It is not too difficult to check that m balls in $\mathbb{R}^3$ cut $\mathbb{R}^3$ into less than m^3 connected components so that there are at most $(2^6 t)^3$ distinct E_y overlaping E_x.

Consider, now, any finite subhypergraph L of H. Each edge of L has at least k vertices, and it intersects at most $d < t^3 2^{18}$ other edges of L. Since, by assumption, $e(d+1) \leq 2^{k-1}$, Theorem 5.2.1 (which is a simple corollary of the local lemma), implies that L is two-colorable. This means that one can color the vertices of L blue and red so that no edge of L is monochromatic. Since this holds for any finite L, a compactness argument, analogous to the one used in the proof of Theorem 5.2.2, shows that H is two-colorable. Given a two-coloring of H with no monochromatic edges, we simply let $\mathcal{F}_1$ be the set of all blue balls, and $\mathcal{F}_2$ be the set of all red ones. Clearly, each $\mathcal{F}_i$ is a covering of $\mathbb{R}^3$, completing the proof of the theorem. ■

It is worth noting that Theorem 5.4.1 can be easily generalized to higher dimensions. We omit the detailed statement of this generalization.

5.5 THE LINEAR ARBORICITY OF GRAPHS

A *linear forest* is a forest (i.e., an acyclic simple graph) in which every connected component is a path. The *linear arboricity* $\mathrm{la}(G)$ of a graph G is the minimum number of linear forests in G, whose union is the set of all edges of G. This notion was introduced by Harary as one of the covering invariants of graphs. The following conjecture, known as the *linear arboricity conjecture*, was raised in Akiyama, Exoo and Harary (1981).

Conjecture 5.5.1 [The linear arboricity conjecture] *The linear arboricity of every d-regular graph is $\lceil (d+1)/2 \rceil$.*

Notice that since every d-regular graph G on n vertices has $nd/2$ edges, and every linear forest in it has at most $n-1$ edges, the inequality

$$\mathrm{la}(G) \geq \frac{nd}{2(n-1)} > \frac{d}{2}$$

is immediate. Since $\mathrm{la}(G)$ is an integer this gives $\mathrm{la}(G) \geq \lceil (d+1)/2 \rceil$. The difficulty in Conjecture 5.5.1 lies in proving the converse inequality: $\mathrm{la}(G) \leq \lceil (d+1)/2 \rceil$. Note also that since every graph G with maximum degree Δ is a subgraph of a Δ-regular graph (which may have more vertices, as well as more edges than G), the linear arboricity conjecture is equivalent to the statement that the linear arboricity of every graph G with maximum degree Δ is at most $\lceil (\Delta+1)/2 \rceil$.

Although this conjecture received a considerable amount of attention, the best general result concerning it, proved without any probabilistic arguments, is that $\mathrm{la}(G) \leq \lceil 3\Delta/5 \rceil$ for even Δ and that $\mathrm{la}(G) \leq \lceil (3\Delta+2)/5 \rceil$ for odd Δ. In this section we prove that for every $\varepsilon > 0$ there is a $\Delta_0 = \Delta_0(\varepsilon)$ such that for every $\Delta \geq \Delta_0$, the linear arboricity of every graph with maximum degree Δ is less than $\left(\frac{1}{2} + \varepsilon\right)\Delta$. This result (with a somewhat more complicated proof) appears in Alon (1988) and its proof relies heavily on the local lemma. We note that this proof is more complicated than the other proofs given in this chapter, and requires certain preparations, some of which are of independent interest.

It is convenient to deduce the result for undirected graphs from its directed version. A *d-regular digraph* is a directed graph in which the indegree and the outdegree of every vertex is precisely d. A linear directed forest is a directed graph in which every connected component is a directed path. The *dilinear arboricity* $\mathrm{dla}(G)$ of a directed graph G is the minimum number of linear directed forests in G whose union covers all edges of G. The directed version of the Linear Arboricity Conjecture, first stated in Nakayama and Peroche (1987) is:

Conjecture 5.5.2 *For every d-regular digraph D,*

$$\mathrm{dla}(D) = d + 1 .$$

Note that since the edges of any (connected) undirected $2d$-regular graph G can be oriented along an Euler cycle, so that the resulting oriented digraph is d-regular, the validity of Conjecture 5.5.2 for d implies that of Conjecture 5.5.1 for $2d$.

It is easy to prove that any graph with n vertices and maximum degree d contains an independent set of size at least $n/(d+1)$. The following proposition shows that at the price of decreasing the size of such a set by a constant factor we can guarantee that it has a certain structure.

Proposition 5.5.3 *Let $H = (V, E)$ be a graph with maximum degree d, and let $V = V_1 \cup V_2 \cup \ldots \cup V_r$ be a partition of V into r pairwise disjoint sets. Suppose each set V_i is of cardinality $|V_i| \geq 2ed$, where e is the basis of the natural logarithm. Then there is an independent set of vertices $W \subseteq V$, that contains a vertex from each V_i.*

Proof. Clearly we may assume that each set V_i is of cardinality precisely $g = \lceil 2ed \rceil$ (otherwise, simply replace each V_i by a subset of cardinality g of it, and replace H by its induced subgraph on the union of these r new sets). Let us pick from each set V_i randomly and independently a single vertex according to a uniform distribution. Let W be the random set of the vertices picked. To complete the proof we show that with positive probability W is an independent set of vertices in H.

For each edge f of H, let A_f be the event that W contains both ends of f. Clearly, $\Pr(A_f) \leq 1/g^2$. Moreover, if the endpoints of f are in V_i and in V_j, then the event A_f is mutually independent of all the events corresponding to edges whose endpoints do not lie in $V_i \cup V_j$. Therefore, there is a dependency digraph for the events in

which the maximum degree is less than $2gd$, and since $e \cdot 2gd \cdot 1/g^2 = 2ed/g < 1$ we conclude, by Corollary 5.1.2, that with positive probability none of the events A_f holds. But this means that W is an independent set containing a vertex from each V_i, completing the proof. ■

Proposition 5.5.3 suffices to prove Conjecture 5.5.2 for digraphs with no short directed cycle. Recall that the directed girth of a digraph is the minimum length of a directed cycle in it.

Theorem 5.5.4 *Let $G = (U, F)$ be a d-regular digraph with directed girth $g \geq 8ed$. Then*

$$\mathrm{dla}(G) = d + 1 .$$

Proof. As is well known, F can be partitioned into d pairwise disjoint 1-regular spanning subgraphs $F_1, \ldots, F_d$ of G. [This is an easy consequence of the Hall-König Theorem; let H be the bipartite graph whose two classes of vertices A and B are copies of U, in which $u \in A$ is joined to $v \in B$ iff $(u, v) \in F$. Since H is d-regular its edges can be decomposed into d perfect matchings, which correspond to d 1-regular spanning subgraphs of G.] Each F_i is a union of vertex disjoint directed cycles $C_{i1}, C_{i2}, \ldots, C_{ir_i}$. Let $V_1, V_2, \ldots, V_r$ be the sets of *edges* of all the cycles $\{C_{ij} : 1 \leq i \leq d, 1 \leq j \leq r_i\}$. Clearly $V_1, V_2, \ldots, V_r$ is a partition of the set F of all edges of G, and by the girth condition, $|V_i| \geq g \geq 8ed$ for all $1 \leq i \leq r$. Let H be the line graph of G, i.e., the graph whose set of vertices is the set F of edges of G in which two edges are adjacent iff they share a common vertex in G. Clearly H is $4d - 2$ regular. As the cardinality of each V_i is at least $8ed \geq 2e(4d - 2)$, there is, by Proposition 5.5.3 , an independent set of H containing a member from each V_i. But this means that there is a matching M in G, containing at least one edge from each cycle C_{ij} of the 1-factors $F_1, \ldots, F_d$. Therefore $M, F_1 \backslash M, F_2 \backslash M, \ldots, F_d \backslash M$ are $d + 1$-directed forests in G (one of which is a matching) that cover all its edges. Hence

$$\mathrm{dla}(G) \leq d + 1 .$$

As G has $|U| \cdot d$ edges and each directed linear forest can have at most $|U| - 1$ edges,

$$\mathrm{dla}(G) \geq |U|d/(|U| - 1) > d .$$

Thus $\mathrm{dla}(G) = d + 1$, completing the proof. ■

The last theorem shows that the assertion of Conjecture 5.5.2 holds for digraphs with sufficiently large (directed) girth. In order to deal with digraphs with small girth, we show that most of the edges of each regular digraph can be decomposed into a relatively small number of almost regular digraphs with high girth. To do this, we need the following statement, which is proved using the local lemma.

Lemma 5.5.5 *Let $G = (V, E)$ be a d-regular directed graph, where d is sufficiently large, and let p be an integer satisfying $10\sqrt{d} \leq p \leq 20\sqrt{d}$. Then, there is a p-coloring of the vertices of G by the colors $0, 1, 2, \ldots, p-1$ with the following property; for each vertex $v \in V$ and each color i, the numbers $N^+(v, i) = \left|\{u \in V; (v, u) \in E\right.$*

and u is colored $i\}$ $\big|$ *and* $N^-(v,i) = \big|\{u \in V : (u,v) \in E$ *and u is colored $i\}\big|$ satisfy,*

$$\begin{aligned} \left|N^+(v,i) - \tfrac{d}{p}\right| &\leq 3\sqrt{d/p}\sqrt{\log d}\,, \\ \left|N^-(v,i) - \tfrac{d}{p}\right| &\leq 3\sqrt{d/p}\sqrt{\log d}\,. \end{aligned} \tag{5.8}$$

Proof. Let $f : V \to \{0, 1, \ldots, p-1\}$ be a random vertex coloring of V by p colors, where for each $v \in V$, $f(v) \in \{0, 1, \ldots, p-1\}$ is chosen according to a uniform distribution. For every vertex $v \in V$ and every color $i, 0 \leq i < p$, let $A^+_{v,i}$ be the event that the number $N^+(v,i)$ of neighbors of v in G whose color is i does not satisfy inequality (5.8). Clearly, $N^+(v,i)$ is a Binomial random variable with expectation $\frac{d}{p}$ and standard deviation $\sqrt{\frac{d}{p}(1-\frac{1}{p})} < \sqrt{\frac{d}{p}}$. Hence, by the standard estimates for Binomial distribution given in Appendix A, for every $v \in V$ and $0 \leq i < p$,

$$\Pr(A^+_{v,i}) < 1/d^4\,.$$

Similarly, if $A^-_{v,i}$ is the event that the number $N^-(v,i)$ violates (5.8) then

$$\Pr(A^-_{v,i}) < 1/d^4\,.$$

Clearly, each of the events $A^+_{v,i}$ or $A^-_{v,i}$ is mutually independent of all the events $A^+_{u,j}$ or $A^-_{u,j}$ for all vertices $u \in V$ that do not have a common neighbor with v in G. Therefore, there is a dependency digraph for all our events with maximum degree $\leq (2d)^2 \cdot p$. Since $e \cdot \frac{1}{d^4}((2d)^2 p + 1) < 1$, Corollary 5.1.2 (i.e., the symmetric form of the Local Lemma), implies that with positive probability no event $A^+_{v,i}$ or $A^-_{v,i}$ occurs. Hence, there is a coloring f which satisfies (5.8) for all $v \in V$ and $0 \leq i < p$, completing the proof. ■

We are now ready to deal with general regular digraphs. Let $G = (V, E)$ be an arbitrary d-regular digraph. Throughout the argument we assume, whenever it is needed, that d is sufficiently large. Let p be a prime satisfying $10d^{1/2} \leq p \leq 20d^{1/2}$ (it is well known that for every n there is a prime between n and $2n$). By Lemma 5.5.5 there is a vertex coloring $f : V \to \{0, 1, \ldots, p-1\}$ satisfying (5.8). For each i, $0 \leq i < p$, let $G_i = (V, E_i)$ be the spanning subdigraph of G defined by $E_i = \{(u,v) \in E : f(v) \equiv (f(u) + i) \bmod p\}$. By inequality (5.8) the maximum indegree Δ^-_i and the maximum outdegree Δ^+_i in each G_i is at most $\frac{d}{p} + 3\sqrt{\frac{d}{p}}\sqrt{\log d}$. Moreover, for each $i > 0$, the length of every directed cycle in G_i is divisible by p. Thus, the directed girth g_i of G_i is at least p. Since each G_i can be completed, by adding vertices and edges, to a Δ_i-regular digraph with the same girth g_i and with $\Delta_i = \max(\Delta^+_i, \Delta^-_i)$, and since $g_i > 8e\Delta_i$ (for all sufficiently large d), we conclude, by Theorem 5.5.4, that $\mathrm{dla}(G_i) \leq \Delta_i + 1 \leq \frac{d}{p} + 3\sqrt{\frac{d}{p}}\sqrt{\log d} + 1$ for all $1 \leq i < p$. For G_0, we only apply the trivial inequality

$$\mathrm{dla}(G_0) \leq 2\Delta_0 \leq 2\frac{d}{p} + 6\sqrt{\frac{d}{p}}\sqrt{\log d}$$

obtained by, e.g., embedding G_0 as a subgraph of a Δ_0-regular graph, splitting the edges of this graph into Δ_0 1-regular spanning subgraphs, and breaking each of these 1-regular spanning subgraphs into two linear directed forests. The last two inequalities, together with the fact that $10\sqrt{d} \leq p \leq 20\sqrt{d}$ imply

$$\mathrm{dla}(G) \leq d + 2\frac{d}{p} + 3\sqrt{pd}\sqrt{\log d} + 3\sqrt{\frac{d}{p}}\sqrt{\log d} + p - 1 \leq d + c \cdot d^{3/4}(\log d)^{1/2}.$$

We have thus proved:

Theorem 5.5.6 *There is an absolute constant $c > 0$ such that for every d-regular digrapn G*

$$\mathrm{dla}(G) \leq d + cd^{3/4}(\log d)^{1/2} .$$

We note that by being a little more careful, we can improve the error term to $c'd^{2/3}(\log d)^{1/3}$. Since the edges of any undirected $d = 2f$-regular graph can be oriented so that the resulting digraph is f-regular, and since any $(2f - 1)$-regular undirected graph is a subgraph of a $2f$-regular graph the last theorem implies:

Theorem 5.5.7 *There is an absolute constant $c > 0$ such that for every undirected d-regular graph G*

$$\mathrm{la}(G) \leq \frac{d}{2} + cd^{3/4}(\log d)^{1/2} .$$

5.6 LATIN TRANSVERSALS

Following the proof of the local lemma we noted that the mutual independency assumption in this lemma can be replaced by the weaker assumption that the conditional probability of each event, given the mutual nonoccurance of an arbitrary set of events, each nonadjacent to it in the dependency digraph, is sufficiently small. In this section we describe an application, from Erdős and Spencer (1991), of this modified version of the lemma. Let $A = (a_{ij})$ be an n of n matrix with, say, integer entries. A permutation π is called a *Latin transversal* (of A) if the entries $a_{i\pi(i)}$ $(1 \leq i \leq n)$ are all distinct.

Theorem 5.6.1 *Suppose $k \leq (n-1)/(4e)$ and suppose that no integer appears in more than k entries of A. Then A has a Latin Transversal.*

Proof. Let π be a random permutation of $\{1, 2, \ldots, n\}$, chosen according to a uniform distribution among all possible $n!$ permutations. Denote by T the set of all ordered four-tuples (i, j, i', j') satisfying $i < i', j \neq j'$ and $a_{ij} = a_{i'j'}$. For each $(i, j, i', j') \in T$, let $A_{iji'j'}$ denote the event that $\pi(i) = j$ and $\pi(i') = j'$. The existence of a Latin transversal is equivalent to the statement that with positive probability none of these events hold. Let us define a symmetric digraph, (i.e., a

graph) G on the vertex set T by making (i,j,i',j') adjacent to (p,q,p',q') if and only if $\{i,i'\} \cap \{p,p'\} \neq \emptyset$ or $\{j,j'\} \cap \{q,q'\} \neq \emptyset$. Thus, these two four-tuples are not adjacent iff the four cells $(i,j),(i',j'),(p,q)$ and (p',q') occupy four distinct rows and columns of A. The maximum degree of G is less than $4nk$; indeed, for a given $(i,j,i',j') \in T$ there are at most $4n$ choices of (s,t) with either $s \in \{i,i'\}$ or $t \in \{j,j'\}$, and for each of these choices of (s,t) there are less than k choices for $(s',t') \neq (s,t)$ with $a_{st} = a_{s't'}$. Each such four-tuple (s,t,s',t') can be uniquely represented as (p,q,p',q') with $p < p'$. Since $e \cdot 4nk \cdot \frac{1}{n(n-1)} \leq 1$, the desired result follows from the above mentioned strengthening of the symmetric version of the Local Lemma, if we can show that

$$\Pr(A_{iji'j'} \mid \bigwedge_S \overline{A}_{pqp'q'}) \leq 1/n(n-1) \tag{5.9}$$

for any $(i,j,i',j') \in T$ and any set S of members of T which are nonadjacent in G to (i,j,i',j'). By symmetry, we may assume that $i = j = 1, i' = j' = 2$ and that hence none of the p's nor q's are either 1 or 2. Let us call a permutation π *good* if it satisfies $\bigwedge_S \overline{A}_{pqp'q'}$, and let S_{ij} denote the set of all good permutations π satisfying $\pi(1) = i$ and $\pi(2) = j$. We claim that $|S_{12}| \leq |S_{ij}|$ for all $i \neq j$. Indeed, suppose first that $i,j > 2$. For each good $\pi \in S_{12}$ define a permutation π^* as follows. Suppose $\pi(x) = i$, $\pi(y) = j$. Then define $\pi^*(1) = i, \pi^*(2) = j, \pi^*(x) = 1, \pi^*(y) = 2$ and $\pi^*(t) = \pi(t)$ for all $t \neq 1,2,x,y$. One can easily check that π^* is good, since the cells $(1,i),(2,j),(x,1),(y,2)$ are not part of any $(p,q,p',q') \in S$. Thus $\pi^* \in S_{ij}$, and since the mapping $\pi \to \pi^*$ is injective $|S_{12}| \leq |S_{ij}|$, as claimed. Similarly one can define injective mappings showing that $|S_{12}| \leq |S_{ij}|$ even when $\{i,j\} \cap \{1,2\} \neq \emptyset$. It follows that $\Pr(A_{1122} \wedge \bigwedge_S \overline{A}_{pqp'q'}) \leq \Pr(A_{1i2j} \wedge \bigwedge_S \overline{A}_{pqp'q'})$ for all $i \neq j$ and hence that $\Pr(A_{1122} \mid \bigwedge_S \overline{A}_{pqp'q'}) \leq 1/n(n-1)$. By symmetry, this implies (5.9) and completes the proof. ■

5.7 THE ALGORITHMIC ASPECT

When the probabilistic method is applied to prove that a certain event holds with high probability, it often supplies an efficient deterministic, or at least randomized, algorithm for the corresponding problem.

By applying the Local Lemma we often manage to prove that a given event holds with positive probability, although this probability may be exponentially small in the dimensions of the problem. Consequently, it is not clear if any of these proofs can provide polynomial algorithms for the corresponding algorithmic problems. For many years there was no known method of converting the proofs of any of the examples discussed in this chapter into an efficient algorithm. In 1991 J. Beck found such a method that works for some of these examples, with a little loss in the constants.

He demonstrated in Beck (1991) his method by considering the problem of hypergraph two-coloring. For simplicity we only describe here the case of fixed edge-size in which each edge intersects a fixed number of other edges.

Let n, d be fixed positive integers. By the (n, d)-problem we mean the following: Given sets $A_1, \ldots, A_N \subseteq \Omega$ with all $|A_i| = n$, such that no set A_i intersects more than d other sets A_j, find a two-coloring of Ω so that no A_i is monochromatic. When $e(d+1) < 2^{n-1}$, Theorem 5.2.1 assures us that this problem always does have a solution. Can we find the coloring in polynomial (in N for fixed n, d) time? Beck has given an affirmative answer under somewhat more restrictive assumptions. We assume Ω is of the form $\Omega = \{1, \ldots, m\}$, $m \leq Nn$ and the initial data structure consists of a list of the elements of the sets A_i and a list giving for each element j those i for which $j \in A_i$. We let G denote the dependency graph with vertices the sets A_i and A_i, A_j adjacent if they overlap.

Theorem 5.7.1 *Let n, d be such that, setting $D = d(d-1)^3$ there exists a decomposition $n = n_1 + n_2 + n_3$ with*

$$\begin{aligned} 16D(1+d) &< 2^{n_1}, \\ 16D(1+d) &< 2^{n_2}, \\ 2e(1+d) &< 2^{n_3}. \end{aligned}$$

Then there is a randomized algorithm with expected running time $O(N(\ln N)^c)$ for the (n, d) problem, where c is a constant (depending only on n and d).

For $\epsilon < 1/11$, fixed, we note that the above conditions are satisfied, for n sufficiently large, when $d < 2^{n\epsilon}$ by taking $n_1 = n_2 \sim 5n/11$ and $n_3 \sim n/11$. We emphasize again that the algorithmic analysis here is for *fixed* n, d and N approaching infinity, although the argument can be extended to the nonfixed case as well.

Beck has given a deterministic algorithm for the (n, d) problem. The randomized algorithm we give may be derandomized using the techniques of Chapter 15. The running time remains polynomial but seemingly no longer $N^{1+o(1)}$. Moreover, the algorithm can even be parallelized using some of the techniques in Chapter 15 together with a certain modification in the algorithm.

Proof. The First Pass. During this pass, points will be either red, blue, uncolored or saved. We move through the points $j \in \Omega$ sequentially, coloring them red or blue at random, flipping a fair coin. After each j is colored we check all $A_i \ni j$. If A_i now has n_1 points in one color and no points in the other color we call A_i *dangerous*. All uncolored $k \in A_i$ are now considered saved. When saved points k are reached in the sequential coloring they are not colored but simply skipped over. At the conclusion of the First Pass points are red, blue or saved. We say a set A_i *survives* if it does not have both red and blue points. Let $S \subseteq G$ denote the (random) set of surviving sets.

Claim 5.7.2 *Almost surely all components C of $G|_S$ have size $O(\ln N)$.*

Proof. An $A_i \in S$ may be dangerous or, possibly, many of its points were saved because neighboring (in G) sets were dangerous. The probability of a particular A_i

becoming dangerous is at most 2^{1-n_1} since for this to occur the first n_1 coin flips determining colors of $j \in A_i$ must come up the same. (We only have inequality since in addition n_1 points of A_i must be reached before being saved.) Let V be an independent set in G; i.e., the $A_i \in V$ are mutually disjoint. Then the probability that all $A_i \in V$ become dangerous is at most $(2^{1-n_1})^{|V|}$ as the coin flips involve disjoint sets. Now let $V \subseteq G$ be such that all distances between the $A_i \in V$ are at least 4, distance being the length of the shortest path in G. We claim that

$$\Pr[V \subseteq S] \leq (d+1)^{|V|}(2^{1-n_1})^{|V|}.$$

This is because for each $A_i \in V$ there are at most $d+1$ choices for a dangerous neighbor $A_{i'}$, giving $(d+1)^{|V|}$ choices for the $A_{i'}$. As the A_i are at least four apart the $A_{i'}$ cannot be adjacent and so the probability that they are all dangerous is at most $(2^{1-n_1})^{|V|}$, as claimed.

Call $T \subseteq G$ a 4-*tree* if the $A_i \in T$ are such that all their mutual distances in G are at least four and so that, drawing an arc between $A_i, A_j \in T$ if their distance is precisely four, the resulting graph is connected. We first bound the number of 4-trees of size u. The "distance-four" graph defined on T must contain a tree. There are less than 4^j trees (up to isomorphism) on j vertices, now fix one. We can label the tree $1, \ldots, u$ so that each $j > 1$ is adjacent to some $i < j$. Now consider the number of $(A^1, \ldots, A^u)$ whose distance-four graph corresponds to this tree. There are N choices for A^1. Having chosen A^i for all $i < j$ the set A^j must be at distance four from A^i in G and there are at most D such points. Hence the number of 4-trees of size u is at most $4^u N D^{u-1} < N(4D)^u$. For any particular 4-tree T we have already that $\Pr[T \subseteq S] \leq [(d+1)2^{1-n_1}]^u$. Hence the expected number of 4-trees $T \subseteq S$ is at most

$$N\left[8D(d+1)2^{-n_1}\right]^u.$$

As the bracketed term is less than $1/2$ by assumption, for $u = c_1 \ln N$ this term is $o(1)$. Thus almost surely $G|_S$ will contain no 4-tree of size bigger than $c_1 \ln N$. We actually want to bound the size of the components C of $G|_S$. A maximal 4-tree T in a component C must have the property that every $A_i \in C$ lies within three of an $A_j \in T$. There are less than d^3 (a constant) A_i within three of any given A_j so that $c_1 \ln N \geq |T| \geq |C|d^{-3}$ and so (since d is a constant),

$$|C| \leq c_2 \ln N$$

proving the Claim. ■

If the First Pass leaves components of size larger than $c_2 \ln N$ we simply repeat the entire procedure. In expected *linear* time the First Pass is successful. The points that are red or blue are now fixed. The sets A_i with both red and blue points can now be ignored. For each surviving A_i fix a subset B_i of $n - n_1$ saved points. It now suffices to color the saved points so that no B_i is monochromatic. The B_i split into components of size $O(\ln N)$ and it suffices to color each component separately. On the Second Pass we apply the method of the First Pass to each component of the B_i. Now we call a set B_i dangerous if it receives n_2 points of one color and none of the

other. The Second Pass takes expected time $O(M)$ to color a component of size M, hence an expected time $O(N)$ to color all the components. (For success we require that a component of size M is broken into components of size at most $c_2 \ln M$. To avoid trivialities, if $M < \ln \ln N$ we skip the Second Pass for the corresponding component.) At the end of the Second Pass (still in linear time!) there is a family of twice surviving sets $C_i \subset B_i \subset A_i$ of size n_3, the largest component of which has size $O(\ln \ln N)$.

We still need to color these $O(N)$ components of sets of size n_3, each component of size $O(\ln \ln N)$. By the Local Lemma (or directly by Theorem 5.2.1), each of these components can be two-colored. *We now find the two-coloring by brute force!* Examining all two-colorings of a component of size M takes time $O(M2^{nM})$ which is $O((\ln N)^c)$ in our case. Doing this for all components takes time $O(N(\ln N)^c)$. This completes the coloring. ■

We note that with slightly more restrictions on n, d, a Third Pass could be made and then the total time would be $O(N(\ln \ln N)^c)$. We note also that a similar technique can be applied for converting several other applications of the Local Lemma into efficient algorithms.

5.8 EXERCISES

1. (*) Prove that for every integer $d > 1$ there is a finite $c(d)$ such that the edges of any bipartite graph with maximum degree d in which every cycle has at least $c(d)$ edges can be colored by $d + 1$ colors so that there are no two adjacent edges with the same color and there is no two-colored cycle.

