THE PROBABILISTIC METHOD

THE PROBABILISTIC METHOD

Second edition, March 2000, Tel Aviv and New York

Noga Alon,
Department of Mathematics,
Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel Aviv University,

Tel Aviv, Israel.

Joel H. Spencer,
Courant Institute of Mathematical Sciences,
New York University,

New York, USA

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.

New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

To Nurit and Mary Ann

Preface

The Probabilistic Method has recently been developed intensively and became one of the most powerful and widely used tools applied in Combinatorics. One of the major reasons for this rapid development is the important role of randomness in Theoretical Computer Science, a field which is recently the source of many intriguing combinatorial problems.

The interplay between Discrete Mathemtics and Computer Science suggests an algorithmic point of view in the study of the Probabilistic Method in Combinatorics and this is the approach we tried to adopt in this book. The manuscript thus includes a discussion of algorithmic techniques together with a study of the classical method as well as the modern tools applied in it. The first part of the book contains a description of the tools applied in probabilistic arguments, including the basic techniques that use expectation and variance, as well as the more recent applications of Martingales and Correlation Inequalities. The second part includes a study of various topics in which probabilistic techniques have been successful. This part contains chapters on discrepancy and random graphs, as well as on several areas in Theoretical Computer Science; Circuit Complexity , Computational Geometry, and Derandomization of randomized algorithms. Scattered between the chapters are gems described under the heading "The Probabilistic Lens". These are elegant proofs that are not necessarily related to the chapters after which they appear and can be usually read separately.

The basic Probabilistic Method can be described as follows: in order to prove the existence of a combinatorial structure with certain properties, we construct an appropriate probability space and show that a randomly chosen element in this space has the desired properties with positive probability. This method has been initiated by Paul Erdős, who contributed so much to its development over the last fifty years, that it seems appropriate to call it "The Erdős Method". His contribution cannot be measured only by his numerous deep results in the subject, but also by his many intriguing problems and conjectures that stimulated a big portion of the research in the area.

It seems impossible to write an encyclopedic book on the Probabilistic Method; too many recent interesting results apply probabilistic arguments, and we do not even try to mention all of them. Our emphasis is on methodology, and we thus try to describe the ideas, and not always to give the best possible results if these are too technical to allow a clear presentation. Many of the results are asymptotic, and we use the standard asymptotic notation: for two functions f and g, we write $f = O(g)$ if $f \leq cg$ for all sufficiently large values of the variables of the two functions, where c is an absolute positive constant. We write $f = \Omega(g)$ if $g = O(f)$ and $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$. If the limit of the ratio f/g tends to zero as the variables of the functions tend to infinity we write $f = o(g)$. Finally, $f \sim g$ denotes that $f = (1 + o(1))g$, i.e., that f/g tends to 1 when the variables tend to infinity. Each chapter ends with a list of exercises. The more difficult ones are marked by a $(*)$. The exercises, which have been added to this new edition of the book, enable the reader to check his/her understanding of the material, and also provide the possibility of using the manuscript as a textbook.

Besides these exercises, the second edition contains several improved results and covers various topics that have not been discussed in the first edition. The additions include a continuous approach to discrete probabilistic problems described in Chapter 3, various novel concentration inequalities introduced in Chapter 7, a discussion of the relation between discrepancy and VC-dimension in Chapter 13 and several combinatorial applications of the entropy function and its properties described in Chapter 14. Further additions are the final two probabilistic lenses and the new extensive appendix on Paul Erdős, his papers, conjectures and personality.

It is a special pleasure to thank our wives, Nurit and Mary Ann. Their patience, understanding and encouragment have been a key-ingredient in the success of this enterprise.

NOGA ALON, JOEL H. SPENCER

Acknowledgments

We are very grateful to all our students and colleagues who contributed to the creation of this second edition by joint research, helpful discussions and useful comments. These include Greg Bachelis, Amir Dembo, Ehud Friedgut, Marc Fossorier, Dong Fu, Svante Janson, Guy Kortzers, Michael Krivelevich, Albert Li, Bojan Mohar, Janos Pach, Yuval Peres, Aravind Srinivasan, Benny Sudakov, Tibor Szábo, Greg Sorkin, John Tromp, David Wilson, Nick Wormald and Uri Zwick, who pointed out various inaccuracies and misprints, and suggested improvements in the presentation as well in the results. Needless to say, the responsibility for the remaining mistakes, as well as the responsibility for the (hopefully very few) new ones, is solely ours.

It is a pleasure to thank Oren Nechushtan, for his great technical help in the preparation of the final manuscript.

Contents

Part I

METHODS

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 in 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 non-constructive way, that it exists. This example appeared in a paper of P. Erdős from 1947. Although Szele 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 has been applying 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 unpractical. 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 non-constructive; it does not suply 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 hand 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, non-constructive 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 2-edge-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

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

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 derivitive 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 non-probabilistic, 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 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 2-*colorable* if there is a 2-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 2 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

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

and there is a 2-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 $m(n) \ge \Omega((\frac{n}{\ln n})^{\frac{1}{2}} 2^n)$.

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

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

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

$$Pr(S \text{ is monochromatic under } \chi) \ge p$$

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

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

There are 2^v colorings so

$$Pr(\vee_\chi A_\chi) \le 2^v (1-p)^m$$

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 2-colorable - and hence $m(n) \le m$.

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

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

then $2^v(1-p)^m < 2^v e^{-pm} \le 1$ so $m(n) \le m$. Now we need 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 $A_i \cap B_j \ne \emptyset$ 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 $A_i \cap B_j \ne \emptyset$. 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 $d_i \equiv xb_i (\bmod\, p)$, $0 \le d_i < p$. 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 $n/3$. Consequently, there is an x, $1 \le x < p$ and a subsequence A of B of cardinality $|A| > \frac{n}{3}$, such that $xa (\bmod\, p) \in C$ 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 $xa_1 + xa_2 \equiv xa_3 (\bmod\, p)$, 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.,

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

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

$$\Pr\left(|A_1 \cup A_2 \cup \ldots \cup A_t| \le n/2\right) \le \sum_{S \subset X, |S| \le n/2} \Pr(A_i \subset S, i = 1, \ldots, t)$$
$$\le 2^n (2^{n/2}/2^{((1/2)+\delta)n})^t = 2^{n(1-\delta t)} . \quad (1.2)$$

Define

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

Clearly

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

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 $(1 \le i \le t)$. By the convexity of z^t the expected value of Y satisfies

$$E(Y) = \sum_{B \in \mathcal{F}} \left(v(B)/m\right)^t = \frac{1}{m^t} \cdot m \left(\frac{\sum v(B)^t}{m}\right)$$
$$\ge \frac{1}{m^t} \cdot m \left(\frac{2d(\mathcal{F})}{m}\right)^t > 2m^{1-t\delta^2/2} . \quad (1.3)$$

Since $Y \le m$ we conclude that

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

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, $0 \le p \le 1$ 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

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

2. Suppose $n \ge 4$ 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 4 colors so that in every edge all 4 colors are represented.

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

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

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, $X = c_1X_1 + \ldots + c_nX_n$. 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 oftimes 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 probabilities 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

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

these probabilities 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

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

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 2 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

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

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 $i = j$

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

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

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

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

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

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

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

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:

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

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

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

■

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 $\epsilon_s \in \{0, 1\}$,

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

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

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

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 $n \ge 2$ 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 4 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
$$((xa_i)\ (mod\ p))\ mod\ n, \quad (1 \leq i \leq m)$$
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 $r_i = \sum_{1 \le j \le n} a_{ij}$ 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 $1 \le i \le n$. Then the permanent $per(A)$ 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]

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

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

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

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 1 $per(A) \leq G[L]$

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 $per(A) = rt$. Conditioning on $\sigma 1 = j$, $R_2 \cdots R_n$ is Lazyman's calculation of $per(A^2)$, 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

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

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

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

■

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

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

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

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

3

Alterations

> Beauty is the first test: there is no premanent 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 $P_iP_jP_k$ 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 halfintegers 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 2-coloring of V so that no edge is monochromatic. Beck (1978) improved Erdős's 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

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

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 one 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 2-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 had bounded $\Pr[C_e] \le k^2 pe^{pn}$. 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 $\sum_e \Pr[C_e] \le \sum_{e \ne f} \Pr[B_{ef}]$. As there are less than $(2^{n-1}k)^2$ pairs $e \ne f$ it now suffices to bound $\Pr[B_{ef}] \le 2^{1-2n}p$.