2. (*) Prove that for every $\epsilon > 0$ there is a finite $l_0 = l_0(\epsilon)$ and an infinite sequence of bits $a_1, a_2, a_3, \ldots \ a_i \in \{0, 1\}$, such that for every $l > l_0$ and every $i \geq 1$ the two binary vectors $u = (a_i, a_{i+1}, \ldots, a_{i+l-1})$ and $v = (a_{i+l}, a_{i+l+1}, \ldots, a_{i+2l-1})$ differ in at least $(\frac{1}{2} - \epsilon)l$ coordinates.

3. Let $G = (V, E)$ be a simple graph and suppose each $v \in V$ is associated with a set $S(v)$ of colors of size at least $10d$, where $d \geq 1$. Suppose, in addition, that for each $v \in V$ and $c \in S(v)$ there are at most d neighbors u of v such that c lies in $S(u)$. Prove that there is a proper coloring of G assigning to each vertex v a color from its class $S(v)$.

4. Let $G = (V, E)$ be a cycle of length $4n$ and let $V = V_1 \cup V_2 \ldots \cup V_n$ be a partition of its $4n$ vertices into n pairwise disjoint subsets, each of cardinality 4. Is it true that there must be an independent set of G containing precisely one vertex from each V_i ? (Prove, or supply a counter-example).

5. (*) Prove that there is an absolute constant $c > 0$ such that for every k there is a set S_k of at least $ck \ln k$ integers, such that for every coloring of the integers by k colors there is an integer x for which the set $x + S$ does not intersect all color classes.

THE PROBABILISTIC LENS:
Directed Cycles

Let $D = (V, E)$ be a simple directed graph with minimum outdegree δ and maximum indegree Δ.

Theorem 1 [Alon and Linial (1989)] *If* $e(\Delta\delta + 1)\left(1 - \frac{1}{k}\right)^{\delta} < 1$ *then* D *contains a (directed, simple) cycle of length* $0 (\mathrm{mod}\, k)$.

Proof. Clearly we may assume that every outdegree is precisely δ, since otherwise we can consider a subgraph of D with this property.

Let $f : V \to \{0, 1, \ldots, k-1\}$ be a random coloring of V, obtained by choosing, for each $v \in V, f(v) \in \{0, \ldots, k-1\}$ independently, according to a uniform distribution. For each $v \in V$, let A_v denote the event that there is no $u \in V$, with $(v, u) \in E$ and $f(u) \equiv (f(v) + 1) (\mathrm{mod}\, k)$. Clearly $\Pr(A_v) = \left(1 - \frac{1}{k}\right)^{\delta}$. One can easily check that each event A_v is mutually independent of all the events A_u but those satisfying

$$N^+(v) \cap \left(u \bigcup N^+(u)\right) \neq \emptyset\,,$$

where here $N^+(v) = \{w \in V : (v, w) \in E\}$. The number of such u's is at most $\Delta\delta$ and hence, by our assumption and by the Local Lemma, (Corollary 5.1.2), $\Pr\left(\bigwedge_{v \in V} \overline{A}_v\right) > 0$. Therefore, there is an $f : V \to \{0, 1, \ldots, k-1\}$ such that for every $v \in V$ there is a $u \in V$ with

$$(v, u) \in E \quad \text{and} \quad f(u) \equiv (f(v) + 1)\,(\mathrm{mod}\, k). \tag{1}$$

Starting at an arbitrary $v = v_0 \in V$ and applying (1) repeatedly we obtain a sequence $v_0, v_1, v_2, \ldots$ of vertices of D so that $(v_i, v_{i+1}) \in E$ and $f(v_{i+1}) \equiv$

$(f(v_i) + 1) \pmod{k}$ for all $i \geq 0$. Let j be the minimum integer so that there is an $\ell < j$ with $v_\ell = v_j$. The cycle $v_\ell v_{\ell+1} v_{\ell+2} \cdots v_j = v_\ell$ is a directed simple cycle of D whose length is divisible by k. ■

This page intentionally left blank

6

Correlation Inequalities

You just keep right on thinking there, Butch, that's what you're good at.
– Robert Redford to Paul Newman in *Butch Cassidy and the Sundance Kid*

Let $G = (V, E)$ be a random graph on the set of vertices $V = \{1, 2, \ldots, n\}$ generated by choosing, for each $i, j \in V, i \neq j$ independently, the pair $\{i, j\}$ to be an edge with probability p, where $0 < p < 1$. Let H be the event that G is Hamiltonian and let P be the event that G is planar. Suppose one wants to compare the two quantities $\Pr(P \wedge H)$ and $\Pr(P) \cdot \Pr(H)$. Intuitively, knowing that G is Hamiltonian suggests that it has many edges and hence seems to indicate that G is less likely to be planar. Therefore it seems natural to expect that $\Pr(P|H) \leq \Pr(P)$ implying

$$\Pr(P \wedge H) \leq \Pr(H) \cdot \Pr(P) .$$

This inequality, which is, indeed, correct, is a special case of the FKG-Inequality of Fortuin, Kasteleyn and Ginibre (1971). In this chapter we present the proof of this inequality and several related results, which deal with the correlation between certain events in probability spaces. The proofs of all these results are rather simple, and still they supply many interesting consequences. The first inequality of this type is due to Harris (1960). A result closer to the ones considered here is a lemma of Kleitman (1966b), stating that if $\mathcal{A}$ and $\mathcal{B}$ are two *monotone decreasing* families of subsets of $\{1, 2, \ldots, n\}$ (i.e., $A \in \mathcal{A}$ and $A' \subseteq A \Rightarrow A' \in \mathcal{A}$ and, similarly $B \in \mathcal{B}$ and $B' \subseteq B \Rightarrow B' \in \mathcal{B}$) then

$$|\mathcal{A} \cap \mathcal{B}| \cdot 2^n \geq |\mathcal{A}| \cdot |\mathcal{B}| .$$

This lemma was followed by many extensions and generalizations until Ahlswede and Daykin (1978) obtained a very general result, which implies all these extensions.

In the next section we present this result and its proof. Some of its many applications are discussed in the rest of the chapter.

6.1 THE FOUR FUNCTIONS THEOREM OF AHLSWEDE AND DAYKIN

Suppose $n \geq 1$ and put $N = \{1, 2, \ldots, n\}$. Let $P(N)$ denote the set of all subsets of N, and let $\mathbb{R}^+$ denote the set of nonnegative real numbers. For a function $\varphi : P(N) \to \mathbb{R}^+$ and for a family $\mathcal{A}$ of subsets of N denote $\varphi(\mathcal{A}) = \sum_{A \in \mathcal{A}} \varphi(A)$. For two families $\mathcal{A}$ and $\mathcal{B}$ of subsets of N define $\mathcal{A} \cup \mathcal{B} = \{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\}$ and $\mathcal{A} \cap \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$.

Theorem 6.1.1 [The Four Functions Theorem] *Let $\alpha, \beta, \gamma, \delta : P(N) \to \mathbb{R}^+$ be four functions from the set of all subsets of N to the nonnegative reals. If, for every two subsets $A, B \subseteq N$ the inequality*

$$\alpha(A)\beta(B) \leq \gamma(A \cup B)\delta(A \cap B) \tag{6.1}$$

holds, then, for every two families of subsets $\mathcal{A}, \mathcal{B} \subseteq P(N)$,

$$\alpha(\mathcal{A})\beta(\mathcal{B}) \leq \gamma(\mathcal{A} \cup \mathcal{B})\delta(\mathcal{A} \cap \mathcal{B}). \tag{6.2}$$

Proof. Observe, first, that we may modify the four functions $\alpha, \beta, \gamma, \delta$ by defining $\alpha(A) = 0$ for all $A \notin \mathcal{A}$, $\beta(B) = 0$ for all $B \notin \mathcal{B}$, $\gamma(C) = 0$ for all $C \notin \mathcal{A} \cup \mathcal{B}$, and $\delta(D) = 0$ for all $D \notin \mathcal{A} \cap \mathcal{B}$. Clearly, (6.1) still holds for the modified functions and in inequality (6.2) we may assume now that $\mathcal{A} = \mathcal{B} = \mathcal{A} \cup \mathcal{B} = \mathcal{A} \cap \mathcal{B} = P(N)$.

To prove this inequality we apply induction on n. The only step that requires some computation is $n = 1$. In this case $P(N) = \{\phi, N\}$. For each function $\varphi \in \{\alpha, \beta, \gamma, \delta\}$ define $\varphi_0 = \varphi(\phi)$ and $\varphi_1 = \varphi(N)$. By (6.1) we have

$$\begin{aligned} \alpha_0\beta_0 &\leq \gamma_0\delta_0, \\ \alpha_0\beta_1 &\leq \gamma_1\delta_0, \\ \alpha_1\beta_0 &\leq \gamma_1\delta_0, \\ \alpha_1\beta_1 &\leq \gamma_1\delta_1. \end{aligned} \tag{6.3}$$

By the above paragraph we only have to prove inequality (6.2), where $\mathcal{A} = \mathcal{B} = P(N)$, i.e., to prove that

$$(\alpha_0 + \alpha_1)(\beta_0 + \beta_1) \leq (\gamma_0 + \gamma_1)(\delta_0 + \delta_1)\,. \tag{6.4}$$

If either $\gamma_1 = 0$ or $\delta_0 = 0$ this follows immediately from (6.3). Otherwise, by (6.3), $\gamma_0 \geq \frac{\alpha_0\beta_0}{\delta_0}$ and $\delta_1 \geq \frac{\alpha_1\beta_1}{\gamma_1}$. It thus suffices to show that $\left(\frac{\alpha_0\beta_0}{\delta_0} + \gamma_1\right)\left(\delta_0 + \frac{\alpha_1\beta_1}{\gamma_1}\right) \geq (\alpha_0 + \alpha_1)(\beta_0 + \beta_1)$, or, equivalently, that $(\alpha_0\beta_0 + \gamma_1\delta_0)(\delta_0\gamma_1 + \alpha_1\beta_1) \geq (\alpha_0 + \alpha_1)(\beta_0 + \beta_1)\delta_0\gamma_1$. The last inequality is equivalent to

$$(\gamma_1\delta_0 - \alpha_0\beta_1)(\gamma_1\delta_0 - \alpha_1\beta_0) \geq 0\,,$$

which follows from (6.3), as both factors in the left-hand side are nonnegative. This completes the proof for $n = 1$.

Suppose, now, that the theorem holds for $n-1$ and let us prove it for n, $(n \geq 2)$. Put $N' = N \backslash \{n\}$ and define for each $\varphi \in \{\alpha, \beta, \gamma, \delta\}$ and each $A \subseteq N'$, $\varphi'(A) = \varphi(A) + \varphi(A \cup \{n\})$. Clearly, for each function $\varphi \in \{\alpha, \beta, \gamma, \delta\}$ $\varphi'(P(N')) = \varphi(P(N))$. Therefore, the desired inequality (6.3) would follow from applying the induction hypothesis to the functions $\alpha', \beta', \gamma', \delta' : P(N') \to \mathbb{R}^+$. However, in order to apply this hypothesis we have to check that these new functions satisfy the assumption of Theorem 6.1.1 on N', i.e., that for every $A', B' \subseteq N'$,

$$\alpha'(A')\beta'(B') \leq \gamma'(A' \cup B')\delta'(A' \cap B') \,. \tag{6.5}$$

Not surprisingly, this last inequality follows easily from the case $n = 1$ which we have already proved. Indeed, let T be a 1-element set and define $\overline{\alpha}(\phi) = \alpha(A')$, $\overline{\alpha}(T) = \alpha(A' \cup \{n\})$, $\overline{\beta}(\phi) = \beta(B')$, $\overline{\beta}(T) = \beta(B' \cup \{n\})$, $\overline{\gamma}(\phi) = \gamma(A' \cup B')$, $\overline{\gamma}(T) = \gamma(A' \cup B' \cup \{n\})$ and $\overline{\delta}(\phi) = \delta(A' \cap B')$, $\overline{\delta}(T) = \delta((A' \cap B') \cup \{n\})$. By the assumption (6.1) $\overline{\alpha}(S)\overline{\beta}(R) \leq \overline{\gamma}(S \cup R)\overline{\delta}(S \cap R)$ for all $S, R \subseteq T$ and hence, by the case $n = 1$ already proved $\alpha'(A')\beta'(B') = \overline{\alpha}(P(T))\overline{\beta}(P(T)) \leq \overline{\gamma}(P(T))\overline{\delta}(P(T)) = \gamma'(A' \cup B')\delta'(A' \cap B')$, which is the desired inequality (6.5). Therefore inequality (6.2) holds, completing the proof. ■

The Ahlswede-Daykin Theorem can be extended to arbitrary finite distributive lattices. A *lattice* in a partially ordered set in which every two elements, x and y, have a unique minimal upper bound, denoted by $x \vee y$ and called the *join* of x and y and a unique maximal lower bound, denoted by $x \wedge y$ and called the *meet* of x and y. A lattice L is *distributive* if for all $x, y, z \in L$,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$$

or, equivalently if for all $x, y, z \in L$,

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z) \,.$$

For two sets $X, Y \subseteq L$ define

$$X \vee Y = \{x \vee y : \, x \in X, \, y \in Y\},$$

and

$$X \wedge Y = \{x \wedge y : \, x \in X, \, y \in Y\} \,.$$

Any subset L of $P(N)$, where $N = \{1, 2, \ldots, n\}$, ordered by inclusion, which is closed under the union and intersection operations is a distributive lattice. Here, the join of two members $A, B \in L$, is simply their union $A \cup B$ and their meet is the intersection $A \cap B$. It is somewhat more surprising (but easy to check) that every finite distributive lattice L is isomorphic to a sublattice of $P(\{1, 2, \ldots, n\})$ for some n. (To see this, call an element $x \in L$ *join-irreducible* if whenever $x = y \vee z$ then either $x = y$ or $x = z$. Let $x_1, x_2, \ldots, x_n$ be the set of all join-irreducible elements in L and associate each element $x \in L$ with the set $A = A(x) \subseteq N$, where $x = \bigvee_{i \in A} x_i$

and $\{x_i : i \in A\}$ are all the join-irreducibles y satisfying $y \leq x$. The mapping $x \to A(x)$ is the desired isomorphism.) This fact enables us to generalize Theorem 6.1.1 to arbitrary finite distributive lattices as follows.

Corollary 6.1.2 *Let L be a finite distributive lattice and let α, β, γ and δ be four functions from L to $\mathbb{R}^+$. If*

$$\alpha(x)\beta(y) \leq \gamma(x \vee y)\delta(x \wedge y)$$

for all $x, y \in L$ then for every $X, Y \subseteq L$,

$$\alpha(X)\beta(Y) \leq \gamma(X \vee Y)\delta(X \wedge Y) .$$

The simplest case in the last Corollary is the case where all the four functions α, β, γ and δ are identically 1, stated below.

Corollary 6.1.3 *Let L be a finite distributive lattice and suppose $X, Y \subseteq L$. Then*

$$|X| \cdot |Y| \leq |X \vee Y| \cdot |X \wedge Y| .$$

We close this section by presenting a very simple consequence of the last Corollary, first proved by Marica and Schonheim (1969).

Corollary 6.1.4 *Let X be a family of subsets of a finite set N and define*

$$X \backslash X = \{F \backslash F' : F,\ F' \in X\} .$$

Then $|X \backslash X| \geq |X|$.

Proof. Let L be the distributive lattice of all subsets of N. By applying Corollary 6.1.3 to X and $Y = \{N \backslash F : F \in X\}$ we obtain

$$|X|^2 = |X| \cdot |Y| \leq |X \cup Y| \cdot |X \cap Y| = |X \backslash X|^2 .$$

The desired result follows. ■

6.2 THE FKG INEQUALITY

A function $\mu : L \to \mathbb{R}^+$, where L is a finite distributive lattice, is called *log-supermodular* if

$$\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y)$$

for all $x, y \in L$. A function $f : L \to \mathbb{R}^+$ is *increasing* if $f(x) \leq f(y)$ whenever $x \leq y$ and is *decreasing* if $f(x) \geq f(y)$ whenever $x \leq y$.

Motivated by a problem from statistical mechanics, Fortuin et al. (1971) proved the following useful inequality which has become known as the FKG-inequality.

Theorem 6.2.1 [The FKG inequality] *Let L be a finite distributive lattice and let $\mu : L \to \mathbb{R}^+$ be a log-supermodular function. Then, for any two increasing functions $f, g : L \to \mathbb{R}^+$ we have*

$$\left(\sum_{x\in L}\mu(x)f(x)\right)\cdot\left(\sum_{x\in L}\mu(x)g(x)\right) \le \left(\sum_{x\in L}\mu(x)f(x)g(x)\right)\cdot\left(\sum_{x\in L}\mu(x)\right) . \tag{6.6}$$

Proof. Define four functions $\alpha, \beta, \gamma, \delta : L \to \mathbb{R}^+$ as follows. For each $x \in L$,

$$\begin{aligned} \alpha(x) &= \mu(x)f(x) , & \beta(x) &= \mu(x)g(x), \\ \gamma(x) &= \mu(x)f(x)g(x) , & \delta(x) &= \mu(x) . \end{aligned}$$

We claim that these functions satisfy the hypothesis of the Ahlswede-Daykin Theorem, stated in Corollary 6.1.2. Indeed, if $x, y \in L$ then, by the supermodularity of μ and since f and g are increasing,

$$\begin{aligned} \alpha(x)\beta(y) &= \mu(x)f(x)\mu(y)g(y) \le \mu(x\vee y)f(x)g(y)\mu(x\wedge y) \\ &\le \mu(x\vee y)f(x\vee y)g(x\vee y)\mu(x\wedge y) = \gamma(x\vee y)\delta(x\wedge y) . \end{aligned}$$

Therefore, by Corollary 6.1.2 (with $X = Y = L$),

$$\alpha(L)\beta(L) \le \gamma(L)\delta(L) ,$$

which is the desired result. ■

Note that the conclusion of Theorem 6.2.1 holds also if both f and g are decreasing (simply interchange γ and δ in the proof). In case f is increasing and g is decreasing (or vice versa) the opposite inequality holds:

$$\left(\sum_{x\in L}\mu(x)f(x)\right)\left(\sum_{x\in L}\mu(x)g(x)\right) \ge \left(\sum_{x\in L}\mu(x)f(x)g(x)\right)\left(\sum_{x\in L}\mu(x)\right) .$$

To prove it, simply apply Theorem 6.2.1 to the two increasing functions $f(x)$ and $k - g(x)$, where k is the constant $\max_{x\in L} g(x)$. (This constant is needed to guarantee that $k - g(x) \ge 0$ for all $x \in L$).

It is helpful to view μ as a measure on L. Assuming μ is not identically zero we can define, for any function $f : L \to \mathbb{R}^+$, its expectation,

$$< f >= \frac{\sum_{x\in L} f(x)\mu(x)}{\sum_{x\in L}\mu(x)} .$$

With this notation, the FKG-inequality asserts that if μ is log-supermodular and $f, g : L \to \mathbb{R}^+$ are both increasing or both decreasing then

$$< fg > \ge < f > \cdot < g > .$$

Similarly, if f is increasing and g is decreasing (or vice versa), then

$$< fg > \le < f > < g > .$$

This formulation demonstrates clearly the probabilistic nature of the inequality, some of whose many interesting consequences are presented in the rest of this chapter.

6.3 MONOTONE PROPERTIES

Recall that a family $\mathcal{A}$ of subsets of $N = \{1, 2, \ldots, n\}$ is *monotone decreasing* if $A \in \mathcal{A}$ and $A' \subseteq A \Rightarrow A' \in \mathcal{A}$. Similarly, it is *monotone increasing* if $A \in \mathcal{A}$ and $A \subseteq A' \Rightarrow A' \in \mathcal{A}$. By considering the power set $P(N)$ as a symmetric probability space, one naturally defines the *probability* of $\mathcal{A}$ by

$$\Pr(\mathcal{A}) = \frac{|\mathcal{A}|}{2^n} .$$

Thus, $\Pr(\mathcal{A})$ is simply the probability that a randomly chosen subset of N lies in $\mathcal{A}$.

Kleitman's Lemma, which was the starting point of all the correlation inequalities considered in this chapter, is the following.

Proposition 6.3.1 *Let $\mathcal{A}$ and $\mathcal{B}$ be two monotone increasing families of subsets of $N = \{1, 2, \ldots, n\}$ and let $\mathcal{C}$ and $\mathcal{D}$ be two monotone decreasing families of subsets of N. Then*

$$\Pr(\mathcal{A} \cap \mathcal{B}) \geq \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}),$$

$$\Pr(\mathcal{C} \cap \mathcal{D}) \geq \Pr(\mathcal{C}) \cdot \Pr(\mathcal{D}),$$

$$\Pr(\mathcal{A} \cap \mathcal{C}) \leq \Pr(\mathcal{A}) \cdot \Pr(\mathcal{C}).$$

In terms of cardinalities, this can be read as follows:

$$\begin{aligned} 2^n|\mathcal{A} \cap \mathcal{B}| &\geq |\mathcal{A}| \cdot |\mathcal{B}|, \\ 2^n|\mathcal{C} \cap \mathcal{D}| &\geq |\mathcal{C}| \cdot |\mathcal{D}|, \\ 2^n|\mathcal{A} \cap \mathcal{C}| &\leq |\mathcal{A}| \cdot |\mathcal{C}|, \end{aligned}$$

where here and in what follows, $\mathcal{A} \cap \mathcal{B}$, $\mathcal{C} \cap \mathcal{D}$ and $\mathcal{A} \cap \mathcal{C}$ denote usual intersections of families.

Proof. Let $f : P(N) \to \mathbb{R}^+$ be the characteristic function of $\mathcal{A}$; i.e., $f(A) = 0$ if $A \notin \mathcal{A}$ and $f(A) = 1$ if $A \in \mathcal{A}$. Similarly, let g be the characteristic function of B. By the assumptions, f and g are both increasing. Applying the FKG-inequality with the trivial measure $\mu \equiv 1$ we get,

$$\Pr(\mathcal{A} \cap \mathcal{B}) = < fg > \geq < f > \cdot < g > = \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}) .$$

The other two inequalities follow similarly from Theorem 6.2.1 and the paragraph following it.

It is worth noting that the Proposition can be also derived easily from the Ahlswede-Daykin Theorem or from Corollary 6.1.3. ■

The last proposition has several interesting combinatorial consequences, some of which appear already in Kleitman's original paper. Since those are direct combinatorial consequences, and do not contain any additional probabilistic ideas, we omit

their exact statement and turn to a version of Proposition 6.3.1 in a more general probability space.

For a real vector $p = (p_1, , \ldots, , p_n)$, where $0 \leq p_i \leq 1$, consider the probability space whose elements are all members of the power set $P(N)$, where, for each $A \subseteq N$, $\Pr(A) = \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j)$. Clearly, this probability distribution is obtained if we choose a random $A \subseteq N$ by choosing each element $i \in N$, independently, with probability p_i. Let us denote, for each $\mathcal{A} \subseteq P(N)$, its probability in this space by $\Pr_p(\mathcal{A})$. In particular, if all the probabilities p_i are $1/2$ then $\Pr_p(\mathcal{A})$ is the quantity denoted as $\Pr(A)$ in Proposition 6.3.1. Define $\mu = \mu_p : P(N) \to \mathbb{R}^+$ by $\mu(A) = \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j)$.

It is easy to check that μ is log-supermodular. This is because for $A, B \subseteq N$, $\mu(A)\mu(B) = \mu(A \cup B)\mu(A \cap B)$, as can be checked by comparing the contribution arising from each $i \in N$ to the left-hand side and to the right-hand side of the last equality. Hence, one can apply the FKG-inequality and obtain the following generalization of Proposition 6.3.1.

Theorem 6.3.2 *Let $\mathcal{A}$ and $\mathcal{B}$ be two monotone increasing families of subsets of N and let $\mathcal{C}$ and $\mathcal{D}$ be two monotone decreasing families of subsets of N. Then, for any real vector $p = (p_1, , \ldots, , p_n)$, $0 \leq p_i \leq 1$,*

$$\begin{aligned} \Pr_p(\mathcal{A} \cap \mathcal{B}) &\geq \Pr_p(\mathcal{A}) \cdot \Pr_p(\mathcal{B}), \\ \Pr_p(\mathcal{C} \cap \mathcal{D}) &\geq \Pr_p(\mathcal{C}) \cdot \Pr_p(\mathcal{D}), \\ \Pr_p(\mathcal{A} \cap \mathcal{C}) &\leq \Pr_p(\mathcal{A}) \cdot \Pr_p(\mathcal{C}) . \end{aligned}$$

This theorem can be applied in many cases and will be used in Chapter 8 to derive the Janson Inequalities. As a simple illustration suppose that $A_1, A_2, \ldots, A_k$ are arbitrary subsets of N and one chooses a random subset A of N by choosing each $i \in N$, independently, with probability p. Then, Theorem 6.3.2 easily implies that

$$\Pr(A \text{ intersects each } A_i) \geq \prod_{i=1}^{k} \Pr(A \text{ intersects } A_i) .$$

Notice that this is false, in general, for other similar probabilistic models. For example, if A is a randomly chosen ℓ-element subset of N then the last inequality may fail.

By viewing the members of N as the $n = \binom{m}{2}$ edges of the complete graph on the set of vertices $V = \{1, 2, \ldots, m\}$ we can derive a correlation inequality for random graphs. Let $G = (V, E)$ be a random graph on the set of vertices V generated by choosing, for each $i, j \in V, i \neq j$, independently, the pair $\{i, j\}$ to be an edge with probability p. (This model of random graphs is discussed in detail in Chapter 10). A *property of graphs* is a subset of the set of all graphs on V, closed under isomorphism. Thus, for example, connectivity is a property (corresponding to all connected graphs on V) and planarity is another property. A property Q is *monotone increasing* if whenever G has Q and H is obtained from G by adding edges then H has Q, too. A *monotone decreasing* property is defined in a similar manner. By interpretting

the members of N in Theorem 6.3.2 as the $\binom{m}{2}$ pairs $\{i,j\}$ with $i,j \in V, i \neq j$ we obtain:

Theorem 6.3.3 *Let Q_1, Q_2, Q_3 and Q_4 be graph properties, where Q_1, Q_2 are monotone increasing and Q_3, Q_4 are monotone decreasing. Let $G = (V,E)$ be a random graph on V obtained by picking every edge, independently, with probability p. Then*

$$\begin{aligned}
\Pr(G \in Q_1 \cap Q_2) &\geq \Pr(G \in Q_1) \cdot \Pr(G \in Q_2), \\
\Pr(G \in Q_3 \cap Q_4) &\geq \Pr(G \in Q_3) \cdot \Pr(G \in Q_4), \\
\Pr(G \in Q_1 \cap Q_3) &\leq \Pr(G \in Q_1) \cdot \Pr(G \in Q_3).
\end{aligned}$$

Thus, for example, the probability that G is both Hamiltonian and planar does not exceed the product of the probability that it is Hamiltonian by that it is planar. It seems hopeless to try and prove such a statement directly, without using one of the correlation inequalities.

6.4 LINEAR EXTENSIONS OF PARTIALLY ORDERED SETS

Let $(P, \leq)$ be a partially ordered set with n elements. A *linear extension* of P is a one to one mapping $\sigma : P \to \{1, 2, \ldots, n\}$, which is order preserving, i.e., if $x, y \in P$ and $x \leq y$ then $\sigma(x) \leq \sigma(y)$. Intuitively, σ is a ranking of the elements of P which preserves the partial order of P. Consider the probability space of all linear extensions of P, where each possible extension is equally likely. In this space we can consider events of the form, e.g., $x \leq y$ or $(x \leq y) \wedge (x \leq z)$ (for $x, y, z \in P$) and compute their probabilities. It turns out that the FKG-inequality is a very useful tool for studying the correlation between such events. The best known result of this form was conjectured by Rival and Sands and proved by Shepp (1982). [See also Fishburn (1992) for a strengthening.] It asserts that for any partially ordered set P and any three elements $x, y, z \in P$: $\Pr(x \leq y \wedge x \leq z) \geq \Pr(x \leq y) \Pr(x \leq z)$.

This result became known as the XYZ-theorem. Although it looks intuitively obvious, its proof is nontrivial and contains a clever application of the FKG-inequality. In this section we present this result and its elegant proof.

Theorem 6.4.1 *Let P be a partially ordered set with n elements $a_1, a_2, \ldots, a_n$. Then*

$$\Pr(a_1 \leq a_2 \wedge a_1 \leq a_3) \geq \Pr(a_1 \leq a_2) \Pr(a_1 \leq a_3) .$$

Proof. Let m be a large integer (which will later tend to infinity) and let L be the set of all ordered n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$, where $x_i \in M = \{1, 2, \ldots, m\}$. (Note that we do *not* assume that the numbers x_i are distinct). Define an order relation $\leq$ on L as follows. For $\mathbf{y} = (y_1, \ldots, y_n) \in L$ and $\mathbf{x}$ as above $\mathbf{x} \leq \mathbf{y}$ iff $x_1 \geq y_1$ and $x_i - x_1 \leq y_i - y_1$ for all $2 \leq i \leq n$. It is not too difficult to check that $(L, \leq)$ is a lattice in which the i-th component of the meet $\mathbf{x} \wedge \mathbf{y}$ is

$(\mathbf{x} \wedge \mathbf{y})_i = \min(x_i - x_1, y_i - y_1) + \max(x_1, y_1)$ and the i-th component of the join $\mathbf{x} \vee \mathbf{y}$ is $(\mathbf{x} \vee \mathbf{y})_i = \max(x_i - x_1,\ y_i - y_1) + \min(x_1, y_1)$.

Moreover, the lattice L is distributive. This follows by an easy computation from the fact that the trivial lattice of integers (with respect to the usual order) is distributive and hence for any three integers a, b and c,

$$\min(a, \max(b, c)) = \max(\min(a, b), \min(a, c)), \tag{6.7}$$

and

$$\max(a, \min(b, c)) = \min(\max(a, b), \max(a, c)) \,. \tag{6.8}$$

Let us show how this implies that L is distributive. Let $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ and $\mathbf{z} = (z_1, \ldots, z_n)$ be three elements of L. We must show that

$$\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z}) = (\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z}) \,.$$

The i-th component of $\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z})$ is

$$\begin{aligned} (\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z}))_i \quad &= \min(x_i - x_1, (\mathbf{y} \vee \mathbf{z})_i - (\mathbf{y} \vee \mathbf{z})_1) \\ &\quad + \max(x_1, (\mathbf{y} \vee \mathbf{z})_1) \\ &= \min(x_i - x_1, \max(y_i - y_1, z_i - z_1)) \\ &\quad + \max(x_1, \min(y_1, z_1)). \end{aligned}$$

Similarly, the i-th component of $(\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z})$ is

$$\begin{aligned} &((\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z}))_i \\ &\quad = \max((\mathbf{x} \wedge \mathbf{y})_i - (\mathbf{x} \wedge \mathbf{y})_1,\ (\mathbf{x} \wedge \mathbf{z})_i - (\mathbf{x} \wedge \mathbf{z})_1) \\ &\quad + \min((\mathbf{x} \wedge \mathbf{y})_1,\ (\mathbf{x} \wedge \mathbf{z})_1) \\ &\quad = \max(\min(x_i - x_1,\ y_i - y_1),\ \min(x_i - x_1,\ z_i - z_1)) \\ &\quad + \min(\max(x_1, y_1), \max(x_1, z_1)) \,. \end{aligned}$$

These two quantities are equal, as follows by applying (6.7) with $a = x_i - x_1$, $b = y_i - y_1$, $c = z_i - z_1$ and (6.8) with $a = x_1$, $b = y_1$, $c = z_1$.

Thus L is distributive. To apply the FKG-inequality we need the measure function μ and the two functions f and g. Let μ be the characteristic function of P, i.e., for $\mathbf{x} = (x_1, \ldots, x_n) \in L$, $\mu(\mathbf{x}) = 1$ if $x_i \leq x_j$ whenever $a_i \leq a_j$ in P, and $\mu(\mathbf{x}) = 0$ otherwise. To show that μ is log-supermodular it suffices to check that if $\mu(\mathbf{x}) = \mu(\mathbf{y}) = 1$ then $\mu(\mathbf{x} \vee \mathbf{y}) = \mu(\mathbf{x} \wedge \mathbf{y}) = 1$. However, if $\mu(\mathbf{x}) = \mu(\mathbf{y}) = 1$ and $a_i \leq a_j$ in P then $x_i \leq x_j$ and $y_i \leq y_j$ and hence

$$\begin{aligned} (x \vee y)_i \quad &= \max(x_i - x_1,\ y_i - y_1) + \min(x_1, y_1) \\ &\leq \max(x_j - x_1,\ y_j - y_1) + \min(x_1, y_1) = (x \vee y)_j \,, \end{aligned}$$

i.e., $\mu(\mathbf{x} \vee \mathbf{y}) = 1$. Similarly, $\mu(\mathbf{x}) = \mu(\mathbf{y}) = 1$ implies $\mu(\mathbf{x} \wedge \mathbf{y}) = 1$, too.

Not surprisingly, we define the functions f and g as the characteristic functions of the two events $x_1 \leq x_2$ and $x_1 \leq x_3$, respectively, i.e., $f(\mathbf{x}) = 1$ if $x_1 \leq x_2$ and $f(\mathbf{x}) = 0$ otherwise, and $g(\mathbf{x}) = 1$ if $x_1 \leq x_3$ and $g(\mathbf{x}) = 0$ otherwise. Trivially, both

f and g are increasing. Indeed, if $\mathbf{x} \leq \mathbf{y}$ and $f(\mathbf{x}) = 1$ then $0 \leq x_2 - x_1 \leq y_2 - y_1$ and hence $f(\mathbf{y}) = 1$, and similarly for g.

We therefore have all the necessary ingredients for applying the FKG-inequality (Theorem 6.2.1). This gives that in L the probability that an n-tuple $(x_1, \ldots, x_n)$ that satisfies the inequalities in P, satisfies both $x_1 \leq x_2$ and $x_1 \leq x_3$ is at least as big as the product of the probability that it satisfies $x_1 \leq x_2$ by that it satisfies $x_1 \leq x_3$. Notice that this is not yet what we wanted to prove; the n-tuples in L are not n-tuples of distinct integers and thus do not correspond to linear extensions of P. However, as $m \to \infty$, the probability that $x_i = x_j$ for some $i \neq j$ in a member $\mathbf{x} = (x_1, \ldots, x_n)$ of L tends to 0 and the assertion of the theorem follows. ■

6.5 EXERCISES

1. Let G be a graph and let P denote the probability that a random subgraph of G obtained by picking each edge of G with probability $1/2$, independently, is connected (and spanning). Let Q denote the probability that in a random two-coloring of G, where each edge is chosen, randomly and independently, to be either red or blue, the red graph and the blue graph are both connected (and spanning). Is $Q \leq P^2$?

2. A family of subsets $\mathcal{G}$ is called *intersecting* if $G_1 \cap G_2 \neq \emptyset$ for all $G_1, G_2 \in \mathcal{G}$. Let $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$ be k intersecting families of subsets of $\{1, 2, \ldots, n\}$. Prove that
$$|\bigcup_{i=1}^{k} \mathcal{F}_i| \leq 2^n - 2^{n-k}.$$

3. Show that the probability that in the random graph $G(2k, 1/2)$ the maximum degree is at most $k - 1$ is at least $1/4^k$.

THE PROBABILISTIC LENS:
Turán's Theorem

In a graph $G = (V, E)$ let d_v denote the degree of a vertex v and let $\alpha(G)$ be the maximal size of an independent set of vertices. The following result was proved by Caro and Wei.

Theorem 1

$$\alpha(G) \geq \sum_{v \in V} \frac{1}{d_v + 1}.$$

Proof. Let $<$ be a uniformly chosen total ordering of V. Define

$$I = \{v \in V : \{v, w\} \in E \Rightarrow v < w\}.$$

Let X_v be the indicator random variable for $v \in I$ and $X = \sum_{v \in V} X_v = |I|$. For each v,

$$E[X_v] = \Pr[v \in I] = \frac{1}{d_v + 1},$$

since $v \in I$ if and only if v is the least element among v and its neighbors. Hence

$$E[X] = \sum_{v \in V} \frac{1}{d_v + 1}$$

and so there exists a specific ordering $<$ with

$$|I| \geq \sum_{v \in V} \frac{1}{d_v + 1}.$$

But if $x, y \in I$ and $\{x, y\} \in E$ then $x < y$ *and* $y < x$, a contradiction. Thus I is independent and $\alpha(G) \geq |I|$. ■

For any $m \leq n$ let q, r satisfy $n = mq + r, 0 \leq r < m$, and let $e = r\binom{q+1}{2} + (m - r)\binom{q}{2}$. Define a graph $G = G_{n,e}$ on n vertices and e edges by splitting the vertex set into m classes as evenly as possible and joining two vertices if and only if they lie in the same class. Clearly $\alpha(G_{n,e}) = m$.

Theorem 2 [Turán (1941)] *Let H have n vertices and e edges. Then $\alpha(H) \geq m$ and $\alpha(H) = m \Leftrightarrow H \cong G_{n,e}$.*

Proof. $G_{n,e}$ has $\sum_{v \in V}(d_v + 1)^{-1} = m$ since each clique contributes 1 to the sum. Fixing $e = \sum_{v \in V} d_v/2$, $\sum_{v \in V}(d_v + 1)^{-1}$ is minimized with the d_v as close together as possible. Thus for any H,

$$\alpha(H) \geq \sum_{v \in V} \frac{1}{d_v + 1} \geq m.$$

For $\alpha(H) = m$ we must have equality on both sides above. The second equality implies the d_v must be as close together as possible. Letting $X = |I|$ as in the previous theorem, assume $\alpha(H) = E[X]$. But $\alpha(H) \geq X$ for all values of $<$ so X must be a constant. Suppose H is not a union of cliques. Then, there exist $x, y, z \in V$ with $\{x, y\}, \{x, z\} \in E, \{y, z\} \notin E$. Let $<$ be an ordering that begins x, y, z and $<'$ the same ordering except that it begins y, z, x, and let I, I' be the corresponding sets of vertices all of whose neighbors are "greater." Then I, I' are identical except that $x \in I, y, z \notin I$ whereas $x \notin I', y, z \in I'$. Thus X is not constant. That is, $\alpha(H) = E[X]$ implies that H is the union of cliques and so $H \cong G_{n,e}$. ■

7

Martingales and Tight Concentration

Mathematics seems much more real to me than business – in the sense that, well, what's the reality in a McDonald's stand? It's here today and gone tomorrow. Now, the integers – that's reality. When you prove a theorem, you've really done something that has substance to it, to which no business venture can compare for reality.
– Jim Simons

7.1 DEFINITIONS

A martingale is a sequence $X_0, \ldots, X_m$ of random variables so that for $0 \leq i < m$,

$$E[X_{i+1}|X_i, X_{i-1}, \ldots, X_0] = X_i.$$

Imagine a gambler walking into a casino with X_0 dollars. The casino contains a variety of games of chance. All games are "fair" in that their expectations are zero. The gambler may allow previous history to determine his choice of game and bet. He might employ the gambler's definition of martingale – double the bet until you win. He might play roulette until he wins three times and then switch to keno. Let X_i be the gambler's fortune at time i. Given that $X_i = a$ the conditional expectation of X_{i+1} must be a and so this is a martingale.

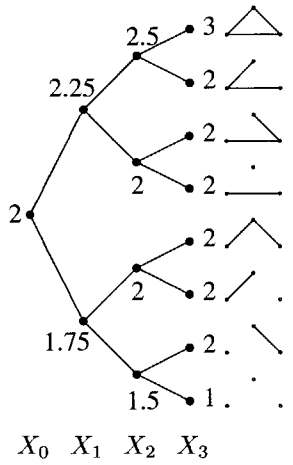
A simple but instructive martingale occurs when the gambler plays "flip a coin" for stakes of one dollar each time. Let $Y_1, \ldots, Y_m$ be independent coin flips, each $+1$ or -1 with probability $\frac{1}{2}$. Normalize so that $X_0 = 0$ is the gambler's initial stake, though he has unlimited credit. Then $X_i = Y_1 + \ldots + Y_i$ has distribution S_i.

Our martingales will look quite different, at least from the outside.

The Edge Exposure Martingale. Let the random graph $G(n,p)$ be the underlying probability space. Label the potential edges $\{i,j\} \subseteq [n]$ by $e_1, \ldots, e_m$, setting $m = \binom{n}{2}$ for convenience, in any specific manner. Let f be any graph theoretic function. We define a martingale $X_0, \ldots, X_m$ by giving the values $X_i(H)$. $X_m(H)$ is simply $f(H)$. $X_0(H)$ is the expected value of $f(G)$ with $G \sim G(n,p)$. Note that X_0 is a constant. In general (including the cases $i = 0$ and $i = m$),

$$X_i(H) = E[f(G) | e_j \in G \longleftrightarrow e_j \in H, 1 \leq j \leq i]$$

In words, to find $X_i(H)$ we first expose the first i pairs $e_1, \ldots, e_i$ and see if they are in H. The remaining edges are not seen and considered to be random. $X_i(H)$ is then the conditional expectation of $f(G)$ with this partial information. When $i = 0$ nothing is exposed and X_0 is a constant. When $i = m$ all is exposed and X_m is the function f. The martingale moves from no information to full information in small steps.



The edge exposure martingale with $n = m = 3$, f the chromatic number, and the edges exposed in the order "bottom, left, right". The values $X_i(H)$ are given by tracing from the central node to the leaf labelled H.

The figure shows why this is a martingale. The conditional expectation of $f(H)$ knowing the first $i-1$ edges is the weighted average of the conditional expectations of $f(H)$ where the i-th edge has been exposed. More generally – in what is sometimes referred to as a Doob martingale process – X_i may be the conditional expectation of $f(H)$ after certain information is revealed as long as the information known at time i includes the information known at time $i-1$.

The Vertex Exposure Martingale. Again let $G(n,p)$ be the underlying probability

space and f any graph theoretic function. Define $X_1, \ldots, X_n$ by

$$X_i(H) = E[f(G)| \text{ for } x, y \leq i, \{x, y\} \in G \longleftrightarrow \{x, y\} \in H].$$

In words, to find $X_i(H)$ we expose the first i vertices and all their internal edges and take the conditional expectation of $f(G)$ with that partial information. By ordering the edges appropriately the vertex exposure martingale may be considered a subsequence of the edge exposure martingale. Note that $X_1(H) = E[f(G)]$ is constant as no edges have been exposed and $X_n(H) = f(H)$ as all edges have been exposed.

7.2 LARGE DEVIATIONS

Maurey (1979) applied a large deviation inequality for martingales to prove an isoperimetric inequality for the symmetric group S_n. This inequality was useful in the study of normed spaces; see Milman and Schechtman (1986) for many related results. The applications of martingales in Graph Theory also all involve the same underlying martingale result used by Maurey, which is the following.

Theorem 7.2.1 [Azuma's Ineqality] *Let* $0 = X_0, \ldots, X_m$ *be a martingale with*

$$|X_{i+1} - X_i| \leq 1$$

for all $0 \leq i < m$. *Let* $\lambda > 0$ *be arbitrary. Then*

$$\Pr[X_m > \lambda\sqrt{m}] < e^{-\lambda^2/2}.$$

In the "flip a coin" martingale X_m has distribution S_m and this result is Theorem A.1.1. Indeed, the general proof is quite similar.

Proof. Set, with foresight, $\alpha = \lambda/\sqrt{m}$. Set $Y_i = X_i - X_{i-1}$ so that $|Y_i| \leq 1$ and $E[Y_i|X_{i-1}, X_{i-2}, \ldots, X_0] = 0$. Then, as in A.1.16,

$$E[e^{\alpha Y_i}|X_{i-1}, X_{i-2}, \ldots, X_0] \leq \cosh(\alpha) \leq e^{\alpha^2/2}.$$

Hence

$$\begin{aligned} E[e^{\alpha X_m}] &= E\left[\prod_{i=1}^{m} e^{\alpha Y_i}\right] \\ &= E\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right) E(e^{\alpha Y_m}|X_{m-1}, X_{m-2}, \ldots, X_0)\right] \\ &\leq E\left[\prod_{i=1}^{m-1} e^{\alpha Y_i}\right] e^{\alpha^2/2} \leq e^{\alpha^2 m/2}. \end{aligned}$$

Therefore

$$\begin{aligned} \Pr[X_m > \lambda\sqrt{m}] &= \Pr[e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}] \\ &< E[e^{\alpha X_m}]e^{-\alpha\lambda\sqrt{m}} \\ &\leq e^{\alpha^2 m/2 - \alpha\lambda\sqrt{m}} \\ &= e^{-\lambda^2/2}, \end{aligned}$$

as needed. ■

Corollary 7.2.2 *Let $c = X_0, \ldots, X_m$ be a martingale with*

$$|X_{i+1} - X_i| \leq 1$$

for all $0 \leq i < m$. Then

$$\Pr[|X_m - c| > \lambda\sqrt{m}] < 2e^{-\lambda^2/2}.$$

A graph theoretic function f is said to satisfy the edge Lipschitz condition if whenever H and H' differ in only one edge then $|f(H) - f(H')| \leq 1$. It satisfies the vertex Lipschitz condition if whenever H and H' differ at only one vertex, $|f(H) - f(H')| \leq 1$.

Theorem 7.2.3 *When f satisfies the edge Lipschitz condition, the corresponding edge exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$. When f satisfies the vertex Lipschitz condition the corresponding vertex exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$.*

We prove these results in a more general context later. They have the intuitive sense that if knowledge of a particular vertex or edge cannot change f by more than one then exposing a vertex or edge should not change the expectation of f by more than one. Now we give a simple application of these results.

Theorem 7.2.4 [Shamir and Spencer (1987)] *Let n, p be arbitrary and let $c = E[\chi(G)]$ where $G \sim G(n,p)$. Then*

$$\Pr[|\chi(G) - c| > \lambda\sqrt{n-1}] < 2e^{-\lambda^2/2}.$$

Proof. Consider the vertex exposure martingale $X_1, \ldots, X_n$ on $G(n,p)$ with $f(G) = \chi(G)$. A single vertex can always be given a new color so the vertex Lipschitz condition applies. Now apply Azuma's Inequality in the form of Corollary 7.2.2 . ■

Letting $\lambda \to \infty$ arbitrarily slowly, this result shows that the distribution of $\chi(G)$ is "tightly concentrated" around its mean. The proof gives no clue as to where the mean is.

7.3 CHROMATIC NUMBER

In Theorem 10.3.1 we prove that $\chi(G) \sim n/2\log_2 n$ almost surely, where $G \sim G(n, 1/2)$. Here we give the original proof of Béla Bollobás using martingales. We follow the notations of Chapter 10, Section 10.3, setting $f(k) = \binom{n}{k}2^{-\binom{k}{2}}$, k_0 so that $f(k_0 - 1) > 1 > f(k_0)$, $k = k_0 - 4$ so that $k \sim 2\log_2 n$ and $f(k) > n^{3+o(1)}$. Our goal is to show

$$\Pr[\omega(G) < k] = e^{-n^{2+o(1)}},$$

where $\omega(G)$ is the size of the maximum clique of G. We shall actually show in Theorem 7.3.2 a more precise bound. The remainder of the argument is given in Chapter 10, Section 10.3.

Let $Y = Y(H)$ be the maximal size of a family of edge disjoint cliques, of size k in H. This ingenious and unusual choice of function is key to the martingale proof.

Lemma 7.3.1 $E[Y] \geq \frac{n^2}{2k^4}(1 + o(1))$.

Proof. Let $\mathcal{K}$ denote the family of k-cliques of G so that $f(k) = \mu = E[|\mathcal{K}|]$. Let W denote the number of unordered pairs $\{A, B\}$ of k-cliques of G with $2 \leq |A \cap B| < k$. Then $E[W] = \Delta/2$, with Δ as described in Chapter 10, Section 10.3 (see also Chapter 4, Section 4.5), $\Delta \sim \mu^2 k^4 n^{-2}$. Let $\mathcal{C}$ be a random subfamily of $\mathcal{K}$ defined by setting, for each $A \in \mathcal{K}$,

$$\Pr[A \in \mathcal{C}] = q,$$

q to be determined. Let W' be the number of unordered pairs $\{A, B\}$, $A, B \in \mathcal{C}$ with $2 \leq |A \cap B| < k$. Then

$$E[W'] = E[W]q^2 = \Delta q^2/2.$$

Delete from $\mathcal{C}$ one set from each such pair $\{A, B\}$. This yields a set $\mathcal{C}^*$ of edge disjoint k-cliques of G and

$$E[Y] \geq E[|\mathcal{C}^*|] \geq E[|\mathcal{C}|] - E[W'] = \mu q - \Delta q^2/2 = \mu^2/2\Delta \sim n^2/2k^4,$$

where we choose $q = \mu/\Delta$ (noting that it is less than one!) to minimize the quadratic. ■

We conjecture that Lemma 7.3.1 may be improved to $E[Y] > cn^2/k^2$. That is, with positive probability there is a family of k-cliques which are edge disjoint and cover a positive proportion of the edges.

Theorem 7.3.2

$$\Pr[\omega(G) < k] < e^{-(c+o(1))\frac{n^2}{\ln^8 n}}$$

with c a positive constant.

Proof. Let $Y_0, \ldots, Y_m, m = \binom{n}{2}$, be the edge exposure martingale on $G(n, 1/2)$ with the function Y just defined. The function Y satisfies the edge Lipschitz condition as adding a single edge can only add at most one clique to a family of edge disjoint cliques. (Note that the Lipschitz condition would not be satisfied for the number of k-cliques as a single edge might yield many new cliques.) G has no k-clique if and only if $Y = 0$. Apply Azuma's Inequality with $m = \binom{n}{2} \sim n^2/2$ and $E[Y] \geq \frac{n^2}{2k^4}(1 + o(1))$. Then

$$\Pr[\omega(G) < k] \quad = \Pr[Y = 0] \leq \Pr[Y - E[Y] \leq -E[Y]]$$

$$\leq e^{-E[Y]^2/2\binom{n}{2}} \leq e^{-(c'+o(1))n^2/k^8}$$

$$= e^{-(c+o(1))n^2/\ln^8 n}$$

as desired. ■

Here is another example where the martingale approach requires an inventive choice of graph theoretic function.

Theorem 7.3.3 *Let $p = n^{-\alpha}$ where α is fixed, $\alpha > \frac{5}{6}$. Let $G = G(n,p)$. Then there exists $u = u(n,p)$ so that almost always*

$$u \leq \chi(G) \leq u + 3.$$

That is, $\chi(G)$ is concentrated in four values.

We first require a technical lemma that has been well known.

Lemma 7.3.4 *Let α, c be fixed $\alpha > \frac{5}{6}$. Let $p = n^{-\alpha}$. Then almost always every $c\sqrt{n}$ vertices of $G = G(n,p)$ may be three-colored.*

Proof. If not, let T be a minimal set which is not three-colorable. As $T - \{x\}$ is three-colorable, x must have internal degree at least 3 in T for all $x \in T$. Thus if T has t vertices it must have at least $\frac{3t}{2}$ edges. The probability of this occuring for some T with at most $c\sqrt{n}$ vertices is bounded from above by

$$\sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{\frac{3t}{2}} p^{3t/2}.$$

We bound

$$\binom{n}{t} \leq \left(\frac{ne}{t}\right)^t \text{ and } \binom{\binom{t}{2}}{\frac{3t}{2}} \leq \left(\frac{te}{3}\right)^{3t/2},$$

so each term is at most

$$\left[\frac{ne}{t}\frac{t^{3/2}e^{3/2}}{3^{3/2}}n^{-3\alpha/2}\right]^t \leq \left[c_1 n^{1-\frac{3\alpha}{2}} t^{1/2}\right]^t \leq \left[c_2 n^{1-\frac{3\alpha}{2}} n^{1/4}\right]^t = \left[c_2 n^{-\epsilon}\right]^t$$

with $\epsilon = \frac{3\alpha}{2} - \frac{5}{4} > 0$ and the sum is therefore $o(1)$. ■

Proof [Theorem 7.3.3] Let $\epsilon > 0$ be arbitrarily small and let $u = u(n,p,\epsilon)$ be the least integer so that

$$\Pr[\chi(G) \leq u] > \epsilon.$$

Now define $Y(G)$ to be the minimal size of a set of vertices S for which $G - S$ may be u-colored. This Y satisfies the vertex Lipschitz condition since at worst one could

add a vertex to S. Apply the vertex exposure martingale on $G(n,p)$ to Y. Letting $\mu = E[Y]$,

$$\Pr[Y \leq \mu - \lambda\sqrt{n-1}] < e^{-\lambda^2/2},$$

$$\Pr[Y \geq \mu + \lambda\sqrt{n-1}] < e^{-\lambda^2/2}.$$

Let λ satisfy $e^{-\lambda^2/2} = \epsilon$ so that these tail events each have probability less than ϵ. We defined u so that with probability at least ϵ, G would be u-colorable and hence $Y = 0$. That is, $\Pr[Y = 0] > \epsilon$. The first inequality therefore forces $\mu \leq \lambda\sqrt{n-1}$. Now employing the second inequality,

$$\Pr[Y \geq 2\lambda\sqrt{n-1}] \leq \Pr[Y \geq \mu + \lambda\sqrt{n-1}] \leq \epsilon.$$

With probability at least $1-\epsilon$ there is a u-coloring of all but at most $c'\sqrt{n}$ vertices. By the Lemma almost always, and so with probability at least $1-\epsilon$, these points may be colored with three further colors, giving a $u+3$-coloring of G. The minimality of u guarantees that with probability at least $1-\epsilon$ at least u colors are needed for G. Altogether

$$\Pr[u \leq \chi(G) \leq u+3] \geq 1-3\epsilon,$$

and ϵ was arbitrarily small. ■

Using the same technique, similar results can be achieved for other values of α. Together with some related ideas it can be shown that for any fixed $\alpha > \frac{1}{2}$, $\chi(G)$ is concentrated on at most two values. See Łuczak (1991) and Alon and Krivelevich (1997) for the detailed proofs.