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

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

Lets take the factors one at a time. Firstly, v itself must start blue and turn red. Secondly, all other $v' \in f$ must start blue. Thirdly, all $v' \in f$ coming before v must have second coin tails. Fourthly, 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,...) 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 one. ■

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 $<$) 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 - $x_w < x$ 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

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

The integrand is always at most one so $\Pr[B_{ef}] \le 2^{1-2n}p$. 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 $|P| \le N/(k+1)$. 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 $P^{FINAL} = P_D$. 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 $|P^{FINAL}|$ 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

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

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 one. 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 (possibly several) set 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,Margaret at time 0.4. There are no other births. Leaves Nancy, Olive, Linda , Margaret, 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 $E[|P^{FINAL}|] \geq E[|P_c|]$. We make $f(c)$ arbitrarily small by taking c appropriately big, so that $E[|P^{FINAL}|] \geq (1 - o(1))n/(k+1)$. As $|P^{FINAL}| \leq n/(k+1)$ 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 proven 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 3-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

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

Let n be sufficiently large so that both these events have probability less than .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 $\alpha(G^*) \le \alpha(G)$. Thus

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

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

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 $Var[X]$. It is defined by

$$Var[X] = E[(X - E[X])^2]$$

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.

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

■

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 7,8 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 $Var[X]$ may be computed by the formula

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

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

$$Cov[Y, Z] = E[YZ] - E[Y]E[Z].$$

In general, if Y, Z are independent then $Cov[Y, Z] = 0$. This often simplifies considerably variance calculations. 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

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

and so

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

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 $\{1, \dots, n\}$. 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 $p \le M$. As no $x \le n$ can have more than ten prime factors larger than M we have $\nu(x) - 10 \le X(x) \le \nu(x)$ 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 $y - 1 < \lfloor y \rfloor \le y$

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

By linearity of expectation

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

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

Now we find an asymptotic expression for

$$Var[X] = \sum_{p \le M} Var[X_p] + \sum_{p \ne q} Cov[X_p, X_q].$$

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

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

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

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

Thus

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

Thus

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

and similarly

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

That is, the covariances do not affect the variance, $Var[X] = \ln\ln n + O(1)$ 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

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

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 $p_1, \ldots, p_s \le M$

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

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 $X_{p_1} \cdots X_{p_s}$. 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 $\tilde{X}$ approach 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 $Var[X]$.

Theorem 4.3.1

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

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* $Var[X] = o(E[X]^2)$ *then* $X > 0$ *a.a.*

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

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

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

Corollary 4.3.3 *If* $Var[X] = o(E[X]^2)$ *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 $i \sim j$

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

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

$$Var[X] \leq E[X] + \Delta.$$

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

Now suppose $p(n) >> n^{-2/3}$ 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

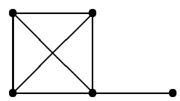
$$\Delta^* = O(n^2 p^5) + O(n p^3) = o(n^4 p^6) = o(E[X])$$

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

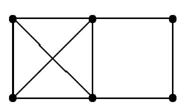
The proof of Theorem 4.4.1 appears to require a fortuitous calculation of Δ^*. The following definitions will allow for a description of when these calculations work out.

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

Now assume $p >> n^{-v/e}$ 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

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

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 $v_1/e_1 < \alpha < v/e$ and set $p = n^{-\alpha}$. The expected number of copies of H_1 is then $o(1)$ so almost always $G(n,p)$ contains no copy of H_1. But if it contains no copy of H_1 then it surely can contain no copy of H. ■

The threshold function for the property of containing a copy of H, for general H, was examined in the original papers of Erdős and Rényi. (Erdős and Rényi (1960) 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 $p >> n^{-v/e}$. 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 $p >> n^{-v/e}$ 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 one (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 one. 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

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

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 has 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 $\sigma^2 = Var[X]$. We bound

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

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

$$\Pr[|X - \mu| \ge \lambda n \sqrt{k}/2] \le \lambda^{-2}$$

Reversing,

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

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

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

and

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

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

Theorem 4.6.1

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

4.7 THE RÖDL NIBBLE

For $2 \le l < k < n$ 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 $M(n, k, l) \ge \binom{n}{l}/\binom{k}{l}$ 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 and 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 Rödl's result may be formulated in terms of either packing or covering here we deal only with the covering problem.)