7.4 TWO GENERAL SETTINGS

The martingales useful in studying Random Graphs generally can be placed in the following general setting which is essentially the one considered in Maurey (1979) and in Milman and Schechtman (1986). Let $\Omega = A^B$ denote the set of functions $g : B \to A$. (With B the set of pairs of vertices on n vertices and $A = \{0,1\}$ we may identify $g \in A^B$ with a graph on n vertices.) We define a measure by giving values p_{ab} and setting

$$\Pr[g(b) = a] = p_{ab},$$

with the values $g(b)$ assumed mutually independent. [In $G(n,p)$ all $p_{1b} = p, p_{0b} = 1-p$.] Now fix a gradation

$$\emptyset = B_0 \subset B_1 \subset \ldots \subset B_m = B.$$

Let $L : A^B \to R$ be a functional (e.g., clique number.) We define a martingale $X_0, X_1, \ldots, X_m$ by setting

$$X_i(h) = E[L(g)|g(b) = h(b) \text{ for all } b \in B_i].$$

X_0 is a constant, the expected value of L of the random g. X_m is L itself. The values $X_i(g)$ approach $L(g)$ as the values of $g(b)$ are "exposed." We say the functional L satisfies the Lipschitz condition relative to the gradation if for all $0 \leq i < m$,

$$h, h' \text{ differ only on } B_{i+1} - B_i \Rightarrow |L(h') - L(h)| \leq 1.$$

Theorem 7.4.1 *Let L satisfy the Lipschitz condition. Then the corresponding martingale satisfies*

$$|X_{i+1}(h) - X_i(h)| \leq 1$$

for all $0 \leq i < m$, $h \in A^B$.

Proof. Let H be the family of h' which agree with h on B_{i+1}. Then

$$X_{i+1}(h) = \sum_{h' \in H} L(h') w_{h'}$$

where $w_{h'}$ is the conditional probability that $g = h'$ given that $g = h$ on B_{i+1}. For each $h' \in H$ let $H[h']$ denote the family of h^* which agree with h' on all points except (possibly) $B_{i+1} - B_i$. The $H[h']$ partition the family of h^* agreeing with h on B_i. Thus we may express

$$X_i(h) = \sum_{h' \in H} \sum_{h^* \in H[h']} [L(h^*) q_{h^*}] w_{h'}$$

where q_{h^*} is the conditional probability that g agrees with h^* on B_{i+1} given that it agrees with h on B_i. (This is because for $h^* \in H[h']$, $w_{h'}$ is also the conditional probability that $g = h^*$ given that $g = h^*$ on B_{i+1}.) Thus

$$\begin{aligned} |X_{i+1}(h) - X_i(h)| &= \left| \textstyle\sum_{h' \in H} w_{h'} [L(h') - \sum_{h^* \in H[h']} L(h^*) q_{h^*}] \right| \\ &\leq \textstyle\sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} |q_{h^*} [L(h') - L(h^*)]| . \end{aligned}$$

The Lipschitz condition gives $|L(h') - L(h^*)| \leq 1$ so

$$|X_{i+1}(h) - X_i(h)| \leq \sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} q_{h^*} = \sum_{h' \in H} w_{h'} = 1.$$

■

Now we can express Azuma's Inequality in a general form.

Theorem 7.4.2 *Let L satisfy the Lipschitz condition relative to a gradation of length m and let $\mu = E[L(g)]$. Then for all $\lambda > 0$,*

$$\Pr[L(g) \geq \mu + \lambda\sqrt{m}] < e^{-\lambda^2/2},$$

$$\Pr[L(g) \leq \mu - \lambda\sqrt{m}] < e^{-\lambda^2/2}.$$

The second general setting is taken from Alon et al. (1997). We assume our underlying probability space is generated by a finite set of mutually independent Yes/No choices, indexed by $i \in I$. We are given a random variable Y on this space . Let p_i denote the probability that choice i is Yes. Let c_i be such that changing choice i (keeping all else the same) can change Y by at most c_i. We call c_i the *effect* of i. Let C be an upper bound on all c_i. We call $p_i(1-p_i)c_i^2$ the *variance* of choice i.

Now consider a solitaire game in which Paul finds the value of Y by making queries of an always truthful oracle Carole. The queries are always of a choice $i \in I$. Paul's choice of query can depend on Carole's previous responses. A strategy for Paul can then naturally be represented in a decision tree form. A "line of questioning" is a path from the root to a leaf of this tree, a sequence of questions and responses that determine Y. The total variance of a line of questioning is the sum of the variances of the queries in it.

Theorem 7.4.3 *For all $\epsilon > 0$ there exists $\delta > 0$ so that the following holds. Suppose Paul has a strategy for finding Y such that every line of questioning has total variance at most σ^2. Then*

$$\Pr[|Y - E[Y]| > \alpha\sigma] \leq 2e^{-\frac{\alpha^2}{2(1+\epsilon)}} \tag{7.1}$$

for all positive α with $\alpha C < \sigma(1+\epsilon)\delta$.

Applications. For a specific suboptimal bound we may take $\epsilon = \delta = 1$. If $C = O(1)$, $\alpha \to \infty$ and $\alpha = o(\sigma)$ the upper bound of (7.1) is $\exp[-\Omega(\alpha^2)]$. In many cases Paul queries *all* $i \in I$. Then we may take σ with $\sigma^2 = \sum_{i \in I} p_i(1-p_i)c_i^2$. For example, consider an edge Lipschitz Y on $G(n,p)$ with $p = p(n) \to 0$. I is the set of $m = \binom{n}{2}$ potential edges, all $p_i = p$, $C = 1$ so that $\sigma = \Theta(\sqrt{n^2 p})$. If $\alpha \to \infty$ with $\alpha = o(\sqrt{n^2 p})$ the upper bound of (7.1) is again $\exp[-\Omega(\alpha^2)]$.

Proof. For simplicity we replace Y by $Y - E[Y]$ so that we shall henceforth assume $E[Y] = 0$. By symmetry we shall bound only the upper tail of Y. We set, with foresight, $\lambda = \alpha/[\sigma(1+\epsilon)]$. Our side assumption gives that $C\lambda < \delta$. We will show

$$E[e^{\lambda Y}] \leq e^{(1+\epsilon)\lambda^2\sigma^2/2}. \tag{7.2}$$

The Martingale Inequality then follows by the Markov bound

$$\Pr[Y > \alpha\sigma] < e^{-\lambda\alpha\sigma}E[e^{\lambda Y}] \leq e^{-\alpha^2/2(1+\epsilon)}.$$

We first claim that for all $\epsilon > 0$ there exists $\delta > 0$ so that for $0 \leq p \leq 1$ and $|a| \leq \delta$

$$pe^{(1-p)a} + (1-p)e^{-pa} \leq e^{(1+\epsilon)p(1-p)a^2/2}. \tag{7.3}$$

Take the Taylor Series in a of the left-hand side. The constant term is 1, the linear term 0, the coefficient of a^2 is $\frac{1}{2}p(1-p)$ and for $j \geq 3$ the coefficient of a^j is at most

$\frac{1}{j!}p(1-p)[p^{j-1}+(1-p)^{j-1}] \leq \frac{1}{j!}p(1-p)$. Pick δ so that $|a| \leq \delta$ implies

$$\sum_{j=3}^{\infty} \frac{a^j}{j!} < \epsilon a^2/2.$$

(In particular this holds for $\epsilon = \delta = 1$.) Then

$$pe^{(1-p)a} + (1-p)e^{-pa} \leq 1 + p(1-p)\frac{a^2}{2}(1+\epsilon)$$

and (7.3) follows from the inequality $1 + x \leq e^x$.

Using this δ we show (7.2) by induction on the depth M of the decision tree. For $M = 0$, Y is constant and (7.2) is immediate. Otherwise, let $p, c, v = p(1-p)c^2$ denote the probability, effect and variance respectively of Paul's first query. Let μ_y, μ_n denote the conditional expectations of Y if Carole's response is Yes or No, respectively. Then $0 = E[Y]$ can be split into

$$0 = p\mu_y + (1-p)\mu_n.$$

The difference $\mu_y - \mu_n$ is the expected *change* in Y when all other choices are made independent with their respective probabilities and the root choice is changed from Yes to No. As this always changes Y by at most c,

$$|\mu_y - \mu_n| \leq c.$$

Thus we may parametrize

$$\mu_y = (1-p)b \quad \text{and} \quad \mu_n = -pb$$

with $|b| \leq c$. From (7.3)

$$pe^{\lambda\mu_y} + (1-p)e^{\lambda\mu_n} \leq e^{(1+\epsilon)p(1-p)b^2\lambda^2/2} \leq e^{(1+\epsilon)v\lambda^2/2}.$$

Let A_y denote the expectation of $e^{\lambda(Y-\mu_y)}$ conditional on Carole's first response being Yes and let A_n denote the analogous quantity for No. Given Carole's first response Paul has a decision tree (one of the two main subtrees) that determines Y with total variation at most $\sigma^2 - v$ and the tree has depth at most $M-1$. So by induction $A_y, A_n \leq A^-$ where we set

$$A^- = e^{(1+\epsilon)\lambda^2(\sigma^2-v)/2}.$$

Now we split

$$\begin{aligned} E[e^{\lambda Y}] &= pe^{\lambda\mu_y}A_y + (1-p)e^{\lambda\mu_n}A_n \\ &\leq [pe^{\lambda\mu_y} + (1-p)e^{\lambda\mu_n}]A^- \\ &\leq e^{(1+\epsilon)\lambda^2(v+(\sigma^2-v))/2} \end{aligned} \tag{7.4}$$

completing the proof of (7.2) and hence of Theorem 7.4.3. ■

We remark that this formal inductive proof somewhat masks the martingale. A martingale $E[Y] = Y_0, \ldots, Y_M = Y$ can be defined with Y_t the conditional expectation of Y after the first t queries and responses. Theorem 7.4.3 can be thought of as bounding the tail of Y by that of a normal distribution of greater or equal variance. For very large distances from the mean, large α, this bound fails.

7.5 FOUR ILLUSTRATIONS

Let g be the random function from $\{1,\ldots,n\}$ to itself, all n^n possible functions equally likely. Let $L(g)$ be the number of values not hit, i.e., the number of y for which $g(x) = y$ has no solution. By Linearity of Expectation,

$$E[L(g)] = n\left(1 - \frac{1}{n}\right)^n \sim \frac{n}{e}.$$

Set $B_i = \{1,\ldots,i\}$. L satisfies the Lipschitz condition relative to this gradation since changing the value of $g(i)$ can change $L(g)$ by at most 1. Thus:

Theorem 7.5.1

$$\Pr[|L(g) - \frac{n}{e}| > \lambda\sqrt{n} + 1] < 2e^{-\lambda^2/2}.$$

Deriving these asymptotic bounds from first principles is quite cumbersome.

As a second illustration let B be any normed space and let $v_1,\ldots,v_n \in B$ with all $|v_i| \leq 1$. Let $\epsilon_1,\ldots,\epsilon_n$ be independent with

$$\Pr[\epsilon_i = +1] = \Pr[\epsilon_i = -1] = \frac{1}{2}$$

and set

$$X = |\epsilon_1 v_1 + \ldots + \epsilon_n v_n|.$$

Theorem 7.5.2

$$\Pr[X - E[X] > \lambda\sqrt{n}] < e^{-\lambda^2/2},$$

$$\Pr[X - E[X] < -\lambda\sqrt{n}] < e^{-\lambda^2/2}.$$

Proof. Consider $\{-1,+1\}^n$ as the underlying probability space with all $(\epsilon_1,\ldots,\epsilon_n)$ equally likely. Then X is a random variable and we define a martingale $X_0,\ldots,X_n = X$ by exposing one ϵ_i at a time. The value of ϵ_i can only change X by 2, so direct application of Theorem 7.4.1 gives $|X_{i+1} - X_i| \leq 2$. But let ϵ, ϵ' be two n-tuples differing only in the i-th coordinate:

$$X_i(\epsilon) = \frac{1}{2}[X_{i+1}(\epsilon) + X_{i+1}(\epsilon')]$$

so that

$$|X_i(\epsilon) - X_{i+1}(\epsilon)| = \frac{1}{2}|X_{i+1}(\epsilon') - X_{i+1}(\epsilon)| \leq 1.$$

Now apply Azuma's Inequality. ■

For a third illustration let ρ be the Hamming metric on $\{0,1\}^n$. For $A \subseteq \{0,1\}^n$ let $B(A,s)$ denote the set of $y \in \{0,1\}^n$ so that $\rho(x,y) \leq s$ for some $x \in A$. ($A \subseteq B(A,s)$ as we may take $x = y$.)

Theorem 7.5.3 *Let $\epsilon, \lambda > 0$ satisfy $e^{-\lambda^2/2} = \epsilon$. Then*

$$|A| \geq \epsilon 2^n \Rightarrow |B(A, 2\lambda\sqrt{n})| \geq (1-\epsilon)2^n.$$

Proof. Consider $\{0,1\}^n$ as the underlying probability space, all points equally likely. For $y \in \{0,1\}^n$ set

$$X(y) = \min_{x \in A} \rho(x, y).$$

Let $X_0, X_1, \ldots, X_n = X$ be the martingale given by exposing one coordinate of $\{0,1\}^n$ at a time. The Lipschitz condition holds for X: If y, y' differ in just one coordinate then $|X(y) - X(y')| \leq 1$. Thus, with $\mu = E[X]$,

$$\Pr[X < \mu - \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon,$$

$$\Pr[X > \mu + \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon.$$

But

$$\Pr[X = 0] = |A|2^{-n} \geq \epsilon,$$

so $\mu \leq \lambda\sqrt{n}$. Thus

$$\Pr[X > 2\lambda\sqrt{n}] < \epsilon$$

and

$$|B(A, 2\lambda\sqrt{n})| = 2^n \Pr[X \leq 2\lambda\sqrt{n}] \geq 2^n(1-\epsilon).$$

■

Actually, a much stronger result is known. Let $B(s)$ denote the ball of radius s about $(0, \ldots, 0)$. The Isoperimetric Inequality proved by Harper (1966) states that

$$|A| \geq |B(r)| \Rightarrow |B(A, s)| \geq |B(r+s)|.$$

One may actually use this inequality as a beginning to give an alternate proof that $\chi(G) \sim n/2\log_2 n$ and to prove a number of the other results we have shown using martingales.

We illustrate Theorem 7.4.3 with a key technical lemma (in simplified form) from Alon et al. (1997). Let $G = (V, E)$ be a graph on N vertices, each vertex having degree D. Asymptotics will be for $N, D \to \infty$. Set $p = 1/D$. Define a random subgraph $H \subseteq G$ by placing each edge $e \in E$ in H with independent probability p. Let M (for matching) be the set of isolated edges of H. Let V^* be those $v \in V$ *not* in any $\{v, w\} \in M$. For $v \in V$ set $\deg^*(v)$ equal the number of $w \in V^*$ with $\{v, w\} \in E$. As

$$\Pr[v \notin V^*] = \sum_{\{v,w\} \in E} p(1-p)^{2D-1} = e^{-2} + O(D^{-1}),$$

linearity of expectation gives

$$E[\deg^*(v)] = D(1 - e^{-2}) + O(1).$$

We want $\deg^*(v)$ tightly concentrated about its mean.

In the notation of Theorem 7.4.3 the probability space is determined by the choices $e \in H$ for all $e \in E$. All $p_i = p$. Changing $e \in H$ to $e \notin H$ can change $\deg^*(v)$ by at most $C = 4$.

Paul needs to find $\deg^*(v)$ by queries of the form "Is $e \in H$?" For each w with $\{v, w\} \in E$ he determines if $w \in V^*$ by the following line of inquiry. First, for all u with $\{w, u\} \in E$ he queries if $\{w, u\} \in H$. If no $\{w, u\} \in H$ then $w \in V^*$. If two (or more) $\{w, u_1\}, \{w, u_2\} \in H$ then w cannot be in an *isolated* edge of H so $w \in V^*$. Now suppose $\{w, u\} \in H$ for precisely one u. Paul then asks (using his acquired knowledge!) for each $z \neq w$ with $\{u, z\} \in E$ if $\{u, z\} \in H$. The replies determine if $\{w, u\}$ is an isolated edge of H and hence if $w \in V^*$. Paul has made at most $D + (D - 1)$ queries for each w for a total of at most $D(2D - 1) = O(D^2)$ queries. We deduce

$$\Pr[|\deg^*(v) - D(1 - e^{-2})| > \lambda D^{1/2}] = \exp[-\Omega(\lambda^2)]$$

when $\lambda \to \infty$ and $\lambda = o(D^{1/2})$.

In application one wishes to iterate this procedure (now applying it to the restriction of G to V^*) in order to find a large matching. This is somewhat akin to the Rödl nibble of §4.7. There are numerous further complications but the tight concentration of $\deg^*(v)$ about its mean plays an indispensable role.

7.6 TALAGRAND'S INEQUALITY

Let $\Omega = \prod_{i=1}^{n} \Omega_i$ where each Ω_i is a probability space and Ω has the product measure. Let $A \subseteq \Omega$ and let $\vec{x} = (x_1, \ldots, x_n) \in \Omega$. Talagrand (1996) gives an unusual, subtle and ultimately powerful notion of the distance – denoted $\rho(A, \vec{x})$ – from $\vec{x}$ to A. We imagine moving from $\vec{x}$ to some $\vec{y} = (y_1, \ldots, y_n) \in A$ by changing coordinates. $\rho(A, \vec{x})$ will measure the minimal cost of such a move when a suitably restricted adversary sets the cost of each change.

Definition 2 *$\rho(A, \vec{x})$ is the least value such that for any $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in R^n$ with $|\vec{\alpha}| = 1$ there exists $\vec{y} = (y_1, \ldots, y_n) \in A$ with*

$$\sum_{x_i \neq y_i} \alpha_i \leq \rho(A, \vec{x}).$$

Note that $\vec{y}$ can, and generally will, depend on $\vec{\alpha}$.

We define for any real $t \geq 0$,

$$A_t = \{\vec{x} \in \Omega : \rho(A, \vec{x}) \leq t\}.$$

Note $A_0 = A$ as when $\vec{x} \in A$ one can select $\vec{y} = \vec{x}$.

Talagrand's Inequality

$$\Pr[A](1 - \Pr[A_t]) \leq e^{-t^2/4}.$$

In particular, if $\Pr[A] \geq \frac{1}{2}$ (or any fixed constant) and t is "very large" then all but a very small proportion of Ω is within "distance" t of A.

Example. Take $\Omega = \{0,1\}^n$ with the uniform distribution and let τ be the Hamming (L^1) metric. Then $\rho(A,\vec{x}) \geq \min_{\vec{y}\in A} \tau(\vec{x},\vec{y})n^{-1/2}$ as the adversary can choose all $\alpha_i = n^{-1/2}$. Suppose to move from $\vec{x}$ to A the values $x_1,\ldots,x_l$ (or any particular l coordinates) must be changed. Then $\rho(A,\vec{x}) \geq l^{1/2}$ as the adversary could choose $\alpha_i = l^{-1/2}$ for $1 \leq i \leq l$ and zero elsewhere.

Define $U(A,\vec{x})$ to be the set of $\vec{s} = (s_1,\ldots,s_n) \in \{0,1\}^n$ with the property that there exists $\vec{y} \in A$ such that

$$x_i \neq y_i \Rightarrow s_i = 1.$$

We may think of $U(A,\vec{x})$ as representing the possible paths from $\vec{x}$ to A. Note that when $s_i = 1$ we, for somewhat technical reasons, do not require $x_i \neq y_i$. With this notation $\rho(A,\vec{x})$ is the least real so that for all $\vec{\alpha}$ with $|\vec{\alpha}| = 1$ there exists $\vec{s} \in U(A,\vec{x})$ with $\vec{\alpha}\cdot\vec{s} \leq \rho(A,\vec{x})$.

Now define $V(A,\vec{x})$ to be the convex hull of $U(A,\vec{x})$. The following result gives an alternate characterization of ρ which supplies the concept with much of its richness.

Theorem 7.6.1

$$\rho(A,\vec{x}) = \min_{\vec{v}\in V(A,\vec{x})} |\vec{v}|.$$

Proof. Let $\vec{v} \in V(A,\vec{x})$ achieve this minimum. The hyperplane through $\vec{v}$ perpendicular to the line from the origin to $\vec{v}$ then separates $V(A,\vec{x})$ from the origin so that all $\vec{s} \in V(A,\vec{x})$ have $\vec{s}\cdot\vec{v} \geq \vec{v}\cdot\vec{v}$. Set $\vec{\alpha} = \vec{v}/|\vec{v}|$. Then all $\vec{s} \in U(A,\vec{x}) \subseteq V(A,\vec{x})$ have $\vec{s}\cdot\vec{\alpha} \geq \vec{v}\cdot\vec{v}/|\vec{v}| = |\vec{v}|$. Conversely, take any $\vec{\alpha}$ with $|\vec{\alpha}| = 1$. Then $\vec{\alpha}\cdot\vec{v} \leq |\vec{v}|$. As $\vec{v} \in V(A,\vec{x})$ we may write $\vec{v} = \sum \lambda_i \vec{s}_i$ for some $\vec{s}_i \in U(A,\vec{x})$, with all $\lambda_i \geq 0$ and $\sum \lambda_i = 1$. Then

$$|\vec{v}| \geq \sum \lambda_i (\vec{\alpha}\cdot\vec{s}_i)$$

and hence some $\vec{\alpha}\cdot\vec{s}_i \leq |\vec{v}|$. ■

The case $\Omega = \{0,1\}^n$ is particularly important and instructive. There $\rho(A,\vec{x})$ is simply the Euclidean distance from $\vec{x}$ to the convex hull of A.

Theorem 7.6.2

$$\int_\Omega \exp\left[\frac{1}{4}\rho^2(A,\vec{x})\right] d\vec{x} \leq \frac{1}{\Pr[A]}.$$

Talagrand's Theorem is an immediate corollary of the above result. Indeed, fix A and consider the random variable $X = \rho(A,\vec{x})$. Then

$$\Pr[\overline{A_t}] = \Pr[X \geq t] = \Pr[e^{X^2/4} \geq e^{t^2/4}] \leq E[e^{X^2/4}]e^{-t^2/4},$$

and the theorem states $E[e^{X^2/4}] \leq \frac{1}{\Pr[A]}$.

Proof [Theorem 7.6.2] We use induction on the dimension n. For $n = 1$, $\rho(A, \vec{x}) = 1$ if $\vec{x} \notin A$, zero otherwise so that

$$\int \exp\left[\frac{1}{4}\rho^2(A, \vec{x})\right] = \Pr[A] + (1 - \Pr[A])e^{1/4} \leq \frac{1}{\Pr[A]},$$

as the inequality $u + (1-u)e^{1/4} \leq u^{-1}$ for $0 < u \leq 1$ is a simple calculus exercise.

Assume the result for n. Write OLD $= \prod_{i=1}^{n} \Omega_i$, $NEW = \Omega_{n+1}$ so that $\Omega =$ OLD $\times$ NEW and any $z \in \Omega$ can be uniquely written $z = (x, \omega)$ with $x \in$ OLD, $\omega \in$ NEW. Set

$$B = \{x \in \text{OLD} : (x, \omega) \in A \text{ for some } \omega \in \text{NEW}\}$$

and for any $\omega \in$ NEW set

$$A_\omega = \{x \in \text{OLD} : (x, \omega) \in A\}.$$

Given $z = (x, \omega) \in \Omega$ we can move to A in two basic ways – either by changing ω, which reduces the problem to moving from x to B, or by not changing ω, which reduces the problem to moving from x to A_ω. Thus

$$\vec{s} \in U(B, x) \Rightarrow (\vec{s}, 1) \in U(A, (x, \omega))$$

and

$$\vec{t} \in U(A_\omega, x) \Rightarrow (\vec{t}, 0) \in U(A, (x, \omega)).$$

Taking the convex hulls, if $\vec{s} \in V(B, x)$ and $\vec{t} \in V(A_\omega, x)$ then $(\vec{s}, 1)$ and $(\vec{t}, 0)$ are in $V(A, (x, \omega))$ and hence for any $\lambda \in [0, 1]$,

$$((1-\lambda)\vec{s} + \lambda\vec{t}, 1 - \lambda) \in V(A, (x, \omega)).$$

Then, by convexity,

$$\rho^2(A, (x, \omega)) \leq (1-\lambda)^2 + |(1-\lambda)\vec{s} + \lambda\vec{t}|^2 \leq (1-\lambda)^2 + (1-\lambda)|\vec{s}|^2 + \lambda|\vec{t}|^2.$$

Selecting $\vec{s}, \vec{t}$ with minimal norms yields the critical inequality

$$\rho^2(A, (x, \omega)) \leq (1-\lambda)^2 + \lambda\rho^2(A_\omega, x) + (1-\lambda)\rho^2(B, x).$$

Quoting from Talagrand, "The main trick of the proof is to resist the temptation to optimize now over λ." Rather, we first fix ω and bound

$$\begin{aligned} &\textstyle\int_x \exp\left[\frac{1}{4}\rho^2(A, (x, \omega))\right] \\ &\leq e^{(1-\lambda)^2/4} \textstyle\int_x \left(\exp\left[\frac{1}{4}\rho^2(A_\omega, x)\right]\right)^\lambda \left(\exp\left[\frac{1}{4}\rho^2(B, x)\right]\right)^{1-\lambda}. \end{aligned}$$

By Hölder's Inequality this is at most

$$e^{(1-\lambda)^2/4} \left[\int_x \exp\left[\frac{1}{4}\rho^2(A_\omega, x)\right]\right]^\lambda \left[\int_x \exp\left[\frac{1}{4}\rho^2(B, x)\right]\right]^{1-\lambda}$$

which by induction is at most

$$e^{(1-\lambda)^2/4}\left(\frac{1}{\Pr[A_\omega]}\right)^{\lambda}\left(\frac{1}{\Pr[B]}\right)^{1-\lambda} = \frac{1}{\Pr[B]}e^{(1-\lambda)^2/4}r^{-\lambda},$$

where $r = \Pr[A_\omega]/\Pr[B] \leq 1$. *Now* we use calculus and minimize $e^{(1-\lambda)^2/4}r^{-\lambda}$ by choosing $\lambda = 1 + 2\ln r$ for $e^{-1/2} \leq r \leq 1$ and $\lambda = 0$ otherwise. Further (somewhat tedius but simple) calculation shows $e^{(1-\lambda)^2/4}r^{-\lambda} \leq 2 - r$ for this $\lambda = \lambda(r)$. Thus

$$\int_x \exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] \leq \frac{1}{\Pr[B]}\left(2 - \frac{\Pr[A_\omega]}{\Pr[B]}\right).$$

We integrate over ω giving

$$\int_\omega \int_x \exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] \leq \frac{1}{\Pr[B]}\left(2 - \frac{\Pr[A]}{\Pr[B]}\right) = \frac{1}{\Pr[A]}x(2-x)$$

where $x = \Pr[A]/\Pr[B] \in [0,1]$. But $x(2-x) \leq 1$, completing the induction and hence the theorem. ■

7.7 APPLICATIONS OF TALAGRAND'S INEQUALITY

Let $\Omega = \prod_{i=1}^n \Omega_i$ where each Ω_i is a probability space and Ω has the product measure. Let $h : \Omega \to R$. Talagrand's Inequality enables us, under certain conditions, to show that the random variable $X = h(\cdot)$ is tightly concentrated. In this sense it can serve the same function Azuma's Inequality does for martingales and there are many cases in which it gives far stronger results.

We call $h : \Omega \to R$ *Lipschitz* if $|h(x) - h(y)| \leq 1$ whenever x, y differ in at most one coordinate. Talagrand's Inequality is most effective on those Lipschitz functions with the property that when $h(x) \geq s$ there are a relatively small number of coordinates that will certify that $h(x) \geq s$. We formalize this notion as follows.

Definition 3 *Let $f : N \to N$. h is f-*certifiable *if whenever $h(x) \geq s$ there exists $I \subseteq \{1,\ldots,n\}$ with $|I| \leq f(s)$ so that all $y \in \Omega$ that agree with x on the coordinates I have $h(y) \geq s$.*

Example. Consider $G(n,p)$ as the product of $\binom{n}{2}$ coin flips and let $h(G)$ be the number of triangles in G. Then h is f-certifiable with $f(s) = 3s$. For if $h(G) \geq s$ there exist s triangles which together have at most $3s$ edges and any other G' with those $3s$ edges has $h(G') \geq s$. Note I, here the indices for those $3s$ edges, very much depends on G. Also note that we need certify only lower bounds for h.

Theorem 7.7.1 *Under the above assumptions and for all b, t,*

$$\Pr[X \leq b - t\sqrt{f(b)}]\Pr[X \geq b] \leq e^{-t^2/4}.$$

Proof. Set $A = \{x : h(x) < b - t\sqrt{f(b)}\}$. Now suppose $h(y) \geq b$. We claim $y \notin A_t$. Let I be a set of indices of size at most $f(b)$ that certifies $h(y) \geq b$ as given above. Define $\alpha_i = 0$ when $i \notin I$, $\alpha_i = |I|^{-1/2}$ when $i \in I$. If $y \in A_t$ there exists a $z \in A$ that differs from y in at most $t|I|^{1/2} \leq t\sqrt{f(b)}$ coordinates of I though at arbitrary coordinates outside of I. Let y' agree with y on I and agree with z outside of I. By the certification $h(y') \geq b$. Now y', z differ in at most $t\sqrt{f(b)}$ coordinates and so, by Lipschitz,

$$h(z) \geq h(y') - t\sqrt{f(b)} \geq b - t\sqrt{f(b)}$$

but then $z \notin A$, a contradiction. So $\Pr[X \geq b] \leq \Pr[\overline{A_t}]$ so from Talagrand's Theorem,

$$\Pr[X < b - t\sqrt{f(b)}] \Pr[X \geq b] \leq e^{-t^2/4}.$$

As the right-hand side is continuous in t we may replace $<$ by $\leq$ giving the Theorem. ■

A small generalization is sometimes useful. Call $h : \Omega \to R$ *K-Lipschitz* if $|h(x) - h(y)| \leq K$ whenever x, y differ in only one coordinate. Applying the above theorem to h/K, which is Lipschitz, we find

$$\Pr[X \leq b - tK\sqrt{f(b)}] \Pr[X \geq b] \leq e^{-t^2/4}.$$

In applications one often takes b to be the median so that for t large the probability of being $t\sqrt{f(b)}$ under the median goes sharply to zero. But it works both ways, by parametrizing so that $m = b - t\sqrt{f(b)}$ is the median one usually gets $b \sim m + t\sqrt{f(m)}$ and that the probability of being $t\sqrt{f(b)}$ above the median goes sharply to zero. Martingales, via Azuma's Inequality, generally produce a concentration result around the mean μ of X while Talagrand's Inequality yields a concentration result about the median m. Means tend to be easy to compute, medians notoriously difficult, but tight concentration result will generally allow us to show that the mean and median are not far away.

Let $x = (x_1, \ldots, x_n)$ where the x_i are independently and uniformly chosen from $[0, 1]$. Set $X = h(x)$ to be the length of the longest increasing subsequence of x. Elementary methods give that $c_1 n^{1/2} < X < c_2 n^{1/2}$ almost surely for some positive constants c_1, c_2 and that the mean μ and median m of X are both in that range. Also X is Lipschitz, as changing one x_i can only change X by at most one. How concentrated is X? We can apply Azuma's Inequality to deduce that if $s \gg n^{1/2}$ then $|X - \mu| \leq s$ almost surely. This is not particularly good since X itself is only of order $n^{1/2}$. Now consider Talagrand's Inequality. X is f-certifiable with $f(s) = s$ since if x has an increasing subsequence of length s then those s coordinates certify that $X \geq s$. Then $\Pr[X < m - tm^{1/2}] \leq e^{-t^2/4}/\Pr[X \geq m] \leq 2e^{-t^2/4}$ as m is the median value. But $m = \Theta(n^{1/2})$. Thus when $s \gg n^{1/4}$ we have $X > m - s$ almost surely. For the other side suppose $t \to \infty$ slowly and let b be such that $b - tb^{1/2} = m$. Then $\Pr[X \geq b] \leq e^{-t^2/4}/\Pr[X \leq m] \leq 2e^{-t^2/4}$. Then $X \leq b$ almost surely. But $b = m + (1 + o(1))tm^{1/2}$ so that $X \leq m + tm^{1/2}$ almost surely. Combining, if $s \gg n^{1/4}$ then $|X - m| < s$ almost surely. A much stronger result,

determining the precise asymptotic distribution of X, has recently been obtained by Baik, Deift and Johansson (1999), using deep analytic tools.

Let's reexamine the bound (Theorem 7.3.2) that $G(n, \frac{1}{2})$ has no clique of size k with k as defined there. We let, as there, Y be the maximal number of edge disjoint k-cliques. From the work there $E[Y] = \Omega(n^2 k^{-4})$ and Y is tightly concentrated about $E[Y]$ so that the median m of Y must also have $m = \Omega(n^2 k^{-4})$. As before Y is Lipschitz. Further Y is f-certifiable with $f(s) = \binom{k}{2} s$ as the edges of the s-cliques certify that $Y \geq s$. Hence

$$\Pr\left[Y \leq m - tm^{1/2}\binom{k}{2}^{1/2}\right] \Pr[Y \geq m] < e^{-t^2/4}.$$

Set $t = \Theta(m^{1/2}/k)$ so that $m = tm^{1/2}\binom{k}{2}^{1/2}$. Then

$$\Pr[\omega(G) < k] = \Pr[Y \leq 0] < 2e^{-t^2/4} < \exp\left[-\Omega\left(\frac{n^2}{\ln^6 n}\right)\right]$$

which improves the bound of Theorem 7.3.2. Still, we should note that application of the Extended Janson Inequalityin §10.3 does even better.

7.8 KIM-VU POLYNOMIAL CONCENTRATION

A recent approach of Kim and Vu (2000) looks highly promising. Let $H = (V(H), E(H))$ be a hypergraph and let each edge $e \in E(H)$ have a nonnegative weight $w(e)$. Let t_i, $i \in V(H)$ be mutually independent indicator random variables with $E[t_i] = p_i$. Consider the random variable polynomial

$$Y = \sum_{e \in E(H)} w_e \prod_{i \in e} t_i.$$

We allow $e = \emptyset$ in which case $\prod_{i \in e} t_i$ is by convention 1. We want to show that Y is concentrated about its mean.

Let $S \subseteq V(H)$ be a random set given by $\Pr[i \in S] = p_i$, these events mutually independent over $i \in V(H)$. Then Y is the weighted number of hyperedges e in the restriction of H to S. In applications we generally have all weights equal one so that Y simply counts the hyperedges in the random S. But we may also think abstractly of Y as simply any polynomial over the indicators t_i having all nonnegative coefficients.

We set $n = |V(H)|$, the number of vertices of H (number of variables t_i). Let k be an upper bound on the size of all hyperedges (upper bound on the degree of the polynomial Y).

Let $A \subseteq V(H)$ with $|A| \leq k$. We truncate Y to Y_A as follows: For those terms $\prod_{i \in e} t_i$ with $A \subseteq e$ we set $t_i = 1$ for all $i \in A$, replacing the term by $\prod_{i \in e - A} t_i$. All other terms (where e does not contain A) are deleted. For example, with $A = \{1\}$, $2t_1t_2 + 5t_1t_3t_4 + 7t_2t_4$ becomes $2t_2 + 5t_3t_4$. Intriguingly, as polynomials in the t_i,

Y_A is the partial derivative of Y with respect to the t_i, $i \in A$. Set $E_A = E[Y_A]$. That is, E_A is the expected number of hyperedges in S that contain A, conditional on all vertices of A being in S. Set E_i equal the maximal E_A over all $A \subseteq V(H)$ of size i. Set $\mu = E[Y]$ for convenience and set

$$E' = \max_{1 \le i \le k} E_i \text{ and } E = \max[\mu, E'].$$

Theorem 7.8.1 [Kim-Vu Polynomial Concentration] *With the above hypotheses*

$$\Pr[|Y - \mu| > a_k (EE')^{1/2} \lambda^k] < d_k e^{-\lambda} n^{k-1}$$

for any $\lambda > 1$.

Here, for definiteness, we may take $a_k = 8^k k!^{1/2}$ and $d_k = 2e^2$.

We omit the proof, which combines martingale inequalities similar to those of Theorem 7.4.3 with a subtle induction on the degree k. There may well be room for improvement in the a_k, d_k and n^{k-1} terms. In applications one generally has k fixed and $\lambda \gg \ln n$ so that the $e^{-\lambda}$ term dominates the probability bound.

Applications of Kim-Vu Polynomial Concentration tend to be straightforward. Let $G \sim G(n,p)$ with $p = n^{-\alpha}$ and assume $0 < \alpha < 2/3$. Fix a vertex x of G and let $Y = Y(x)$ be the number of triangles containing x. Set $\mu = E[Y] = \binom{n-1}{2} p^3 \sim \frac{1}{2} n^{2-3\alpha}$. Let $\delta > 0$ be fixed. We want to bound $\Pr[|Y - \mu| < \delta\mu]$.

The random graph G is defined by the random variables t_{ij}, one for each unordered pair of vertices, which are indicators of the adjacency of the two vertices. In that context

$$Y = \sum_{i,j \ne x} t_{xi} t_{xj} t_{ij}.$$

This is a polynomial of degree $k = 3$. When A consists of a single edge xi we find $E_A = (n-2)p^2$; when it consists of three edges forming a triangle containing x we find $E_A = 1$. When $A = \emptyset$, $E_A = \mu$. Other cases give smaller E_A. Basically $E' \sim \max[np^2, 1]$. Calculation gives $E' \sim c\mu n^{-\epsilon}$ for some positive ϵ (dependent on α) throughout our range. We apply Kim-Vu Polynomial Concentration with $\lambda = c' n^{\epsilon/6}$, c' a small positive constant, to bound $\Pr[|Y - \mu| < \delta\mu]$ by $\exp[-\Omega(n^{\epsilon/6})]$. Note that the n^{k-1} factor is absorbed by the exponential.

In particular, as this probability is $o(n^{-1})$, we have that almost surely every vertex x is in $\sim \mu$ triangles. This result generalizes. Fix $\alpha \in (0,1)$ and suppose (R,H) is a rooted graph, safe, in the sense of §10.7, with respect to α. Let $G \sim G(n,p)$ with $p = n^{-\alpha}$. For distinct vertices $x_1, \ldots, x_r$ let $Y = Y(x_1, \ldots, x_r)$ denote the number of extensions in G to H. Set $\mu = E[Y]$. Kim-Vu Polynomial Concentration gives an exponentially small upper bound on the probability that Y is not near μ. In particular, this probability is $o(n^{-r})$. Hence almost surely every r vertices have $\sim \mu$ extensions to H.

7.9 EXERCISES

1. Let $G = (V, E)$ be the graph whose vertices are all 7^n vectors of length n over Z_7, in which two vertices are adjacent iff they differ in precisely one coordinate. Let $U \subset V$ be a set of 7^{n-1} vertices of G, and let W be the set of all vertices of G whose distance from U exceeds $(c+2)\sqrt{n}$, where $c > 0$ is a constant. Prove that $|W| \leq 7^n \cdot e^{-c^2/2}$.

2. (*) Let $G = (V, E)$ be a graph with chromatic number $\chi(G) = 1000$. Let $U \subset V$ be a random subset of V chosen uniformly among all $2^{|V|}$ subsets of V. Let $H = G[U]$ be the induced subgraph of G on U. Prove that

$$\Pr(\chi(H) \leq 400) < 1/100.$$

3. Prove that there is an absolute constant c such that for every $n > 1$ there is an interval I_n of at most $c\sqrt{n}/\log n$ consecutive integers such that the probability that the chromatic number of $G(n, 0.5)$ lies in I_n is at least 0.99.

THE PROBABILISTIC LENS:
Weierstrass Approximation Theorem

The well-known Weierstrass Approximation Theorem asserts that the set of real polynomials over $[0,1]$ is dense in the space of all continuous real functions over $[0,1]$. This is stated in the following theorem.

Theorem 1 [Weierstrass Approximation Theorem] *For every continuous real function $f : [0,1] \mapsto R$ and every $\epsilon > 0$, there is a polynomial $p(x)$ such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in [0,1]$.*

Bernstein (1912) gave a charming probabilistic proof of this theorem, based on the properties of the Binomial distribution . His proof is the following.

Proof. Since a continuous $f : [0,1] \mapsto R$ is uniformly continuous there is a $\delta > 0$ such that if $x, x' \in [0,1]$ and $|x - x'| \leq \delta$ then $|f(x) - f(x')| \leq \epsilon/2$. In addition, since f must be bounded there is an $M > 0$ such that $|f(x)| \leq M$ in $[0,1]$.

Let $B(n,x)$ denote the Binomial random variable with n independent trials and probability of success x for each of them. Thus, the probability that $B(n,x) = j$ is precisely $\binom{n}{j} x^j (1-x)^{n-j}$. The expectation of $B(n,x)$ is nx and its standard deviation is $\sqrt{nx(1-x)} \leq \sqrt{n}$. Therefore, by Chebyschev's Inequality discussed in Chapter 4, for every integer n, $\Pr(|B(n,x) - nx| > n^{2/3}) \leq \frac{1}{n^{1/3}}$. It follows that there is an integer n such that

$$\Pr(|B(n,x) - nx| > n^{2/3}) < \frac{\epsilon}{4M}$$

and

$$\frac{1}{n^{1/3}} < \delta.$$

Define

$$P_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right).$$

We claim that for every $x \in [0,1]$, $|P_n(x) - f(x)| \leq \epsilon$. Indeed, since $\sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} = 1$, we have

$$\begin{aligned} |P_n(x) - f(x)| &\leq \textstyle\sum_{i;|i-nx|\leq n^{2/3}} \binom{n}{i} x^i (1-x)^{n-i} |f(\frac{i}{n}) - f(x)| \\ &+ \textstyle\sum_{i;|i-nx|>n^{2/3}} \binom{n}{i} x^i (1-x)^{n-i} (|f(\frac{i}{n})| + |f(x)|) \\ &\leq \textstyle\sum_{i;|i/n-x|\leq n^{-1/3}<\delta} \binom{n}{i} x^i (1-x)^{n-i} |f(\frac{i}{n}) - f(x)| \\ &+ \Pr(|B(n,x) - nx| > n^{2/3}) 2M \\ &\leq \tfrac{\epsilon}{2} + \tfrac{\epsilon}{4M} 2M = \epsilon. \end{aligned}$$

This completes the proof. ■

8

The Poisson Paradigm

> One of the things that attracts us most when we apply ourselves to a mathematical problem is precisely that within us we always hear the call: here is the problem, search for the solution, you can find it by pure thought, for in mathematics there is no *ignorabimus*.
> – David Hilbert

When X is the sum of many rare indicator "mostly independent" random variables and $\mu = E[X]$ we would like to say that X is close to a Poisson distribution with mean μ and, in particular, that $\Pr[X = 0]$ is nearly $e^{-\mu}$. We call this rough statement the Poisson Paradigm. In this chapter we give a number of situations in which this Paradigm may be rigorously proven.

8.1 THE JANSON INEQUALITIES

In many instances we would like to bound the probability that none of a set of bad events $B_i, i \in I$ occur. If the events are mutually independent then

$$\Pr[\wedge_{i \in I} \overline{B_i}] = \prod_{i \in I} \Pr[\overline{B_i}].$$

When the B_i are "mostly" independent the Janson Inequalities allow us, sometimes, to say that these two quantities are "nearly" equal.

Let Ω be a finite universal set and let R be a random subset of Ω given by

$$\Pr[r \in R] = p_r,$$

these events mutually independent over $r \in \Omega$. Let $A_i, i \in I$, be subsets of Ω, I a finite index set. Let B_i be the *event* $A_i \subseteq R$. (That is, each point $r \in \Omega$ " flips a coin" to determine if it is in R. B_i is the event that the coins for all $r \in A_i$ came up "heads.") Let X_i be the indicator random variable for B_i and $X = \sum_{i \in I} X_i$ the number of $A_i \subseteq R$. The event $\wedge_{i \in I} \overline{B_i}$ and $X = 0$ are then identical. For $i, j \in I$ we write $i \sim j$ if $i \neq j$ and $A_i \cap A_j \neq \emptyset$. Note that when $i \neq j$ and not $i \sim j$ then B_i, B_j are independent events since they involve separate coin flips. Furthermore, and this plays a crucial role in the proofs, if $i \notin J \subset I$ and not $i \sim j$ for all $j \in J$ then B_i is mutually independent of $\{B_j | j \in J\}$, i.e., independent of any Boolean function of those B_j. This is because the coin flips on A_i and on $\cup_{j \in J} A_j$ are independent. We define

$$\Delta = \sum_{i \sim j} \Pr[B_i \wedge B_j].$$

Here the sum is over ordered pairs so that $\Delta/2$ gives the same sum over unordered pairs. We set

$$M = \prod_{i \in I} \Pr[\overline{B_i}],$$

the value of $\Pr[\wedge_{i \in I} \overline{B_i}]$ if the B_i were independent. Finally, we set

$$\mu = E[X] = \sum_{i \in I} \Pr[B_i].$$

Theorem 8.1.1 [The Janson Inequality] *Let $B_i, i \in I$, Δ, M, μ be as above and assume all $\Pr[B_i] \leq \epsilon$. Then*

$$M \leq \Pr[\wedge_{i \in I} \overline{B_i}] \leq M e^{\frac{1}{1-\epsilon} \frac{\Delta}{2}}$$

and, further,

$$\Pr[\wedge_{i \in I} \overline{B_i}] \leq e^{-\mu + \frac{\Delta}{2}}.$$

For each $i \in I$

$$\Pr[\overline{B_i}] = 1 - \Pr[B_i] \leq e^{-\Pr[B_i]}$$

so, multiplying over $i \in I$,

$$M \leq e^{-\mu}.$$

The two upper bounds for Theorem 8.1.1 are generally quite Similar; we tend to use the second for convenience. In many asymptotic instances a simple calculation gives $M \sim e^{-\mu}$. In particular, this is always the case when $\epsilon = o(1)$ and $\epsilon\mu = o(1)$.

Perhaps the simplest example of Theorem 8.1.1 is the asymptotic probability that $G(n, c/n)$ is triangle-free, given in §10.1. There, as is often the case, $\epsilon = o(1)$, $\Delta = o(1)$ and μ approaches a constant k. In those instances $\Pr[\wedge_{i \in I} \overline{B_i}] \to e^{-k}$. This is no longer the case when Δ becomes large. Indeed, when $\Delta \geq 2\mu$ the upper bound of Theorem 8.1.1 becomes useless. Even for Δ slightly less it is improved by the following result.

Theorem 8.1.2 [The Extended Janson Inequality] *Under the assumptions of Theorem 8.1.1 and the further assumption that* $\Delta \geq \mu$,

$$\Pr[\wedge_{i\in I}\overline{B_i}] \leq e^{-\frac{\mu^2}{2\Delta}}.$$

Theorem 8.1.2 (when it applies) often gives a much stronger result than Chebyschev's Inequality as used in Chapter 4. In §4.3 we saw $\mathrm{Var}[X] \leq \mu + \Delta$ so that

$$\Pr[\wedge_{i\in I}\overline{B_i}] = \Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{E[X]^2} \leq \frac{\mu+\Delta}{\mu^2}.$$

Suppose $\mu \to \infty$, $\mu \ll \Delta$, and $\gamma = \frac{\mu^2}{\Delta} \to \infty$. Chebyschev's upper bound on $\Pr[X = 0]$ is then roughly γ^{-1} while Janson's upper bound is roughly $e^{-\gamma}$.

8.2 THE PROOFS

The original proofs of Janson are based on estimates of the Laplace transform of an appropriate random variable. The proof we present here follows that of Boppana and Spencer (1989). We shall use the inequalities

$$\Pr\left[B_i|\wedge_{j\in J}\overline{B_j}\right] \leq \Pr[B_i]$$

valid for all index sets $J \subset I, i \notin J$ and

$$\Pr[B_i|B_k \wedge \bigwedge_{j\in J}\overline{B_j}] \leq \Pr[B_i|B_k]$$

valid for all index sets $J \subset I, i, k \notin J$. The first follows from Theorem 6.3.2. The second is equivalent to the first since conditioning on B_k is the same as assuming $p_r = \Pr[r \in R] = 1$ for all $r \in A_k$.

Proof [Theorem 8.1.1] The lower bound follows immediately. Order the index set $I = \{1, \ldots, m\}$ for convenience. For $1 \leq i \leq m$,

$$\Pr[B_i|\wedge_{1\leq j<i}\overline{B_j}] \leq \Pr[B_i]$$

so

$$\Pr[\overline{B_i}|\wedge_{1\leq j<i}\overline{B_j}] \geq \Pr[\overline{B_i}]$$

and

$$\Pr[\wedge_{i\in I}\overline{B_i}] = \prod_{i=1}^{m}\Pr[\overline{B_i}|\wedge_{1\leq j<i}\overline{B_j}] \geq \prod_{i=1}^{m}\Pr[\overline{B_i}].$$

Now the first upper bound. For a given i renumber, for convenience, so that $i \sim j$ for $1 \leq j \leq d$ and not for $d+1 \leq j < i$. We use the inequality $\Pr[A|B \wedge C] \geq \Pr[A \wedge B|C]$, valid for any A, B, C. With $A = B_i$, $B = \overline{B_1} \wedge \ldots \wedge \overline{B_d}$, $C = \overline{B_{d+1}} \wedge \ldots \wedge \overline{B_{i-1}}$,

$$\Pr[B_i|\wedge_{1\leq j<i}\overline{B_j}] = \Pr[A|B \wedge C] \geq \Pr[A \wedge B|C] = \Pr[A|C]\Pr[B|A \wedge C].$$

From the mutual independence $\Pr[A|C] = \Pr[A]$. We bound

$$\Pr[B|A \wedge C] \geq 1 - \sum_{j=1}^{d} \Pr[B_j|B_i \wedge C] \geq 1 - \sum_{j=1}^{d} \Pr[B_j|B_i]$$

from the Correlation Inequality. Thus

$$\Pr[B_i| \wedge_{1 \leq j < i} \overline{B_j}] \geq \Pr[B_i] - \sum_{j=1}^{d} \Pr[B_j \wedge B_i].$$

Reversing

$$\begin{aligned} \Pr[\overline{B_i}| \wedge_{1 \leq j < i} \overline{B_j}] &\leq \Pr[\overline{B_i}] + \textstyle\sum_{j=1}^{d} \Pr[B_j \wedge B_i] \\ &\leq \Pr[\overline{B_i}] \left(1 + \tfrac{1}{1-\epsilon} \textstyle\sum_{j=1}^{d} \Pr[B_j \wedge B_i]\right) \end{aligned}$$

since $\Pr[\overline{B_i}] \geq 1 - \epsilon$. Employing the inequality $1 + x \leq e^x$,

$$\Pr[\overline{B_i}| \wedge_{1 \leq j < i} \overline{B_j}] \leq \Pr[\overline{B_i}] e^{\frac{1}{1-\epsilon} \sum_{j=1}^{d} \Pr[B_j \wedge B_i]}.$$

For each $1 \leq i \leq m$ we plug this inequality into

$$\Pr[\wedge_{i \in I} \overline{B_i}] = \prod_{i=1}^{m} \Pr[\overline{B_i}| \wedge_{1 \leq j < i} \overline{B_j}].$$

The terms $\Pr[\overline{B_i}]$ multiply to M. The exponents add: for each $i, j \in I$ with $j < i$ and $j \sim i$ the term $\Pr[B_j \wedge B_i]$ appears once so they add to $\Delta/2$.