Several researchers realized that Rödl's method applies in a much more general setting, dealing with covers in uniform hypergraphs. This has first been 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 Z. 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 $\delta_1, \delta_2, \cdots$ 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 $Var(|E'|) = |E|p(1-p) \leq (1 \pm \delta_1)\frac{\epsilon n}{r}$. 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

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

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

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

However,

$$Cov(I_x, I_y) = \mathbf{E}(I_x I_y) - \mathbf{E}(I_x)\mathbf{E}(I_y) = (1-p)^{d(x)+d(y)-d(x,y)} - (1-p)^{d(x)+d(y)}$$

$$\leq (1-p)^{-d(x,y)} - 1 \leq (1 - \frac{\epsilon}{D})^{-\delta D} \leq \delta_5.$$

It follows that

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

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

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

It remains to bound the sum $\sum_{x\in e, x \in f, x, f\, good} Cov(I_e, I_f)$. For each fixed good e this sum is a sum of the form $\sum_{x \in f, f\, good} Cov(I_e, I_f)$. 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, $Cov(I_e, I_f) \leq (1-p)^{-t(e,f)} - 1 \leq \delta_{13}$, implying that for each fixed good edge e

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

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

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

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 $V = V_0, V_1, \cdots, V_t$, 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

$$(1 + 4\delta)\sum_{i=0}^{t-1}\frac{\epsilon n}{r}e^{-i\epsilon} + |V_t| \leq (1 + 4\delta)\frac{\epsilon n}{r}\frac{1}{1 - e^{-\epsilon}} + (1 + 2\delta)ne^{-\epsilon t}$$

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

This completes the proof. ■

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

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

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

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 $(1 + o(1))\frac{\binom{n}{l}}{\binom{k}{l}}$, 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 $Var(X)$ denote its variance. Prove that
$$Prob(X = 0) \leq \frac{Var(X)}{E(X^2)}.$$

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

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

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

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

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

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

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 $\{ax \ (mod\ p) : x \in X\}$ 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

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

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 Brégman's proof of Minc Conjecture for the permanent of a $(0,1)$-matrix, described in the Probabilistic Lens; Brégman's 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

$$C(T) \leq F(T). \tag{4.4}$$

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,

$$\sum_{i}^{n} r_i = \binom{n}{2}. \tag{4.5}$$

By interpreting combinatorially the terms in the expansion of the permanent $Per(A_T)$, it follows that

$$Per(A_T) = F(T). \tag{4.6}$$

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.,

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

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 (4.5) and (4.6), $F(T)$ is at most the value of the function $\prod_{i=1}^{n} (r_i!)^{1/r_i}$, where the integral variables r_i satisfy (4.5) 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 Hamilton 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 Hamilton path in S, the probability that it can be extended to a Hamilton cycle in T is precisely $1/4$. Thus, the expected number of Hamilton 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 $Pr\left(\bigwedge_{i=1}^{n} \overline{A}_j\right) \geq \prod_{i=1}^{n}(1 - x_i)$. 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 $i \notin S$

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

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

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

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

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

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

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

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

The assertion of Lemma 5.1.1 now follows easily, as

$$\begin{aligned} Pr\left(\bigwedge_{i=1}^{n}\overline{A}_i\right) &= (1-Pr(A_1))\cdot(1-Pr(A_2|\overline{A}_1))\cdot\ldots\cdot(1-Pr(A_n|\bigwedge_{i=1}^{n-1}\overline{A}_i) \\ &\ge \prod_{i=1}^{n}(1-x_i)\,, \end{aligned} \tag{5.5}$$

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 $Pr(A_i) \le p$ for all $1 \le i \le n$. If*

$$ep(d+1) \le 1 \tag{5.6}$$

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, $\left|\{j : (i,j) \in E\}\right| \le d$. 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 $d \ge 1$, $\left(1 - \frac{1}{d+1}\right)^d > 1/e$. ■