For the second upper bound we instead bound

$$\begin{aligned} \Pr[\overline{B_i}| \wedge_{1 \leq j < i} \overline{B_j}] &\leq 1 - \Pr[B_i] + \textstyle\sum_{j=1}^{d} \Pr[B_j \wedge B_i] \\ &\leq \exp\left(-\Pr[B_i] + \textstyle\sum_{j=1}^{d} \Pr[B_j \wedge B_i]\right). \end{aligned}$$

Now the $-\Pr[B_i]$ terms add to $-\mu$ while the $\Pr[B_j \wedge B_i]$ terms again add to $\Delta/2$. ■

Proof [Theorem 8.1.2] The second upper bound of Theorem 8.1.1 may be rewritten

$$-\ln[\Pr[\wedge_{i \in I} \overline{B_i}]] \geq \sum_{i \in I} \Pr[B_i] - \frac{1}{2} \sum_{i \sim j} \Pr[B_i \wedge B_j].$$

For any set of indices $S \subset I$ the same inequality applied only to the $B_i, i \in S$ gives

$$-\ln[\Pr[\wedge_{i \in S} \overline{B_i}]] \geq \sum_{i \in S} \Pr[B_i] - \frac{1}{2} \sum_{i,j \in S, i \sim j} \Pr[B_i \wedge B_j].$$

Let now S be a random subset of I given by

$$\Pr[i \in S] = p,$$

with p a constant to be determined, the events mutually independent. (Here we are using probabilistic methods to prove a probability theorem!) Each term $\Pr[B_i]$ then appears with probability p and each term $\Pr[B_i \wedge B_j]$ with probability p^2 so that

$$\begin{aligned} E\left[-\ln[\Pr[\wedge_{i\in S}\overline{B_i}]\right] &\geq E\left[\textstyle\sum_{i\in S}\Pr[B_i]\right] - \tfrac{1}{2}E\left[\textstyle\sum_{i,j\in S, i\sim j}\Pr[B_i \wedge B_j]\right] \\ &= p\mu - p^2\tfrac{\Delta}{2}. \end{aligned}$$

We set

$$p = \frac{\mu}{\Delta}$$

so as to maximize this quantity. The added assumption of Theorem 8.1.2 assures us that the probability p is at most 1. Then

$$E\left[-\ln[\Pr[\wedge_{i\in S}\overline{B_i}]]\right] \geq \frac{\mu^2}{2\Delta}.$$

Therefore there is a specific $S \subset I$ for which

$$-\ln[\Pr[\wedge_{i\in S}\overline{B_i}]] \geq \frac{\mu^2}{2\Delta}.$$

That is,

$$\Pr[\wedge_{i\in S}\overline{B_i}] \leq e^{-\frac{\mu^2}{2\Delta}}.$$

But

$$\Pr[\wedge_{i\in I}\overline{B_i}] \leq \Pr[\wedge_{i\in S}\overline{B_i}]$$

completing the proof. ■

8.3 BRUN'S SIEVE

The more traditional approach to the Poisson Paradigm is called Brun's Sieve, for its use by the number theorist T. Brun. Let $B_1, \ldots, B_m$ be events, X_i the indicator random variable for B_i and $X = X_1 + \ldots + X_m$ the number of B_i that hold. Let there be a hidden parameter n (so that actually $m = m(n)$, $B_i = B_i(n)$, $X = X(n)$) which will define our o, O notation. Define

$$S^{(r)} = \sum \Pr[B_{i_1} \wedge \ldots \wedge B_{i_r}],$$

the sum over all sets $\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, m\}$. The Inclusion-Exclusion Principle gives that

$$\Pr[X = 0] = \Pr[\overline{B_1} \wedge \ldots \wedge \overline{B_m}] = 1 - S^{(1)} + S^{(2)} - \ldots + (-1)^r S^{(r)} \ldots .$$

Theorem 8.3.1 *Suppose there is a constant μ so that*

$$E[X] = S^{(1)} \to \mu$$

and such that for every fixed r,

$$E[X^{(r)}/r!] = S^{(r)} \to \mu^r/r!.$$

Then

$$\Pr[X = 0] \to e^{-\mu}$$

and indeed for every t

$$\Pr[X = t] \to \frac{\mu^t}{t!}e^{-\mu}.$$

Proof. We do only the case $t = 0$. Fix $\epsilon > 0$. Choose s so that

$$\left|\sum_{r=0}^{2s}(-1)^r\frac{\mu^r}{r!} - e^{-\mu}\right| \leq \frac{\epsilon}{2}.$$

The Bonferroni Inequalities state that, in general, the inclusion-exclusion formula alternately over and underestimates $\Pr[X = 0]$. In particular,

$$\Pr[X = 0] \leq \sum_{r=0}^{2s}(-1)^r S^{(r)}.$$

Select n_0 (the hidden variable) so that for $n \geq n_0$,

$$\left|S^{(r)} - \frac{\mu^r}{r!}\right| \leq \frac{\epsilon}{2(2s+1)}$$

for $0 \leq r \leq 2s$. For such n

$$\Pr[X = 0] \leq e^{-\mu} + \epsilon.$$

Similarly, taking the sum to $2s + 1$ we find n_0 so that for $n \geq n_0$,

$$\Pr[X = 0] \geq e^{-\mu} - \epsilon.$$

As ϵ was arbitrary $\Pr[X = 0] \to e^{-\mu}$. ■

The threshold functions for $G \sim G(n,p)$ to contain a copy of a given graph H, derived in §10.1 via the Janson Inequality, were originally found using Brun's Sieve. Here is an example where both methods are used. Let $G \sim G(n,p)$, the random graph of Chapter 10. Let $EPIT$ represent the statement that *every* vertex lies in a triangle.

Theorem 8.3.2 *Let $c > 0$ be fixed and let $p = p(n)$, $\mu = \mu(n)$ satisfy*

$$\binom{n-1}{2}p^3 = \mu,$$

$$e^{-\mu} = \frac{c}{n}.$$

Then

$$\lim_{n\to\infty} \Pr[G(n,p) \models EPIT] = e^{-c}.$$

In Spencer (1990a) threshold functions are found for a very wide class of "extension statements" that every r vertices lie in a copy of some fixed H.

Proof. First fix $x \in V(G)$. For each unordered $y, z \in V(G) - \{x\}$ let B_{xyz} be the event that $\{x, y, z\}$ is a triangle of G. Let C_x be the event $\wedge\overline{B_{xyz}}$ and X_x the corresponding indicator random variable. We use Janson's Inequality to bound $E[X_x] = \Pr[C_x]$. Here $p = o(1)$ so $\epsilon = o(1)$. $\sum \Pr[B_{xyz}] = \mu$ as defined above. Dependency $xyz \sim xuv$ occurs if and only if the sets overlap (other than in x). Hence

$$\Delta = \sum_{y,z,z'} \Pr[B_{xyz} \wedge B_{xyz'}] = O(n^3)p^5 = o(1)$$

since $p = n^{-2/3+o(1)}$. Thus

$$E[X_x] \sim e^{-\mu} = \frac{c}{n}.$$

Now define

$$X = \sum_{x\in V(G)} X_x,$$

the number of vertices x not lying in a triangle. Then from Linearity of Expectation,

$$E[X] = \sum_{x\in V(G)} E[X_x] \to c.$$

We need to show that the Poisson Paradigm applies to X. Fix r. Then

$$E[X^{(r)}/r!] = S^{(r)} = \sum \Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}],$$

the sum over all sets of vertices $\{x_1, \ldots, x_r\}$. All r-sets look alike so

$$E[X^{(r)}/r!] = \binom{n}{r} \Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}] \sim \frac{n^r}{r!} \Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}]$$

where $x_1, \ldots, x_r$ are some particular vertices. But

$$C_{x_1} \wedge \ldots \wedge C_{x_r} = \wedge\overline{B_{x_i yz}},$$

the conjunction over $1 \le i \le r$ and all y, z. We apply Janson's Inequality to this conjunction. Again $\epsilon = p^3 = o(1)$. The number of $\{x_i, y, z\}$ is $r\binom{n-1}{2} - O(n)$, the overcount coming from those triangles containing two (or three) of the x_i. (Here it is crucial that r is fixed.) Thus

$$\sum \Pr[B_{x_i yz}] = p^3 \left(r\binom{n-1}{2} - O(n) \right) = r\mu + O(n^{-1+o(1)}).$$

As before Δ is p^5 times the number of pairs $x_iyz \sim x_jy'z'$. There are $O(rn^3) = O(n^3)$ terms with $i = j$ and $O(r^2n^2) = O(n^2)$ terms with $i \neq j$ so again $\Delta = o(1)$. Therefore

$$\Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}] \sim e^{-r\mu}$$

and

$$E[X^{(r)}/r!] \sim \frac{(ne^{-\mu})^r}{r!} = \frac{c^r}{r!}.$$

Hence the conditions of Theorem 8.3.1 are met for X. ■

8.4 LARGE DEVIATIONS

We return to the formulation of §8.1. Our object is to derive large deviation results on X similar to those in Appendix A. Given a point in the probability space (i.e., a selection of R) we call an index set $J \subseteq I$ a disjoint family (abbreviated disfam) if

- B_j for every $j \in J$.
- For no $j, j' \in J$ is $j \sim j'$.

If, in addition,

- If $j' \notin J$ and $B_{j'}$ then $j \sim j'$ for some $j \in J$,

then we call J a maximal disjoint family (maxdisfam). We give some general results on the possible sizes of maxdisfams. The connection to X must then be done on an *ad hoc* basis.

Lemma 8.4.1 *With the above notation and for any integer s,*

$$\Pr[\textit{there exists a disfam } J, |J| = s] \leq \frac{\mu^s}{s!}.$$

Proof. Let $\sum^*$ denote the sum over all s-sets $J \subseteq I$ with no $j \sim j'$. Let $\sum^o$ denote the sum over ordered s-tuples $(j_1, \ldots, j_s)$ with $\{j_1, \ldots, j_s\}$ forming such a J. Let $\sum^a$ denote the sum over *all* ordered s-tuples $(j_1, \ldots, j_s)$. Then

$$\Pr[\text{there exists a disfam } J, |J| = s] \leq \sum{}^* \Pr[\wedge_{j \in J} B_j]$$

$$= \sum{}^* \prod_{j \in J} \Pr[B_j] = \tfrac{1}{s!} \sum{}^o \Pr[B_{j_1}] \ldots \Pr[B_{j_s}]$$

$$\leq \tfrac{1}{s!} \sum{}^a \Pr[B_{j_1}] \ldots \Pr[B_{j_s}] \leq \tfrac{1}{s!} [\textstyle\sum_{i \in I} \Pr[B_i]]^s = \mu^s/s!.$$

■

Lemma 8.4.1 gives an effective upper bound when $\mu^s \ll s!$ – basically if $s > \mu\alpha$ for $\alpha > e$. For smaller s we look at the further conditon of J being a *max*disfam. To that end we let μ_s denote the minimum, over all $j_1, \ldots, j_s \in I$ of $\sum \Pr[B_i]$, the sum

taken over all $i \in I$ except those i with $i \sim j_l$ for some $1 \leq l \leq s$. In application s will be small (otherwise we use Lemma 8.4.1) and μ_s will be close to μ. For some applications it is convenient to set

$$\nu = \max_{j \in I} \sum_{i \sim j} \Pr[B_i]$$

and note that $\mu_s \geq \mu - s\nu$.

Lemma 8.4.2 *With the above notation and for any integer s,*

$$\begin{aligned} \Pr[\textit{there exists a maxdisfam } J, |J| = s] &\leq \frac{\mu^s}{s!} e^{-\mu_s} e^{\frac{\Delta}{2}} \\ &\leq \frac{\mu^s}{s!} e^{-\mu} e^{s\nu} e^{\frac{\Delta}{2}}. \end{aligned}$$

Proof. As in Lemma 8.4.1 we bound this probability by $\sum^*$ of $J = \{j_1, \ldots, j_s\}$ being a maxdisfam. For this to occur J must first be a disfam and then $\wedge^* \overline{B_i}$, where $\wedge^*$ is the conjunction over all $i \in I$ except those with $i \sim j_l$ for some $1 \leq l \leq s$. We apply Janson's Inequalityto give an upper bound to $\Pr[\wedge^* \overline{B_i}]$. The associated values μ^*, Δ^* satisfy

$$\mu^* \geq \mu_s,$$

$$\Delta^* \leq \Delta,$$

the latter since Δ^* has simply fewer addends. Thus

$$\Pr[\wedge^* \overline{B_i}] \leq e^{-\mu_s} e^{\frac{\Delta}{2}}$$

and

$$\begin{aligned} \sum^* \Pr[J \text{ maxdisfam}] &\leq e^{-\mu_s} e^{\frac{\Delta}{2}} \sum^* \Pr[\wedge_{j \in J} B_j] \\ &\leq e^{-\mu_s} e^{\frac{\Delta}{2}} \mu^s / s! \,. \end{aligned}$$

■

When $\Delta = o(1)$ and $\nu\mu = o(1)$ or, more generally, $\mu_{3\mu} = \mu + o(1)$, then Lemma 8.4.2 gives a close approximation to the Poisson Distribution since

$$\Pr[\text{there exists a maxdisfam } J, |J| = s] \leq (1 + o(1)) \frac{\mu^s}{s!} e^{-\mu}$$

for $s \leq 3\mu$ and the probability is quite small for larger s by Lemma 8.4.1 .

8.5 COUNTING EXTENSIONS

We begin with a case that uses the basic large deviation results of Appendix A.

Theorem 8.5.1 *Set* $p = \frac{\ln n}{n} \omega(n)$ *where* $\omega(n) \to \infty$ *arbitrarily slowly. Then in* $G(n,p)$ *almost always*

$$\deg(x) \sim (n-1)p$$

for all *vertices* x.

This is actually a large deviation result. It suffices to show the following.

Theorem 8.5.2 *Set* $p = \frac{\ln n}{n}\omega(n)$ *where* $\omega(n) \to \infty$ *arbitrarily slowly. Let* $x \in G$ *be fixed. Fix* $\epsilon > 0$. *Then*

$$\Pr[|\deg(x) - (n-1)p| > \epsilon(n-1)p] = o(n^{-1}).$$

Proof. As $\deg(x) \sim B(n-1,p)$, i.e., it is a Binomial random variable with the above parameters, we have from A.1.14 that

$$\Pr[|\deg(x) - (n-1)p| > \epsilon(n-1)p] < 2e^{-c_\epsilon(n-1)p} = o(n^{-1}),$$

as c_ϵ is fixed and $(n-1)p \gg \ln n$. ■

This result illustrates why logarithmic terms appear so often in the study of Random Graphs. We want *every* x to have a property, hence we try to get the failure probability down to $o(n^{-1})$. When the Poisson Paradigm applies the failure probability is roughly an exponential, and hence we want the exponent to be logarithmic. This often leads to a logarithmic term for the edge probability p.

In §3 we found the threshold function for every vertex to lie on a triangle. It basically occured when the expected number of extensions of a given vertex to a triangle reached $\ln n$. Now set $N(x)$ to be the number of triangles containing x. Set $\mu = \binom{n-1}{2}p^3 = E[N(x)]$.

Theorem 8.5.3 *Let* p *be such that* $\mu \gg \ln n$. *Then almost always*

$$N(x) \sim \mu$$

for all $x \in G(n,p)$.

As above, this is actually a large deviation result. We actually show the following.

Theorem 8.5.4 *Let* p *be such that* $\mu \gg \ln n$. *Let* $x \in G$ *be fixed. Fix* $\epsilon > 0$. *Then*

$$\Pr[|N(x) - \mu| > \epsilon\mu] = o(n^{-1}).$$

Proof. We shall prove this under the further assumption $p = n^{-2/3+o(1)}$ (or, equivalently, $\mu = n^{o(1)}$) which could be removed by technical methods. We now have, in the notation of Lemmas 8.4.1, 8.4.2 $\nu\mu, \Delta = o(1)$. Let P denote the Poisson Distribution with mean μ. Then

$$\Pr[\text{there exists a maxdisfam } J, |J| \le \mu(1-\epsilon)] \le (1+o(1))\Pr[P \le \mu(1-\epsilon)],$$

$$\begin{aligned}&\Pr[\text{there exists a maxdisfam } J, \mu(1+\epsilon) \le |J| \le 3\mu] \\ &\le (1+o(1))\Pr[\mu(1+\epsilon) \le P \le 3\mu]\end{aligned},$$

$$\Pr[\text{there exists a maxdisfam } J, |J| \geq 3\mu]$$
$$\leq \Pr[\text{there exists a disfam } J, |J| \geq 3\mu] \leq \sum_{s=3\mu}^{\infty} \frac{\mu^s}{s!} = O((1-c)^{\mu})$$

where $c > 0$ is an absolute constant. Since $\mu \gg \ln n$ the third term is $o(n^{-1})$. The first and second terms are $o(n^{-1})$ by A.1.15. With probability $1 - o(n^{-1})$ every maxdisfam J has size between $(1-\epsilon)\mu$ and $(1+\epsilon)\mu$.

Fix one such J. (There *always* is some maximal disfam – even if no B_i held we could take $J = \emptyset$.) The elements of J are triples xyz which form triangles, hence $N(x) \geq |J| \geq (1-\epsilon)\mu$. The upper bound is *ad hoc*. The probability that there exist five triangles of the form $xyz_1, xyz_2, xyz_3, xyz_4, xyz_5$ is at most $n^6 p^{11} = o(n^{-1})$. The probability that there exist triangles xy_iz_i, xy_iz_i', $1 \leq i \leq 4$, all vertices distinct is at most $n^{12}p^{20} = o(n^{-1})$. Consider the graph whose vertices are the triangles xyz, with $\sim$ giving the edge relation. There are $N(x)$ vertices, the maxdisfam J are the maximal independent sets. In this graph, with probability $1 - o(n^{-1})$, each vertex xyz has degree at most nine and there is no set of four disjoint edges. This implies that for any J, $|J| \geq N(x) - 27$ and

$$N(x) \leq (1+\epsilon)\mu + 27 \leq (1+\epsilon')\mu.$$

■

For any graph H with "roots" $x_1, \ldots, x_r$ we can examine in $G(n,p)$ the number of extensions $N(x_1, \ldots, x_r)$ of a given set of r vertices to a copy of H. In Spencer (1990b) some general results are given that generalize Theorems 8.5.2, 8.5.4. Under fairly wide assumptions (see Exercise 5, Chapter 10), when the expected number μ of extensions satisfies $\mu \gg \ln n$ then almost always all $N(x_1, \ldots, x_r) \sim \mu$.

8.6 COUNTING REPRESENTATIONS

The results of this section shall use the following very basic and very useful result.

Lemma 8.6.1 [The Borel-Cantelli Lemma] *Let $A_n, n \in N$ be events with*

$$\sum_{n=1}^{\infty} \Pr[A_n] < \infty.$$

Then

$$\Pr[\bigwedge_{i=1}^{\infty} \bigvee_{j=i}^{\infty} A_j] = 0.$$

That is, almost always A_n is false for all sufficiently large n. In application we shall aim for $\Pr[A_n] < n^{-c}$ with $c > 1$ in order to apply this Lemma.

Again we begin with a case that involves only the Large Deviation results of Appendix A. For a given set S of natural numbers let (for every $n \in N$) $f(n) = f_S(n)$ denote the number of representations $n = x + y$, $x, y \in S, x < y$.

Theorem 8.6.2 [Erdős (1956)] *There is a set S for which $f(n) = \Theta(\ln n)$. That is, there is a set S and constants c_1, c_2 so that for all sufficiently large n*

$$c_1 \ln n \leq f(n) \leq c_2 \ln n.$$

Proof. Define S randomly by

$$\Pr[x \in S] = p_x = \min\left[10\sqrt{\frac{\ln x}{x}}, 1\right].$$

Fix n. Now $f(n)$ is a random variable with mean

$$\mu = E[f(n)] = \frac{1}{2}\sum_{x+y=n, x\neq y} p_x p_y.$$

Roughly there are n addends with $p_x p_y > p_n^2 = 100\frac{\ln n}{n}$. We have $p_x p_x = \Theta(\frac{\ln n}{n})$ except in the regions $x = o(n), y = o(n)$ and care must be taken that those terms don't contribute significantly to μ. Careful asymptotics (and first year Calculus!) yield

$$\mu \sim (50 \ln n)\int_0^1 \frac{dx}{\sqrt{x(1-x)}} = 50\pi \ln n.$$

The negligible effect of the $x = o(n), y = o(n)$ terms reflects the finiteness of the indefinite integral at poles $x = 0$ and $x = 1$. The possible representations $x + y = n$ are mutually independent events so that from A.1.14,

$$\Pr[|f(n) - \mu| > \epsilon\mu] < 2e^{-\delta\mu}$$

for constants $\epsilon, \delta = \delta(\epsilon)$. To be specific we can take $\epsilon = 0.9, \delta = 0.1$ and

$$\Pr[|f(n) - \mu| > 0.9\mu] < 2e^{-5\pi \ln n} < n^{-1.1}$$

for n sufficiently large. Take $c_1 < 0.1(50\pi)$ and $c_2 > 1.9(50\pi)$.

Let A_n be the event that $c_1 \ln n \leq f(n) \leq c_2 \ln n$ does *not* hold. We have $\Pr[A_n] < n^{-1.1}$ for n sufficiently large. The Borel-Cantelli Lemma applies, almost always all A_n fail for n sufficiently large. Therefore there exists a specific point in the probability space, i.e., a specific set S, for which $c_1 \ln n \leq f(n) \leq c_2 \ln n$ for all sufficiently large n. ■

The development of the infinite probability space used here, and below, has been carefully done in the book *Sequences* by H. Halberstam and K. F. Roth.

The use of the infinite probability space leaves a number of questions about the existential nature of the proof that go beyond the algorithmic. For example, does there exist a recursive set S having the property of Theorem 8.6.2? An affirmative answer is given in Kolountzakis (1999).

Now for a given set S of natural numbers let $g(n) = g_S(n)$ denote the number of representations $n = x + y + z$, $x, y, z \in S$, $x < y < z$. The following result was

actually proven for representations of n as the sum of k terms for any fixed k. For simplicity we present here only the proof for $k = 3$.

Theorem 8.6.3 [Erdős and Tetali (1990)] *There is a set S for which $g(n) = \Theta(\ln n)$. That is, there is a set S and constants c_1, c_2 so that for all sufficiently large n,*

$$c_1 \ln n \leq g(n) \leq c_2 \ln n.$$

Proof. Define S randomly by

$$\Pr[x \in S] = p_x = \min\left[10\left(\frac{\ln x}{x^2}\right)^{1/3}, \frac{1}{2}\right].$$

Fix n. Now $g(n)$ is a random variable and

$$\mu = E[g(n)] = \sum_{x+y+z=n} p_x p_y p_z.$$

Careful asymptotics give

$$\mu \sim \frac{10^3}{6} \ln n \int_{x=0}^{1} \int_{y=0}^{1-x} \frac{dxdy}{[xy(1-x-y)]^{2/3}} = K \ln n,$$

where K is large. (We may make K arbitrarily large by increasing "10.") We apply Lemma 8.4.2. Here

$$\Delta = \sum p_x p_y p_z p_{y'} p_{z'},$$

the sum over all five-tuples with $x + y + z = x + y' + z' = n$. Roughly there are n^3 terms, each $\sim p_n^5 = n^{-10/3+o(1)}$ so that the sum is $o(1)$. Again, care must be taken that those terms with one (or more) small variables don't contribute much to the sum. We bound $s \leq 3\mu = \Theta(\ln n)$ and consider μ_s. This is the minimal possible $\sum p_x p_y p_z$ over all those x, y, z with $x + y + z = n$ that do not intersect a given s representations; let us weaken that and say a given set of $3s$ elements. Again one needs that the weight of $\sum_{x+y+z=n} p_x p_y p_z$ is not on the edges but "spread" in the center and one shows $\mu_s \sim \mu$. Now, as in §8.5, let P denote the Poisson distibution with mean μ. The probability that there exists a maxdisfam J of size less than $\mu(1-\epsilon)$ or between $\mu(1+\epsilon)$ and 3μ is asymptotically the probability that P lies in that range. For moderate ϵ, as K is large, these – as well as the probability of having a disfam of size bigger than 3μ – will be $o(n^{-c})$ with $c > 1$. By the Borel-Cantelli Lemma almost always all sufficiently large n will have all maxdisfam J of size between $c_1 \ln n$ and $c_2 \ln n$. Then $g(n) \geq c_1 \ln n$ immediately.

The upper bound is again *ad hoc*. With this p let $f(n)$ be, as before, the number of representations of n as the sum of two elements of S. We use only that $p_x = x^{-2/3+o(1)}$. We calculate

$$E[f(n)] = \sum_{x+y=n} (xy)^{-2/3+o(1)} = n^{-1/3+o(1)},$$

again watching the "pole" at 0. Here the possible representations are mutually independent so

$$\Pr[f(n) \geq 4] \leq E[f(n)]^4/4! = n^{-4/3+o(1)},$$

and by the Borel-Cantelli Lemma almost always $f(n) \leq 3$ for all sufficiently large n. But then almost always there is a C so that $f(n) \leq C$ for all n. For all sufficiently large n there is a maxdisfam (with representations as the sum of three terms) of size less than $c_2 \ln n$. Every triple $x, y, z \in S$ with $x+y+z = n$ must contain at least one of these at most $3c_2 \ln n$ points. The number of triples $x, y, z \in S$ with $x+y+z = n$ for a particular x is simply $f(n-x)$, the number of representations $n - x = y + z$ (possibly one less since $y, z \neq x$), and so is at most C. But then there are at most $C(3c_2 \ln n)$ total representations $n = x + y + z$. ■

8.7 FURTHER INEQUALITIES

Here we discuss some further results that allow one, sometimes, to apply the Poisson Paradigm. Let $B_i, i \in I$ be events in an arbitrary probability space. As in the Lovász Local Lemma of Chapter 5 we say that a symmetric binary relation $\sim$ on I is a *dependency digraph* if for each $i \in I$ the event B_i is mutually independent of $\{B_j | \text{not } i \sim j\}$. [The digraph of Chapter 5, §5.1 has $E = \{(i,j) | i \sim j\}$.] Suppose the events B_i satisfy the inequalities of §8.2:

$$\Pr[B_i | \wedge_{j \in J} \overline{B_j}] \leq \Pr[B_i]$$

valid for all index sets $J \subset I, i \notin J$ and

$$\Pr\left[B_i | B_k \wedge \bigwedge_{j \in J} \overline{B_j}\right] \leq \Pr[B_i | B_k]$$

valid for all index sets $J \subset I, i, k \notin J$. Then the Janson inequalities Theorems 8.1.1 and 8.1.2 and also Lemmas 8.4.1 and 8.4.2 hold as stated. The proofs are identical, the above are the only properties of the events B_i that were used.

Suen (1990) [see also Janson (1998) for significant variations] has given a very general result that allows the approximation of $\Pr[\wedge_{i \in I} \overline{B_i}]$ by $M = \prod_{i \in I} \Pr[\overline{B_i}]$. Again let $B_i, i \in I$ be events in an arbitrary probability space. We say that a binary relation $\sim$ on I is a *superdependency digraph* if the following holds: Let $J_1, J_2 \subset I$ be disjoint subsets so that $j_1 \sim j_2$ for no $j_1 \in J_1, j_2 \in J_2$. Let B^1 be any Boolean combination of the events $B_j, j \in J_1$ and let B^2 be any Boolean combination of the events $B_j, j \in J_2$. Then B^1, B^2 are independent. Note that the $\sim$ of §8.1 is indeed a superdependency digraph.

Theorem 8.7.1 [Suen] *Under the above conditions,*

$$\left|\Pr[\wedge_{i \in I} \overline{B_i}] - M\right| \leq M\left[e^{\sum_{i \sim j} y(i,j)} - 1\right]$$

where

$$y(i,j) = (\Pr[B_i \wedge B_j] + \Pr[B_i]\Pr[B_j]) \prod_{l \sim i \text{ or } l \sim j} (1 - \Pr[B_l])^{-1}.$$

We shall not prove Theorem 8.7.1. In many instances the above product is not large. Suppose it is less than two for all $i \sim j$. In that instance

$$\sum_{i \sim j} y(i,j) \leq 2[\Delta + \sum_{i \sim j} \Pr[B_i]\Pr[B_j]].$$

In many instances $\sum_{i \sim j} \Pr[B_i]\Pr[B_j]$ is small relative to Δ (as in many instances when $i \sim j$ the events B_i, B_j are positively correlated). When, furthermore, $\Delta = o(1)$ Suen's Theorem gives the approximation of $\Pr[\wedge_{i \in I} \overline{B_i}]$ by M. Suen has applied this result to examinations of the number of *induced* copies of a fixed graph H in the random $G(n,p)$.

Janson (1990) has given a one-way large deviation result on the X of §8.1 which is somewhat simpler to apply than Lemmas 8.4.1 and 8.4.2.

Theorem 8.7.2 [Janson] *With $\mu = E[X]$ and $\gamma > 0$ arbitrary,*

$$\Pr[X \leq (1-\gamma)\mu] < e^{-\gamma^2 \mu/(2+\frac{\Delta}{\mu})}.$$

When $\Delta = o(\mu)$ this bound on the tail approximates that of the normal curve with mean and standard deviation μ. We shall not prove Theorem 8.7.2 here. The proofs of Theorems 8.7.1 and 8.7.2 as well as the original proofs by Janson of Theorems 8.1.1 and 8.1.2 are based on estimations of the Laplace transform of X, bounding $E[e^{-tX}]$.

8.8 EXERCISES

1. Prove that for every $\epsilon > 0$ there is some $n_0 = n_0(\epsilon)$ so that for every $n > n_0$ there is a graph on n vertices containing every graph on $k \leq (2-\epsilon)\log_2 n$ vertices as an induced subgraph.

2. Find a threshold function for the property: $G(n,p)$ contains at least $n/6$ pairwise vertex disjoint triangles.

THE PROBABILISTIC LENS:
Local Coloring

This result of Erdős (1962) gives further probabilistic evidence that the chromatic number $\chi(G)$ cannot be deduced from local considerations.

Theorem 1 *For all k there exists $\epsilon > 0$ so that for all sufficiently large n there exist graphs G on n vertices with $\chi(G) > k$ and yet $\chi(G|_S) \leq 3$ for every set S of vertices of size at most ϵn.*

Proof. For a given k let $c, \epsilon > 0$ satisfy (with foresight)

$$\begin{aligned} c &> 2k^2 H(1/k) \ln 2, \\ \epsilon &< e^{-5} 3^3 c^{-3}, \end{aligned}$$

where $H(x) = -x \log_2 x - (1-x)\log_2(1-x)$ is the entropy function. Set $p = c/n$ and let $G \sim G(n,p)$. We show that G almost surely satisfies the two conditions of the Theorem.

If $\chi(G) \leq k$ there would be an independent set of size n/k. The expected number of such sets is

$$\binom{n}{n/k}(1-p)^{\binom{n/k}{2}} < 2^{n(H(1/k)+o(1))} e^{-cn/2k^2(1+o(1))}$$

which is $o(1)$ by our condition on c. Hence almost surely $\chi(G) > k$.

Suppose some set S with $t \leq \epsilon n$ vertices required at least four colors. Then as in the proof of Lemma 7.3.4 there would be a minimal such set S. For any $v \in S$ there would be a three-coloring of $S - \{v\}$. If v had two or fewer neighbors in S then this could be extended to a three-coloring of S. Hence every $v \in S$ would have degree

at least three in $G|_S$ and so $G|_S$ would have at least $3t/2$ edges. The probability that some $t \leq \epsilon n$ vertices have at least $3t/2$ edges is less than

$$\sum_{t \leq \epsilon n} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} \left(\frac{c}{n}\right)^{3t/2}.$$

We outline the analysis. When $t = O(1)$ the terms are negligible. Otherwise we bound each term from above by

$$\left[\frac{ne}{t}\left(\frac{te}{3}\right)^{3/2}\left(\frac{c}{n}\right)^{3/2}\right]^t \leq \left[e^{5/2}3^{-3/2}c^{3/2}\sqrt{t/n}\right]^t.$$

Now since $t \leq \epsilon n$ the bracketed term is at most $e^{5/2}3^{-3/2}c^{3/2}\epsilon^{1/2}$ which is less than one by our condition on ϵ. The full sum is $o(1)$, i.e., almost surely no such S exists. ■

Many tempting conjectures are easily *dis*proved by the Probabilistic Method. If every $n/(\ln n)$ vertices may be three-colored then can a graph G on n vertices be four-colored? This result shows that the answer is no.

This page intentionally left blank

9

Pseudorandomness

'A knot!', said Alice, already to make herself useful, and looking anxiously about her. 'Oh, do let me help to undo it!'
– from *Alice in Wonderland*, by Lewis Carroll

As shown in the various chapters of this book, the probabilistic method is a powerful tool for establishing the existence of combinatorial structures with certain properties. It is often the case that such an existence proof is not sufficient; we actually prefer an *explicit construction*. This is not only because an explicit construction may shed more light on the corresponding problem, but also because it often happens that a random-looking structure is useful for a certain algorithmic procedure; in this case we would like to have an algorithm and not merely to prove that it exists.

The problem of finding explicit constructions may look trivial; after all, since we are mainly dealing with finite cases, once we have a probabilistic proof of existence we can find an explicit example by exhaustive search. Moreover, many of the probabilistic proofs of existence actually show that most members of a properly chosen random space have the desired properties. We may thus expect that it would not be too difficult to find one such member. Although this is true in principle, it is certainly not practical to check all possibilities; it is thus common to define an explicit construction of a combinatorial object as one that can be performed efficiently; say, in time which is polynomial in the parameters of the object.

Let us illustrate this notion by one of the best known open problems in the area of explicit constructions, the problem of constructing explicit *Ramsey graphs*. The first example given in Chapter 1 is the proof of Erdős that for every n there are graphs on n vertices containing neither a clique nor an independent set on $2 \log_2 n$ vertices. This proof is an existence proof; can we actually describe such graphs explicitly? Erdős

offered a prize of $500 for the explicit construction of an infinite family of graphs, in which there is neither a clique nor an independent set of size more than a constant times the logarithm of the number of vertices, for some absolute constant. Of course, we can, in principle, for every fixed n, check all graphs on n vertices until we find a good one, but this does not give an efficient way of producing the desired graphs and hence is not explicit. Although the problem mentioned above received a considerable amount of attention, it is still open. The best known explicit construction is due to Frankl and Wilson (1981), who describe explicit graphs on n vertices which contain neither a clique nor an independent set on more than $2^{c\sqrt{\log n \log\log n}}$ vertices, for some absolute positive constant c.

Although the problem of constructing explicit Ramsey graphs is still open, there are several other problems for which explicit constructions are known. In this chapter we present a few examples and discuss briefly some of their algorithmic applications. We also describe several seemingly unrelated properties of a graph, which all turn out to be equivalent. All these are properties of the random graph and it is thus common to call a graph that satisfies these properties *quasi random*. The equivalence of all these properties enables one to show, in several cases, that certain explicit graphs have many pseudorandom properties by merely showing that they posses one of them.

9.1 THE QUADRATIC RESIDUE TOURNAMENTS

Recall that a *tournament* on a set V of n players is an orientation $T = (V, E)$ of the set of edges of the complete graph on the set of vertices V. If (x, y) is a directed edge we say that x *beats* y. Given a permutation π of the set of players, a (directed) edge (x, y) of the tournament is *consistent* with π if x precedes y in π. If π is viewed as a ranking of the players, then it is reasonable to try and find rankings with as many consistent arcs as possible. Let $c(\pi, T)$ denote the number of arcs of T which are consistent with π, and define $c(T) = \max(c(\pi, T))$, where the maximum is taken over all permutations π of the set of vertices of T. For every tournament T on n players, if $\pi = 1, 2, \ldots, n$ and and $\pi' = n, n-1, \ldots, 1$ then $c(\pi, T) + c(\pi', T) = \binom{n}{2}$. Therefore $c(T) \geq \frac{1}{2}\binom{n}{2}$. In fact, it can be shown that for every such T, $c(T) \geq \frac{1}{2}\binom{n}{2} + \Omega(n^{3/2})$. On the other hand, a simple probabilistic argument shows that there are tournaments T on n players for which $c(T) \leq (1 + o(1))\frac{1}{2}\binom{n}{2}$. [The best known estimate, which gives the right order of magnitude for the largest possible value of the difference of $c(T) - \frac{1}{2}\binom{n}{2}$ is more complicated and was given by de la Vega in 1983, where he showed that there are tournaments T on n players for which $c(T) \leq \frac{1}{2}\binom{n}{2} + O(n^{3/2})$.]

Can we describe explicitly tournaments T on n vertices in which $c(T) \leq (1 + o(1))\frac{1}{2}\binom{n}{2}$? This problem was mentioned by Erdős and Moon (1965) and by Spencer (1985b). It turns out that several such constructions can be given. Let us describe one.

Let $p \equiv 3 \pmod 4$ be a prime and let $T = T_p$ be the tournament whose vertices are all elements of the finite field $GF(p)$ in which (i, j) is a directed edge iff $i - j$ is

a quadratic residue. (Since $p \equiv 3 \pmod 4$, -1 is a quadratic nonresidue modulo p and hence T_p is a well-defined tournament).

Theorem 9.1.1 *For the tournaments T_p described above,*

$$c(T_p) \leq \frac{1}{2}\binom{p}{2} + O(p^{3/2}\log p).$$

In order to prove this theorem we need some preparations. Let χ be the quadratic residue character defined on the elements of the finite field $GF(p)$ by $\chi(y) = y^{(p-1)/2}$. Equivalently, $\chi(y)$ is 1 if y is a nonzero square, 0 if y is 0 and -1 otherwise. Let $D = (d_{ij})_{i,j=0}^{p-1}$ be the p by p matrix defined by $d_{ij} = \chi(i-j)$.

Fact 1 *For every two distinct j and l, $\sum_{i \in GF(p)} d_{ij}d_{il} = -1$.*

Proof.

$$\begin{aligned}\textstyle\sum_i d_{ij}d_{il} &= \textstyle\sum_i \chi(i-j)\chi(i-l) = \sum_{i \neq j,l} \chi(i-j)\chi(i-l) \\ &= \textstyle\sum_{i \neq j,l} \chi((i-j)/(i-l)) = \sum_{i \neq j,l} \chi 1 + (l-j)/(i-l)).\end{aligned}$$

As i ranges over all elements of $GF(p)$ besides j and l the quantity $(1 + (l - j)/(i-l))$ ranges over all elements of $GF(p)$ besides 0 and 1. Since the sum of $\chi(r)$ over all r in $GF(p)$ is 0 this implies that the right-hand side of the last equation is $0 - \chi(0) - \chi(1) = -1$, completing the proof of the fact. ■

For two subsets A and B of $GF(p)$, let $e(A,B)$ denote the number of directed edges of T_p that start in a vertex of A and end in a vertex of B. By the definition of the matrix D it follows that

$$\sum_{i \in A}\sum_{j \in B} d_{ij} = e(A,B) - e(B,A).$$

The following lemma is proved in Alon (1986b).

Lemma 9.1.2 *For any two subsets A and B of $GF(p)$;*

$$|\sum_{i \in A}\sum_{j \in B} d_{ij}| \leq |A|^{1/2}|B|^{1/2}p^{1/2}.$$

Proof. By the Cauchy-Schwarz Inequality and by the fact above,

$$\begin{aligned}(\textstyle\sum_{i \in A}\sum_{j \in B} d_{ij})^2 &\leq |A|(\textstyle\sum_{i \in A}(\sum_{j \in B} d_{ij})^2) \\ &\leq |A|(\textstyle\sum_{i \in GF(p)}(\sum_{j \in B} d_{ij})^2) \\ &= |A|(\textstyle\sum_{i \in GF(p)}(|B| + 2\sum_{j<l \in B} d_{ij}d_{il})) \\ &= |A||B|p + 2|A|\textstyle\sum_{j<l \in B}\sum_{i \in GF(P)} d_{ij}d_{il} \\ &= |A||B|p - |A||B|(|B|-1) = |A||B|(p - |B| + 1) \\ &\leq |A||B|p,\end{aligned}$$

completing the proof of the lemma. ■

Proof [Theorem 9.1.1] Let r be the smallest integer satisfying $2^r \geq p$. Let $\pi = \pi_1, \ldots, \pi_p$ be an arbitrary permutation of the vertices of T_p, and define $\pi' = \pi_p, \ldots, \pi_1$. We must show that $c(\pi, T_p) \leq \frac{1}{2}\binom{p}{2} + O(p^{3/2} \log p)$, or equivalently, that $c(\pi, T_p) - c(\pi', T_p) \leq O(p^{3/2} \log p)$. Let a_1 and a_2 be two integers satisfying $p = a_1 + a_2$ and $a_1 \leq 2^{r-1}, a_2 \leq 2^{r-1}$. Let A_1 be the set of the first a_1 vertices in the permutation π and let A_2 be the set of the last a_2 vertices in π. By Lemma 9.1.2,

$$e(A_1, A_2) - e(A_2, A_1) \leq (a_1 a_2 p)^{1/2} \leq 2^{r-1} p^{1/2}.$$

Next, let $a_{11}, a_{12}, a_{21}, a_{22}$ be integers each of which does not exceed 2^{r-2} such that $a_1 = a_{11} + a_{12}$ and $a_2 = a_{21} + a_{22}$. Let A_{11} be the subset of A_1 consisting of those a_{11} elements of A_1 that appear first in π, and let A_{12} be the set of the a_{12} remaining elements of A_1. The partition of A_2 into the two sets A_{21} and A_{22} is defined similarly. By applying Lemma 9.1.2 we obtain

$$\begin{aligned} & e(A_{11}, A_{12}) - e(A_{12}, A_{11}) + e(A_{21}, A_{22}) - e(A_{22}, A_{21}) \\ & \leq (a_{11} a_{12} p)^{1/2} + (a_{21} a_{22} p)^{1/2} \\ & \leq 2 \cdot 2^{r-2} p^{1/2}. \end{aligned}$$

Continuing in the same manner we obtain, in the i-th step, a partition of the set of vertices into 2^i blocks, each consisting of at most 2^{r-i} consecutive elements in the permutation π. This partition is obtained by splitting each block in the partition corresponding to the previous step into two parts. By applying Lemma 9.1.2 to each such pair $A_{\epsilon 1}, A_{\epsilon 2}$, (where here ϵ is a vector of length $i-1$ with $\{1,2\}$-entries), and by summing we conclude that the sum over all these 2^{i-1} vectors ϵ of the differences $e(A_{\epsilon 1}, A_{\epsilon 2}) - e(A_{\epsilon 2}, A_{\epsilon 1})$ does not exceed

$$2^{i-1} 2^{r-i} p^{1/2} \leq 2^{r-1} p^{1/2}.$$

Observe that the sum of the left-hand sides of all these inequalities as i ranges from 1 to r is precisely the difference $c(\pi, T_p) - c(\pi', T_p)$. Therefore, by summing we obtain

$$c(\pi, T_p) - c(\pi', T_p) \leq 2^{r-1} p^{1/2} r = O(p^{3/2} \log p),$$

completing the proof. ■

We note that any antisymmetric matrix with $\{1, -1\}$-entries in which each two rows are roughly orthogonal can be used to give a construction of a tournament as above. Some related results appear in Frankl, Rödl and Wilson (1988). The tournaments T_p, however, have stronger pseudorandom properties than do some of these other tournaments. For example, for every $k \leq \frac{1}{4} \log p$, and for every set S of k vertices of T_p, the number of vertices of T_p that beat all the members of S is $(1 + o(1))p/2^k$. This was proved by Graham and Spencer (1971) by applying Weil's

famous theorem known as the Riemann hypotheses for curves over finite fields [Weil (1948)]. Taking a sufficiently large p this supplies an explicit construction for the Schütte problem mentioned in Chapter 1.

9.2 EIGENVALUES AND EXPANDERS

A graph $G = (V, E)$ is called an (n, d, c)-*expander* if it has n vertices, the maximum degree of a vertex is d, and for every set of vertices $W \subset V$ of cardinality $|W| \leq n/2$, the inequality $|N(W)| \geq c|W|$ holds, where $N(W)$ denotes the set of all vertices in $V \setminus W$ adjacent to some vertex in W. We note that sometimes a slightly different definition is used, but the difference is not essential. Expanders share many of the properties of sparse random graphs, and are the subject of an extensive literature. A family of *linear expanders of density d and expansion c* is a sequence $\{G_i\}_{i=1}^{\infty}$, where G_i is an (n_i, d, c)-expander and n_i tends to infinity as i tends to infinity.

Such a family is the main component of the parallel sorting network of Ajtai, Komlós and Szemerédi (1983), and can be used for constructing certain fault tolerant linear arrays. It also forms the basic building block used in the construction of graphs with special connectivity properties and small number of edges. Some other examples of the numerous applications of these graphs to various problems in theoretical computer science can be found in, e.g., Alon (1986b) and its references.

It is not too difficult to prove the existence of a family of linear expanders using probabilistic arguments. This was first done by Pinsker (1973). An explicit construction is much more difficult to find, and was first given by Margulis (1973). This construction was later improved by various authors; most known constructions are Cayley graphs of certain groups of matrices, and their expansion properties are proved by estimating the eigenvalues of the adjacency matrices of the graphs and by relying on the close correspondence between the expansion properties of a graph and its spectral properties. This correspondence was first studied, independently, by Tanner (1984) and by Alon and Milman (1984). Since it is somewhat simpler for the case of regular graphs we restrict our attention here to this case.

Let $G = (V, E)$ be a d-regular graph and let $A = A_G = (a_{uv})_{u,v \in V}$ be its adjacency matrix given by $a_{uv} = 1$ if $uv \in E$ and $a_{uv} = 0$ otherwise. Since G is d-regular the largest eigenvalue of A is d, corresponding to the all 1 eigenvector. Let $\lambda = \lambda(G)$ denote the second largest eigenvalue of G. For two (not necessarily disjoint) subsets B and C of V let $e(B, C)$ denote the number of ordered pairs (u, v), where $u \in B$, $v \in C$ and uv is an edge of G. (Note that if B and C are disjoint this is simply the number of edges of G that connect a vertex of B with a vertex of C.)

Theorem 9.2.1 *For every partition of the set of vertices V into two disjoint subsets B and C,*

$$e(B, C) \geq \frac{(d - \lambda)|B||C|}{n}.$$

Proof. Put $|V| = n, b = |B|, c = |C| = n - b$. Let $D = dI$ be the n by n scalar matrix with the degree of regularity of G on its diagonal. Observe that for any real

vector x of length n (considered as a function $x : V \mapsto R$) we have

$$((D-A)x,x) = \sum_{u\in V}(d(x(u))^2 - \sum_{v;uv\in E} x(v)x(u))$$
$$= d\sum_{u\in V}(x(u))^2 - 2\sum_{uv\in E} x(v)x(u) = \sum_{uv\in E}(x(v)-x(u))^2.$$

Define, now, a vector x by $x(v) = -c$ if $v \in B$ and $x(v) = b$ if $v \in C$. Notice that A and $D-A$ have the same eigenvectors, and that the eigenvalues of $D-A$ are precisely $d-\mu$, as μ ranges over all eigenvalues of A. Note, also, that $\sum_{v\in V} x(v) = 0$; i.e., x is orthogonal to the eigenvector of the smallest eigenvalue of $D-A$. Since $D-A$ is a symmetric matrix its eigenvectors are orthogonal to each other and form a basis of the n-dimensional space. It follows that x is a linear combination of the other eigenvectors of $D-A$ and hence, by the definition of λ and the fact that $d-\lambda$ is the second smallest eigenvalue of $D-A$ we conclude that $((D-A)x,x) \geq (d-\lambda)(x,x) = (d-\lambda)(bc^2+cb^2) = (d-\lambda)bcn$.

By the second paragraph of the proof the left-hand side of the last inequality is $\sum_{uv\in E}(x(u)-x(v))^2 = e(B,C)\cdot(b+c)^2 = e(B,C)\cdot n^2$. Thus

$$e(B,C) \geq \frac{(d-\lambda)bc}{n},$$

completing the proof. ■

Corollary 9.2.2 *If λ is the second largest eigenvalue of a d-regular graph G with n vertices, then G is an (n,d,c)-expander for $c = \frac{d-\lambda}{2d}$.*

Proof. Let W be a set of $w \leq n/2$ vertices of G. By Theorem 9.2.1 there are at least $\frac{(d-\lambda)w(n-w)}{n} \geq \frac{(d-\lambda)w}{2}$ edges from W to its complement. Since no vertex in the complement is adjacent to more than d of these edges it follows that $|N(W)| \geq \frac{(d-\lambda)w}{2d}$. ■

The estimate for c in the last corollary can in fact be improved to $\frac{2(d-\lambda)}{3d-2\lambda}$, as shown by Alon and Milman (1984). Each of these estimates shows that if the second largest eigenvalue of G is far from the first, then G is a good expander. The converse of this is also true, although more complicated. This is given in the following result, proved in Alon (1986a), which we state without its proof.

Theorem 9.2.3 *If G is a d-regular graph which is an (n,d,c)-expander then $\lambda(G) \leq d - \frac{c^2}{4+2c^2}$.*

The last two results supply an efficient algorithm for approximating the expanding properties of a d-regular graph; we simply compute (or estimate) its second largest eigenvalue. The larger the difference between this eigenvalue and d is, the better expanding properties of G follow. It is thus natural to ask how far from d this second eigenvalue can be. It is known [see Nilli (1991)] that the second largest eigenvalue of any d-regular graph with diameter k is at least $2\sqrt{d-1}(1-O(1/k))$. Therefore, in any infinite family of d-regular graphs, the limsup of the second largest eigenvalue is at least $2\sqrt{d-1}$. Lubotzky, Phillips and Sarnak (1986), and independently,

Margulis (1988), gave, for every $d = p + 1$ where p is a prime congruent to 1 modulo 4, explicit constructions of infinite families of d-regular graphs G_i with second largest eigenvalues $\lambda(G_i) \leq 2\sqrt{d-1}$. These graphs are Cayley graphs of factor groups of the group of all two by two invertible matrices over a finite field, and their eigenvalues are estimated by applying results of Eichler and Igusa concerning the Ramanujan conjecture. Eichler's proof relies on Weil's theorem mentioned in the previous section. The nonbipartite graphs G constructed in this manner satisfy a somewhat stronger assertion than $\lambda(G) \leq 2\sqrt{d-1}$. In fact, besides their largest eigenvalue d, they do not have eigenvalues whose absolute value exceed $2\sqrt{d-1}$. This fact implies some strong pseudo-random properties, as shown in the next results.

Theorem 9.2.4 *Let $G = (V, E)$ be a d-regular graph on n vertices, and suppose the absolute value of each of its eigenvalues but the first one is at most λ. For a vertex $v \in V$ and a subset B of V denote by $N(v)$ the set of all neighbors of v in G, and let $N_B(v) = N(v) \cap B$ denote the set of all neighbors of v in B. Then, for every subset B of cardinality bn of V,*

$$\sum_{v \in V} (|N_B(v)| - bd)^2 \leq \lambda^2 b(1-b)n.$$

Observe that in a random d-regular graph each vertex v would tend to have about bd neighbors in each set of size bn. The above theorem shows that if λ is much smaller than d then for most vertices v, $N_B(v)$ is not too far from bd.

Proof. Let A be the adjacency matrix of G and define a vector $f : V \mapsto R$ by $f(v) = 1 - b$ for $v \in B$ and $f(v) = -b$ for $v \notin B$. Clearly $\sum_{v \in V} f(v) = 0$; i.e., f is orthogonal to the eigenvector of the largest eigenvalue of A. Therefore

$$(Af, Af) \leq \lambda^2 (f, f).$$

The right-hand side of the last inequality is $\lambda^2(bn(1-b)^2 + (1-b)nb^2) = \lambda^2 b(1-b)n$. The left-hand side is

$$\sum_{v \in V} ((1-b)|N_B(v)| - b(d - |N_B(v)|))^2 = \sum_{v \in V} (|N_B(v)| - bd)^2.$$

The desired result follows. ■

Corollary 9.2.5 *Let $G = (V, E), d, n$ and λ be as in Theorem 9.2.4. Then for every two sets of vertices B and C of G, where $|B| = bn$ and $|C| = cn$ we have*

$$|e(B, C) - cbdn| \leq \lambda \sqrt{bc}\, n.$$

Proof. By Theorem 9.2.4,

$$\sum_{v \in C} (|N_B(v)| - bd)^2 \leq \sum_{v \in V} (|N_B(v)| - bd)^2 \leq \lambda^2 b(1-b)n.$$

Thus, by the Cauchy-Schwarz Inequality,

$$|e(B,C) - cbdn| \leq \sum_{v \in C} |N_B(v) - bd|$$

$$\leq \sqrt{cn}(\sum_{v \in C}(|N_B(v)| - bd)^2)^{1/2} \leq \sqrt{cn}\lambda\sqrt{b(1-b)n} \leq \lambda\sqrt{bc}\, n.$$

■

The special case $B = C$ gives the following result. A slightly stronger estimate is proved in a similar way in Alon and Chung (1988).

Corollary 9.2.6 *Let $G = (V, E), d, n$ and λ be as in Theorem 9.2.4. Let B be an arbitrary set of bn vertices of G and let $e(B) = \frac{1}{2}e(B, B)$ be the number of edges in the induced subgraph of G on B. Then*

$$|e(B) - \frac{1}{2}b^2dn| \leq \frac{1}{2}\lambda bn.$$

A *walk of length l* in a graph G is a sequence $v_0, \ldots, v_l$ of vertices of G, where for each $1 \leq i \leq l$, $v_{i-1}v_i$ is an edge of G. Obviously, the total number of walks of length l in a d-regular graph on n vertices is precisely $n \cdot d^l$. Suppose, now , that C is a subset of, say, $n/2$ vertices of G. How many of these walks do not contain any vertex of C? If G is disconnected it may happen that half of these walks avoid C. However, as shown byAjtai, Komlós and Szemerédi (1987), there are many fewer such walks if all the eigenvalues of G but the largest are small. This result and some of its extensions have several applications in theoretical computer science, as shown in the above-mentioned paper (see also Cohen and Wigderson (1989)). We conclude this section by stating and proving the result and one of its applications.

Theorem 9.2.7 *Let $G = (V, E)$ be a d-regular graph on n vertices, and suppose that each of its eigenvalues but the first one is at most λ. Let C be a set of cn vertices of G. Then, for every l, the number of walks of length l in G that avoid C does not exceed $(1-c)n((1-c)d + c\lambda))^l$.*

Proof. Let A be the adjacency matrix of G and let A' be the adjacency matrix of its induced subgraph on the complement of C. We claim that the maximum eigenvalue of A' is at most $(1-c)d + c\lambda$. To prove this claim we must show that for every vector $f : V \mapsto R$ satisfying $f(v) = 0$ for each $v \in C$ and $\sum_{v \in V} f(v)^2 = 1$, the inequality $(Af, f) \leq (1-c)d + c\lambda$ holds. Let $f_1, f_2, \ldots, f_n$ be an orthonormal basis of eigenvectors of A, where f_1 is the eigenvector of λ_i, $\lambda_1 = d$ and each entry of f_1 is $1/\sqrt{n}$. Then $f = \sum_{i=1}^n c_i f_i$, where $\sum_{i=1}^n c_i^2 = 1$ and

$$c_1 = \sum_{v \in V} f(v)/\sqrt{n} = \sum_{v \in V-C} f(v)/\sqrt{n}$$

$$\leq (\sum_{v \in V-C} f(v)^2)^{1/2}((1-c)n/n)^{1/2} = \sqrt{1-c}\,,$$

where here we used the Cauchy-Schwarz Inequality. Therefore $\sum_{i=2}^{n} c_i^2 = c$ and

$$(Af, f) = \sum_{i=1}^{n} c_i^2 \lambda_i \leq (1-c)d + c\lambda,$$

supplying the desired estimate for the largest eigenvalue of A'.

Let $\gamma_1 \geq \gamma_2 \ldots \geq \gamma_m$ be the eigenvalues of A', where $m = (1-c)n$. By the Perron-Frobenius Theorem it follows that the absolute value of each of them is at most $\gamma_1 \leq (1-c)d + c\lambda$. The total number of walks of length l that avoid C is precisely $(A'^l g, g)$, where g is the all 1-vector indexed by the vertices in $V - C$. By expressing g as a linear combination of the eigenvectors of A', $g = \sum_{i=1}^{m} b_i g_i$ where g_i is the eigenvector of γ_i, we conclude that this number is precisely

$$\sum_{i=1}^{m} b_i^2 \gamma_i^l \leq \gamma_1^l \sum_{i=1}^{m} b_i^2 = m\gamma_1^l \leq m((1-c)d + c\lambda)^l.$$

Substituting $m = (1-c)n$ the desired result follows. ■

A *randomly chosen walk* of length l in a graph G is a walk of length l in G chosen according to a uniform distribution among all walks of that length. Notice that if G is d-regular such a walk can be chosen by choosing randomly its starting point v_0, and then by choosing, for each $1 \leq i \leq l$, v_i randomly among the d neighbors of v_{i-1}.