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 $Pr(A_i) \le x_i \prod_{(ij)\in E}(1-x_j)$ by the weaker assumption that for each i and each $S_2 \subset \{1,\ldots,n\}\backslash\{j : (i,j) \in E\}$, $\quad Pr\left(A_i \middle| \bigwedge_{j\in S_2}\overline{A}_j\right) \le x_i \prod_{(i,j)\in E}(1-x_j)$.
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 does 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 2-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 $e(d+1) \le 2^{k-1}$ 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*

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

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 2-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 2-color the edges of K_n randomly and independently, where each edge is colored blue with probability p. For each set of 3 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 $\binom{k}{2}(n-k) < k^2n/2$ 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 3-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 non-decomposable 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 3 dimensional Euclidean space by open unit balls. Suppose, further, than 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 2-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 2-colorable. Given a 2-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* $la(G)$ 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

$la(G) \le \lceil 3\Delta/5 \rceil$ for even Δ and that $la(G) \le \lceil (3\Delta+2)/5 \rceil$ 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 $\Delta \ge \Delta_0$, 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 *di-linear 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 $V = V_1 \cup V_2 \cup \cdots \cup V_r$ be a partition of V into r pairwise disjoint sets. Suppose each set V_i is of cardinality $|V_i| \ge 2ed$, 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, $Pr(A_f) \le 1/g^2$. 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 $N^+(v, i) = \big|\{u \in V; (v, u) \in E$*

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 $10\sqrt{d} \le p \le 20\sqrt{d}$ imply

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

We have thus proved:

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

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

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*

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

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 non-occurance 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 $a_{i\pi(i)}$ $(1 \le i \le n)$ are all distinct.

Theorem 5.6.1 *Suppose $k \le (n-1)/(4e)$ 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 fourtuples (i, j, i', j') satisfying $i < i', j \ne j'$ 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 fourtuples 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 fourtuple (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

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

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 $Pr(A_{1122} \bigwedge \bigwedge_S \overline{A}_{pqp'q'}) \leq Pr(A_{1i2j} \bigwedge \bigwedge_S \overline{A}_{pqp'q'})$ for all $i \neq j$ and hence that $Pr(A_{1122} | \bigwedge \bigwedge_S \overline{A}_{pqp'q'}) \leq 1/n(n-1)$. 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 a 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 2-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? J. 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*

$$16D(1+d) < 2^{n_1}$$

$$16D(1+d) < 2^{n_2}$$

$$2e(1+d) < 2^{n_3}.$$

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 non-fixed 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 survived sets.
Claim 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 while still uncolored.) 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

$$(v, u) \in E \quad \text{and} \quad f(u) \equiv (f(v) + 1)\,(\mathrm{mod}\, k). \tag{5.10}$$

Starting at an arbitrary $v = v_0 \in V$ and applying (5.10) 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. ■

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 $I\!R^+$ denote the set of nonnegative real numbers. For a function $\varphi : P(N) \to I\!R^+$ 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 $\alpha, \beta, \gamma, \delta : P(N) \to I\!R^+$ 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 $\alpha', \beta', \gamma', \delta' : P(N') \to I\!R^+$. 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 proven $\alpha'(A')\beta'(B') = \overline{\alpha}(P(T))\overline{\beta}(P(T)) \leq \overline{\gamma}(P(T))\overline{\delta}(P(T)) = \gamma'(A' \cup B')\delta'(A' \cap B')$, 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 $y \le x$. 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*

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

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

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

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*

$$|X| \cdot |Y| \le |X \vee Y| \cdot |X \wedge Y|\,.$$

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 $|X \backslash X| \ge |X|$.

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:

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

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

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

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

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 $x \in L$

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

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 $f, g : L \to I\!R^+$ are both increasing or both decreasing then

$$< fg > \geq < f > \cdot < g > .$$

Similarly, if f is increasing and g is decreasing (or vice versa), then

$$< fg > \leq < f > < g > .$$

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 $f : P(N) \to I\!R^+$ 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 $\mu = \mu_p : P(N) \to I\!R^+$ 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 $p = (p_1, , \ldots, , p_n)$, $0 \leq p_i \leq 1$*

$$\begin{gathered} Pr_p(\mathcal{A} \cap \mathcal{B}) \geq Pr_p(\mathcal{A}) \cdot Pr_p(\mathcal{B}) \\ Pr_p(\mathcal{C} \cap \mathcal{D}) \geq Pr_p(\mathcal{C}) \cdot Pr_p(\mathcal{D}) \\ Pr_p(\mathcal{A} \cap \mathcal{C}) \leq Pr_p(\mathcal{A}) \cdot Pr_p(\mathcal{C}) . \end{gathered}$$