Corollary 9.2.8 *Let $G = (V, E), d, n, \lambda, C$ and c be as in Theorem 9.2.7 and suppose*

$$(1-c)d + c\lambda \leq \frac{d}{\sqrt{2}}.$$

Then, for every l, the probability that a randomly chosen walk of length l in G avoids C is at most $2^{-l/2}$.

Proof. The number of walks of length l in G that avoid C is at most $(1-c)n((1-c)d + c\lambda)^l \leq nd^l 2^{-l/2}$, by Theorem 9.2.7. Since the total number of walks is nd^l, the desired result follows. ■

The results above are useful for amplification of probabilities in randomized algorithms. Although such an amplification can be achieved for any Monte Carlo algorithm we prefer, for simplicity, to consider one representative example: the primality testing algorithm of Rabin (1980).

For an odd integer q, define two integers a and b by $q - 1 = 2^a b$, where b is odd. An integer x, $1 \leq x \leq q-1$ is called a *witness* (for the nonprimality of q) if for the sequence $x_0, \ldots, x_a$ defined by $x_0 = x^b \pmod q$ and $x_i = x_{i-1}^2 \pmod q$ for $1 \leq i \leq a$, either $x_a \neq 1$ or there is an i such that $x_i \neq -1, 1$ and $x_{i+1} = 1$. One can show that if q is a prime then there are no such witnesses for q, whereas if q is an odd nonprime then at least half of the numbers between 1 and $q-1$ are witnesses for q. (In fact, at least $3/4$ are witnesses, as shown by Rabin). This suggests the following randomized algorithm for testing if an odd integer q is a prime (for even integers there is a simpler algorithm !):

Choose, randomly, an integer x between 1 and $q-1$ and check if it is a witness. If it is, report that q is not a prime. Otherwise, report that q is a prime.

Observe that if q is a prime, the algorithm certainly reports it is a prime, whereas if q is not a prime, the probability that the algorithm makes a mistake and reports it as a prime is at most $1/2$. What if we wish to reduce the probability of making such a mistake? Clearly, we can simply repeat the algorithm. If we repeat it l independent times, then the probability of making an error (i.e., reporting a nonprime as a prime) decreases to $1/2^l$. However, the number of random bits required for this procedure is $l \cdot \log(q-1)$.

Suppose we wish to use fewer random bits. By applying the properties of a randomly chosen walk on an appropriate graph, proved in the last two results, we can obtain the same estimate for the error probability by using only $\log(q-1) + O(l)$ random bits. This is done as follows.

Let G be a d-regular graph with $q-1$ vertices, labelled by all integers between 1 and $q-1$. Suppose G has no eigenvalue, but the first one, that exceeds λ and suppose that

$$\frac{d+\lambda}{2} \leq \frac{d}{\sqrt{2}}. \tag{9.1}$$

Now choose randomly a walk of length $2l$ in the graph G, and check, for each of the numbers labelling its vertices, if it is a witness. If q is a nonprime, then at least half of the vertices of G are labelled by witnesses. Hence, by Corollary 9.2.8 and by (9.1), the probability that no witness is on the walk is at most $2^{-2l/2} = 2^{-l}$. Thus we obtain the same reduction in the error-probability as the one obtained by choosing l independent witnesses. Let us estimate the number of random bits required for choosing such a random walk.

The known constructions of expanders given by Lubotzky et al. (1986) or by Margulis (1988) give explicit families of graphs with degree d and with $\lambda \leq 2\sqrt{d-1}$, for each $d = p+1$, where p is a prime congruent to 1 modulo 4. [We note that these graphs will not have exactly $q-1$ vertices but this does not cause any real problem as we can take a graph with n vertices, where $q-1 \leq n \leq (1+o(1))(q-1)$, and label its i-th vertex by $i \pmod{(q-1)}$. In this case the number of vertices labelled by witnesses would still be at least $(\frac{1}{2} + o(1))n$.) One can easily check that, e.g., $d = 30$ and $\lambda = 2\sqrt{29}$ satisfy (9.1), and thus we can use a 30-regular graph. The number of random bits required for choosing a random walk of length $2l$ in it is less than $\log(q-1) + 10l + 1$, much less than the $l \log(q-1)$ bits which are needed in the repetition procedure.

9.3 QUASI RANDOM GRAPHS

In this section we describe several pseudorandom properties of graphs which, somewhat surprisingly, turn out to be all equivalent. All the properties are ones satisfied, almost surely, by a random graph in which every edge is chosen, independently, with probability $1/2$. The equivalence between some of these properties were first proved by several authors; see Thomason (1987), Frankl et al. (1988) and Alon and Chung

(1988), but the first paper in which all of them (and some others) appear is the one by Chung, Graham and Wilson (1989). Our presentation here follows that paper, although, in order to simplify the presentation, we consider only the case of regular graphs.

We first need some notation. For two graphs G and H, let $N_G^*(H)$ be the number of labelled occurrences of H as an induced subgraph of G (i.e., the number of adjacency preserving injections $f : V(H) \mapsto V(G)$ whose image is the set of vertices of an induced copy of H in G.) Similarly, $N_G(H)$ denotes the number of labelled copies of H as a (not necessarily induced) subgraph of G. Note that $N_G(H) = \sum_L N_G^*(L)$, where L ranges over all graphs on the set of vertices of H obtained from H by adding to it a (possibly empty) set of edges.

Throughout this section G always denotes a graph with n vertices. We denote the eigenvalues of its adjacency matrix (taken with multiplicities) by $\lambda_1, \ldots, \lambda_n$, where $|\lambda_1| \geq \ldots \geq |\lambda_n|$. [Since we consider in this section only the eigenvalues of G we simply write λ_1 and not $\lambda_1(G)$.] Recall also the following notation, used in the previous section: for a vertex v of G, $N(v)$ denotes the set of its neighbors in G. If S is a set of vertices of G, $e(S)$ denotes the number of edges in the induced subgraph of G on S. If B and C are two (not necessarily disjoint) subsets of vertices of G, $e(B, C)$ denotes the number of ordered pairs (b, c) where $b \in B, c \in C$, and bc is an edge of G. Thus $e(S) = \frac{1}{2}e(S, S)$.

We can now state the pseudorandom properties considered here. All the properties refer to a graph $G = (V, E)$ with n vertices. Throughout the section, we use the $o(\cdot)$-notation, without mentioning the precise behavior of each $o(\cdot)$. Thus, occurrences of two $o(1)$, say, need not mean that both are identical and only mean that if we consider a family of graphs G and let their number of vertices n tend to infinity then each $o(1)$ tends to 0.

Property $P_1(s)$**:** For every graph $H(s)$ on s vertices

$$N_G^*(H(s)) = (1 + o(1))n^s 2^{-\binom{s}{2}}.$$

Property P_2**:** For the cycle $C(4)$ with 4 vertices $N_G(C(4)) \leq (1 + o(1))(n/2)^4$.

Property P_3**:** $|\lambda_2| = o(n)$.

Property P_4**:** For every set S of vertices of G, $e(S) = \frac{1}{4}|S|^2 + o(n^2)$.

Property P_5**:** For every two sets of vertices B and C; $e(B, C) = \frac{1}{2}|B||C| + o(n^2)$.

Property P_6**:** $\sum_{u,v \in V} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| = o(n^3)$.

It is easy to check that all the properties above are satisfied, almost surely, by a random graph on n vertices. In this section we show that all these properties are equivalent for a regular graph with n vertices and degree of regularity about $n/2$. The fact that the innocent-looking property P_2 is strong enough to imply for such graphs $P_1(s)$ for every $s \geq 1$ is one of the interesting special cases of this result.

Graphs that satisfy any (and thus all) of the properties above are called *quasi random*. As noted above the assumption that G is regular can be dropped (at the expense of slightly modifying property P_2 and slightly complicating the proofs).

Theorem 9.3.1 *Let G be a d-regular graph on n vertices, where $d = (\frac{1}{2} + o(1))n$.*

If G satisfies any one of the seven properties $P_1(4), P_1(s)$ for all $s \geq 1$, P_2, P_3, P_4, P_5, P_6 then it satisfies all seven.

Proof. We show that

$$\begin{aligned} P_1(4) &\Longrightarrow P_2 \Longrightarrow P_3 \Longrightarrow P_4 \Longrightarrow P_5 \\ &\Longrightarrow P_6 \Longrightarrow P_1(s) \textit{ for all } s \geq 1 \ (\Longrightarrow P_1(4)). \end{aligned}$$

1. $P_1(4) \Longrightarrow P_2$.
Suppose G satisfies $P_1(4)$. Then $N_G(C(4)) = \sum_L N_G^*(L)$, as L ranges over the four labelled graphs obtained from a labelled $C(4)$ by adding to it a (possibly empty) set of edges. Since G satisfies $P_1(4)$, $N_G^*(L) = (1 + o(1))n^4 2^{-16}$ for each of these graphs L and hence $N_G(C(4)) = (1 + o(1))n^4 2^{-4}$, showing that G satisfies P_2.
2. $P_2 \Longrightarrow P_3$.
Suppose G satisfies P_2 and let A be its adjacency matrix. The trace of A^4 is precisely $\sum_{i=1}^n \lambda_i^4$. On the other hand it is easy to see that this trace is precisely the number of (labelled) closed walks of length 4 in G, i.e., the number of sequences $v_0, v_1, v_2, v_3, v_4 = v_0$ of vertices of G such that $v_i v_{i+1}$ is an edge for each $0 \leq i \leq 3$. This number is $N_G((C(4))$ plus the number of such sequences in which $v_2 = v_0$, which is nd^2, plus the number of such sequences in which $v_2 \neq v_0$ and $v_3 = v_1$, which is $nd(d-1)$. Thus

$$\begin{aligned} \sum_{i=1}^n \lambda_i^4 &= d^4 + \sum_{i=2}^n \lambda_i^4 \\ &= (1 + o(1))(n/2)^4 + \sum_{i=2}^n \lambda_i^4 = N_G(C(4)) + O(n^3) \\ &= (1 + o(1))(n/2)^4. \end{aligned}$$

It follows that $\sum_{i=2}^n \lambda_i^4 = o(n^4)$, and hence that $|\lambda_2| = o(n)$, as needed.
3. $P_3 \Longrightarrow P_4$.
This is an immediate consequence of Corollary 9.2.6.
4. $P_4 \Longrightarrow P_5$.
Suppose G satisfies P_4. We first claim that it satisfies property P_5 for disjoint sets of vertices B and C. Indeed, if B and C are disjoint then

$$\begin{aligned} e(B,C) &= e(B \cup C) - e(B) - e(C) \\ &= \tfrac{1}{4}(|B| + |C|)^2 - \tfrac{1}{4}|B|^2 - \tfrac{1}{4}|C|^2 + o(n^2) = \tfrac{1}{2}|B||C| + o(n^2), \end{aligned}$$

proving the claim.

In case B and C are not disjoint we have

$$e(B,C) = e(B \setminus C, C \setminus B) + e(B \cap C, C \setminus B) + e(B \cap C, B \setminus C) + 2e(B \cap C).$$

Put $|B| = b, |C| = c, |B \cap C| = x$. By the above expression for $e(B,C)$ and by the fact that G satisfies P_4 and P_5 for disjoint B and C we get

$$\begin{aligned} e(B,C) &= \tfrac{1}{2}(b-x)(c-x) + \tfrac{1}{2}x(c-x) + \tfrac{1}{2}x(b-x) + \tfrac{2}{4}x^2 + o(n^2) \\ &= \tfrac{1}{2}bc + o(n^2) = \tfrac{1}{2}|B||C| + o(n^2), \end{aligned}$$

showing that G satisfies P_5.
5. $P_5 \Longrightarrow P_6$.

Suppose that G satisfies P_5 and recall that G is d-regular, where $d = (\frac{1}{2} + o(1))n$. Let v be a fixed vertex of G, and let us estimate the sum

$$\sum_{u \in V, u \neq v} \left| |N(u) \cap N(v)| - \frac{n}{4} \right|.$$

Define

$$B_1 = \left\{ u \in V, u \neq v : |N(u) \cap N(v)| \geq \frac{n}{4} \right\},$$

and similarly

$$B_2 = \left\{ u \in V, u \neq v : |N(u) \cap N(v)| < \frac{n}{4} \right\}.$$

Let C be the set of all neighbors of v in G. Observe that

$$\begin{aligned} &\textstyle\sum_{u \in B_1} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| \\ &= \textstyle\sum_{u \in B_1} |N(u) \cap N(v)| - |B_1| \frac{n}{4} \\ &= e(B_1, C) - |B_1| \frac{n}{4}. \end{aligned}$$

Since G satisfies P_5 and since $d = (\frac{1}{2} + o(1))n$ the last difference is $\frac{1}{2}|B_1|d + o(n^2) - |B_1|\frac{n}{4} = o(n^2)$.

A similar argument implies that

$$\sum_{u \in B_2} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| = o(n^2).$$

It follows that for every vertex v of G,

$$\sum_{u \in V, u \neq v} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| = o(n^2),$$

and by summing over all vertices v we conclude that G satisfies property P_6.

6. $P_6 \Longrightarrow P_1(s)$ *for all* $s \geq 1$.

Suppose $G = (V, E)$ satisfies P_6. For any two distinct vertices u and v of G let $a(u, v)$ be 1 if $uv \in E$ and 0 otherwise. Also, define $s(u, v) = |\{w \in V : a(u, w) = a(v, w)\}|$. Since G is $d = (\frac{1}{2} + o(1))n$-regular, $s(u, v) = 2|N(u) \cap N(v)| + n - 2d = 2|N(u) \cap N(v)| + o(n)$. Therefore, the fact that G satisfies P_6 implies that

$$\sum_{u,v \in V} \left| s(u, v) - \frac{n}{2} \right| = o(n^3). \tag{9.2}$$

Let $H = H(s)$ be an arbitrary fixed graph on s vertices, and put $N_s = N_G^*(H(s))$. We must show that

$$N_s = (1 + o(1))n^s 2^{-\binom{s}{2}}.$$

Denote the vertex set of $H(s)$ by $\{v_1, \ldots, v_s\}$. For each $1 \leq r \leq s$, put $V_r = \{v_1, \ldots, v_r\}$, and let $H(r)$ be the induced subgraph of H on V_r. We prove, by induction on r, that for $N_r = N_G^*(H(r))$,

$$N_r = (1 + o(1))n_{(r)} 2^{-\binom{r}{2}}, \tag{9.3}$$

where $n_{(r)} = n(n-1)\cdots(n-r+1)$.

This is trivial for $r = 1$. Assuming it holds for r, where $1 \leq r < s$, we prove it for $r+1$. For a vector $\alpha = (\alpha_1, \ldots, \alpha_r)$ of distinct vertices of G, and for a vector $\epsilon = (\epsilon_1, \ldots, \epsilon_r)$ of $(0,1)$-entries, define

$$f_r(\alpha, \epsilon) = |\{v \in V : v \neq \alpha_1, \ldots \alpha_r \text{ and } a(v, \alpha_j) = \epsilon_j \text{ for all } 1 \leq j \leq r\}|.$$

Clearly N_{r+1} is the sum of the N_r quantities $f_r(\alpha, \epsilon)$ in which $\epsilon_j = a(v_{r+1}, v_j)$ and α ranges over all N_r induced copies of $H(r)$ in G.

Observe that altogether there are precisely $n_{(r)}2^r$ quantities $f_r(\alpha, \epsilon)$. It is convenient to view $f_r(\alpha, \epsilon)$ as a random variable defined on a sample space of $n_{(r)}2^r$ points, each having an equal probability. To complete the proof we compute the expectation and the variance of this random variable. We show that the variance is so small that most of the quantities $f_r(\alpha, \epsilon)$ are very close to the expectation, and thus obtain a sufficiently accurate estimate for N_{r+1} which is the sum of N_r such quantities.

We start with the simple computation of the expectation $E(f_r)$ of $f_r(\alpha, \epsilon)$. We have

$$\begin{aligned} E(f_r) &= \frac{1}{n_{(r)}2^r} \sum_{\alpha,\epsilon} f_r(\alpha, \epsilon) = \frac{1}{n_{(r)}2^r} \sum_{\alpha} \sum_{\epsilon} f_r(\alpha, \epsilon) \\ &= \frac{1}{n_{(r)}2^r} \sum_{\alpha} (n-r) = \frac{n-r}{2^r}, \end{aligned}$$

where we used the fact that every vertex $v \neq \alpha_1, \ldots, \alpha_r$ defines ϵ uniquely.

Next, we estimate the quantity S_r defined by

$$S_r = \sum_{\alpha,\epsilon} f_r(\alpha, \epsilon)(f_r(\alpha, \epsilon) - 1).$$

We claim that

$$S_r = \sum_{u \neq v} s(u, v)_{(r)}. \tag{9.4}$$

To prove this claim, observe that S_r can be interpreted as the number of ordered triples $(\alpha, \epsilon, (u, v))$, where $\alpha = (\alpha_1, \ldots, \alpha_r)$ is an ordered set of r distinct vertices of G, $\epsilon = (\epsilon_1, \ldots, \epsilon_r)$ is a binary vector of length r, and u, v is an ordered pair of additional vertices of G so that

$$a(u, \alpha_k) = a(v, \alpha_k) = \epsilon_k \text{ for all } 1 \leq k \leq r.$$

For each fixed α and ϵ, there are precisely $f_r(\alpha, \epsilon)(f_r(\alpha, \epsilon) - 1)$ choices for the pair (u, v) and hence S_r counts the number of these triples.

Now, let us compute this number by first choosing u and v. Once u, v are chosen, the additional vertices $\alpha_1, \ldots, \alpha_r$ must all belong to the set $\{w \in V : a(u, w) = a(v, w)\}$. Since the cardinality of this set is $s(u, v)$ it follows that there are $s(u, v)_{(r)}$ choices for $\alpha_1, \ldots \alpha_r$. Once these are chosen the vector ϵ is determined and thus (9.4) follows.

We next claim that (9.2) implies

$$\sum_{u \neq v} s(u, v)_{(r)} = (1 + o(1))n^{r+2}2^{-r}. \tag{9.5}$$

To prove this claim define $\epsilon_{uv} = s(u,v) - \frac{n}{2}$. Observe that by (9.2), $\sum_{u \neq v} |\epsilon_{uv}| = o(n^3)$, and that $|\epsilon_{uv}| \leq n/2 \leq n$ for each u, v. Hence, for every fixed $a \geq 1$,

$$\sum_{u \neq v} |\epsilon_{uv}|^a \leq n^{a-1} \sum_{u \neq v} |\epsilon_{uv}| = o(n^{a+2}).$$

This implies that

$$\sum_{u \neq v} s(u,v)_{(r)} =$$

$$\sum_{u \neq v} (\tfrac{n}{2} + \epsilon_{uv})_{(r)}$$

$$= \sum_{k=0}^{r} \sum_{u \neq v} c_k (\tfrac{n}{2})^k \epsilon_{uv}^{r-k} \quad \text{(for appropriate constants } c_k\text{)}$$

$$= (\tfrac{n}{2})^r n_{(2)} + \sum_{k=0}^{r-1} \sum_{u \neq v} c_k (\tfrac{n}{2})^k \epsilon_{uv}^{r-k}$$

$$\leq (\tfrac{n}{2})^r n_{(2)} + \sum_{k=0}^{r-1} \sum_{u \neq v} |c_k| n^k |\epsilon_{uv}|^{r-k}$$

$$\leq n^{r+2} 2^{-r} + c \sum_{k=0}^{r-1} n^k \sum_{u \neq v} |\epsilon_{uv}|^{r-k} \quad \text{(for an appropriate constant } c\text{)}$$

$$\leq n^{r+2} 2^{-r} + c \sum_{k=0}^{r-1} n^k \cdot o(n^{r-k+2})$$

$$= n^{r+2} 2^{-r} (1 + o(1)),$$

implying (9.5).

By (9.4) and (9.5)

$$S_r = (1 + o(1)) n^{r+2} 2^{-r}.$$

Therefore,

$$\sum_{\alpha,\epsilon} (f_r(\alpha,\epsilon) - E(f_r))^2$$

$$= \sum_{\alpha,\epsilon} f_r^2(\alpha,\epsilon) - \sum_{\alpha,\epsilon} E(f_r)^2$$

$$\sum_{\alpha,\epsilon} (f_r^2(\alpha,\epsilon) - f_r(\alpha,\epsilon)) + \sum_{\alpha,\epsilon} f_r(\alpha,\epsilon) - n_{(r)} 2^r (n-r)^2 2^{-2r}$$

$$= S_r + n_{(r)} 2^r E(f_r) - n_{(r)} 2^r (n-r)^2 2^{-2r}$$

$$= S_r + n_{(r+1)} - n_{(r)} 2^r (n-r)^2 2^{-2r} = o(n^{r+2}).$$

Recall that N_{r+1} is the summation of N_r quantities of the form $f_r(\alpha,\epsilon)$. Thus:

$$|N_{r+1} - N_r E(f_r)|^2 = \left| \sum_{N_r \text{ terms}} (f_r(\alpha,\epsilon) - E(f_r)) \right|^2.$$

By Cauchy-Schwarz, the last expression is at most

$$N_r \sum_{N_r \text{ terms}} (f_r(\alpha,\epsilon) - E(f_r))^2 \\ \leq N_r \sum_{\alpha,\epsilon} (f_r(\alpha,\epsilon) - E(f_r))^2 \\ = N_r \cdot o(n^{r+2}) = o(n^{2r+2}).$$

It follows that

$$|N_{r+1} - N_r E(f_r)| = o(n^{r+1}),$$

and hence, by the induction hypothesis,

$$N_{r+1} = N_r E(f_r) + o(n^{r+1}) \\ = (1+o(1)) n_{(r)} 2^{-\binom{r}{2}} \cdot (n-r) 2^{-r} + o(n^{r+1}) \\ = (1+o(1)) n_{(r+1)} 2^{-\binom{r+1}{2}}.$$

This completes the proof of the induction step and establishes Theorem 9.3.1. ■

There are many examples of families of quasi random graphs. The most widely used is probably the family of Paley graphs G_p defined as follows. For a prime p congruent to 1 modulo 4, let G_p be the graph whose vertices are the integers $0, 1, 2, \ldots, p-1$ in which i and j are adjacent if and only if $i - j$ is a quadratic residue modulo p. The graphs G_p, which are the undirected analogues of the quadratic residue tournaments discussed in §9.1, are $(p-1)/2$-regular. For any two distinct vertices i and j of G_p, the number of vertices k which are either adjacent to both i and j or nonadjacent to both is precisely the number of times the quotient $\frac{k-i}{k-j}$ is a quadratic residue. As k ranges over all numbers between 0 and $p-1$ but i and j, this quotient ranges over all numbers but 1 and 0 and hence it is a quadratic residue precisely $\frac{1}{2}(p-1) - 1$ times. (This is essentially the same assertion as that of the first fact given in the proof of Theorem 9.1.1) We have thus shown that for every two vertices i and j of G_p, $s(i,j) = (p-3)/2$, and this, together with the fact that G_p is $(p-1)/2$-regular, easily implies that it satisfies Property P_6. Therefore it is quasi random. As is the case with the quadratic residue tournaments , G_p satisfies, in fact, some stronger pseudorandom properties which are not satisfied by every quasi random graph, and which can be proved by applying Weil's Theorem.

9.4 EXERCISES

1. By considering a random bipartite three-regular graph on $2n$ vertices obtained by picking three random permutations between the two color classes, prove that there is a $c > 0$ such that for every n there exists a $(2n, 3, c)$-expander.
2. Let $G = (V, E)$ be an (n, d, λ)-graph, suppose n is divisible by k, and let $C : V \mapsto \{1, 2, \ldots, k\}$ be a coloring of V by k colors, so that each color appears precisely n/k times. Prove that there is a vertex of G which has a neighbor of each of the k colors, provided $k\lambda \leq d$.
3. Let $G = (V, E)$ be a graph in which there is at least one edge between any two disjoint sets of size $a+1$. Prove that for every set Y of $5a$ vertices, there is a set

X of at most a vertices, such that for every set Z satisfying $Z \cap (X \cup Y) = \emptyset$ and $|Z| \leq a$, the inequality $|N(Z) \cap Y| \geq 2|Z|$ holds.

4. Prove that for every $\epsilon > 0$ there exists an $n_0 = n_0(\epsilon)$ so that for every $(n, n/2, 2\sqrt{n})$- graph $G = (V, E)$ with $n > n_0$, the number of triangles M in G satisfies $|M - n^3/48| \leq \epsilon n^3$.

THE PROBABILISTIC LENS:
Random Walks

A *vertex-transitive* graph is a graph $G = (V, E)$ such that for any two vertices $u, v \in V$ there is an automorphism of G that maps u into v. A *random walk* of length l in G starting at a vertex v is a randomly chosen sequence $v = v_0, v_1, \ldots, v_l$, where each v_{i+1} is chosen, randomly and independently, among the neighbours of v_i $(0 \leq i < l)$.

The following theorem states that for every vertex-transitive graph G, the probability that a random walk of even length in G ends at its starting point is at least as big as the probability that it ends at any other vertex. Note that the proof requires almost no computation. We note also that the result does not hold for general regular graphs, and the vertex transitivity assumption is necessary.

Theorem 1 *Let $G = (V, E)$ be a vertex-transitive graph. For an integer k and for two (not neccessarily distinct) vertices u, v of G, let $P^k(u, v)$ denote the probability that a random walk of length k starting at u ends at v. Then, for every integer k and for every two vertices $u, v \in V$,*

$$P^{2k}(u, u) \geq P^{2k}(u, v).$$

Proof. We need the following simple inequality, sometimes attributed to Chebyschev.

Claim 9.4.1 *For every sequence $(a_1, \ldots, a_n)$ of n reals and for any permutation π of $\{1, \ldots, n\}$,*

$$\sum_{i=1}^{n} a_i a_{\pi(i)} \leq \sum_{i=1}^{n} a_i^2.$$

Proof. The inequaility follows immediately from the fact that

$$\sum_{i=1}^{n} a_i^2 - \sum_{i=1}^{n} a_i a_{\pi(i)} = \frac{1}{2}\sum_{i=1}^{n}(a_i - a_{\pi(i)})^2 \geq 0.$$

■

Consider, now, a random walk of length $2k$ starting at u. By summing over all the possibilities of the vertex the walk reaches after k steps we conclude that for every vertex v:

$$P^{2k}(u,v) = \sum_{w \in V} P^k(u,w)P^k(w,v) = \sum_{w \in V} P^k(u,w)P^k(v,w), \qquad (1)$$

where the last equality follows from the fact that G is an undirected regular graph.

Since G is vertex-transitive, the two vectors $(P^k(u,w))_{w \in V}$ and $(P^k(v,w))_{w \in V}$ can be obtained from each other by permuting the coordinates. Therefore, by the claim above, the maximum possible value of the sum in the right-hand side of (1) is when $u = v$, completing the proof of the theorem. ■