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 $i, j \in V, \quad i \neq j$, 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 $i, j \in V, \quad i \neq j$ 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:

$$\begin{aligned} (\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z}))_i &= \min(x_i - x_1, (\mathbf{y} \vee \mathbf{z})_i - (\mathbf{y} \vee \mathbf{z})_1) + \max(x_1, (\mathbf{y} \vee \mathbf{z})_1) = \\ &= \min(x_i - x_1,\ \max(y_i - y_1, z_i - z_1)) + \max(x_1, \min(y_1, z_1)) . \end{aligned}$$

Similarly, the i-th component of $(\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z})$ is:

$$\begin{gathered} ((\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z}))_i = \\ \max((\mathbf{x} \wedge \mathbf{y})_i - (\mathbf{x} \wedge \mathbf{y})_1,\ (\mathbf{x} \wedge \mathbf{z})_i - (\mathbf{x} \wedge \mathbf{z})_1) + \min((\mathbf{x} \wedge \mathbf{y})_1,\ (\mathbf{x} \wedge \mathbf{z})_1) = \\ \max(\min(x_i - x_1,\ y_i - y_1),\ \min(x_i - x_1,\ z_i - z_1)) + \\ \min(\max(x_1, y_1), \max(x_1, z_1)) . \end{gathered}$$

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

$$\begin{aligned}(x \vee y)_i \qquad &= \max(x_i - x_1,\ y_i - y_1) + \min(x_1, y_1) \\ &\leq \max(x_j - x_1,\ y_j - y_1) + \min(x_1, y_1) = (x \vee y)_j\ ,\end{aligned}$$

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 2-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
$$|\cup_{i=1}^k \mathcal{F}_i| \leq 2^n - 2^{n-k}.$$

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 $0 \le i < m$,

$$E[X_{i+1}|X_i, X_{i-1}, \ldots, X_0] = X_i.$$

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.

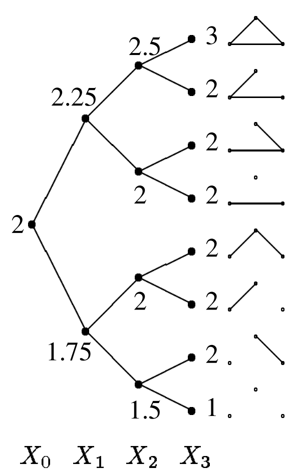
A simple but instructive martingale occurs when the gambler plays "flip a coin" for stakes of one dollar each time. Let $Y_1, \ldots, Y_m$ be independent coin flips, each $+1$ or -1 with probability one half. Normalize so that $X_0 = 0$ is the gambler's initial stake, though he has unlimited credit. 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

$$X_i(H) = E[f(G) | \text{for } x,y \le i, \{x,y\} \in G \longleftrightarrow \{x,y\} \in H].$$

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 *Let* $0 = X_0, \ldots, X_m$ *be a martingale with*

$$|X_{i+1} - X_i| \le 1$$

for all $0 \le i < m$. *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 $|Y_i| \le 1$ and $E[Y_i | X_{i-1}, X_{i-2}, \ldots, X_0] = 0$. Then, as in A.1.16,

$$E[e^{\alpha Y_i} | X_{i-1}, X_{i-2}, \ldots, X_0] \le \cosh(\alpha) \le e^{\alpha^2/2}.$$

Hence

$$E[e^{\alpha X_m}] = E[\prod_{i=1}^{m} e^{\alpha Y_i}]$$

$$= E\big[\big(\prod_{i=1}^{m-1} e^{\alpha Y_i}\big) E(e^{\alpha Y_m} | X_{m-1}, X_{m-2}, \ldots, X_0)\big]$$

$$\leq E\big[\prod_{i=1}^{m-1} e^{\alpha Y_i}\big] e^{\alpha^2/2} \leq e^{\alpha^2 m/2}.$$

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

$$\begin{aligned} \Pr[\omega(G) < k] &= \Pr[Y = 0] & &\leq \Pr[Y - E[Y] \leq -E[Y]] \\ &\leq e^{-E[Y]^2/2\binom{n}{2}} & &\leq e^{-(c'+o(1))n^2/k^8} \\ &= e^{-(c+o(1))n^2/\ln^8 n} \end{aligned}$$

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 had 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 3-colored.*

Proof. If not, let T be a minimal set which is not 3-colorable. As $T - \{x\}$ is 3-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

$$\binom{n}{t} \leq (\frac{ne}{t})^t \text{ and } \binom{\binom{t}{2}}{\frac{3t}{2}} \leq (\frac{te}{3})^{3t/2},$$

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 3 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 $0 \le i < m$

$$h, h' \text{ differ only on } B_{i+1} - B_i \Rightarrow |L(h') - L(h)| \le 1.$$

Theorem 7.4.1 *Let L satisfy the Lipschitz condition. Then the corresponding martingale satisfies*

$$|X_{i+1}(h) - X_i(h)| \le 1$$

for all $0 \le i < m$, $h \in A^B$.

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

$$\begin{aligned} |X_{i+1}(h) - X_i(h)| &= \left| \textstyle\sum_{h' \in H} w_{h'} [L(h') - \sum_{h^* \in H[h']} L(h^*) q_{h^*}] \right| \\ &\le \textstyle\sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} |q_{h^*} [L(h') - L(h^*)]| \,. \end{aligned}$$

The Lipschitz condition gives $|L(h') - L(h^*)| \le 1$ so

$$|X_{i+1}(h) - X_i(h)| \le \sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} q_{h^*} = \sum_{h' \in H} w_{h'} = 1.$$

■

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 $\lambda > 0$*

$$\Pr[L(g) \ge \mu + \lambda\sqrt{m}] < e^{-\lambda^2/2},$$

$$\Pr[L(g) \le \mu - \lambda\sqrt{m}] < e^{-\lambda^2/2}.$$

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*

$$\Pr[|Y - E[Y]| > \alpha\sigma] \le 2e^{-\frac{\alpha^2}{2(1+\epsilon)}} \tag{7.1}$$

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 $\sigma = \Theta(\sqrt{n^2p})$. If $\alpha \to \infty$ with $\alpha = o(\sqrt{n^2p})$ 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

$$E[e^{\lambda Y}] \le e^{(1+\epsilon)\lambda^2\sigma^2/2}. \tag{7.2}$$

The Martingale Inequality then follows by the Markov bound

$$\Pr[Y > \alpha\sigma] < e^{-\lambda\alpha\sigma}E[e^{\lambda Y}] \le e^{-\alpha^2/2(1+\epsilon)}.$$

We first claim that for all $\epsilon > 0$ there exists $\delta > 0$ so that for $0 \le p \le 1$ and $|a| \le \delta$

$$pe^{(1-p)a} + (1-p)e^{-pa} \le e^{(1+\epsilon)p(1-p)a^2/2}. \tag{7.3}$$

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 $j \ge 3$ 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

$$\mu_y = (1-p)b \text{ and } \mu_n = -pb$$

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 one. 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 two 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.

$$X_i(\epsilon) = \frac{1}{2}\left[X_{i+1}(\epsilon) + X_{i+1}(\epsilon')\right]$$

so that

$$|X_i(\epsilon) - X_{i+1}(\epsilon)| = \frac{1}{2}\left|X_{i+1}(\epsilon') - X_{i+1}(\epsilon)\right| \leq 1$$

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 1 $\rho(A, \vec{x})$ *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

$$\int_{\Omega} \exp[\frac{1}{4}\rho^2(A, \vec{x})] d\vec{x} \leq \frac{1}{\Pr[A]}.$$

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

$$\int \exp[\frac{1}{4}\rho^2(A, \vec{x})] = \Pr[A] + (1 - \Pr[A])e^{1/4} \leq \frac{1}{\Pr[A]},$$

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

$$B = \{x \in OLD : (x, \omega) \in A \text{ for some } \omega \in NEW\}$$

and for any $\omega \in NEW$ set

$$A_\omega = \{x \in OLD : (x, \omega) \in A\}.$$

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

$$\int_x \exp[\frac{1}{4}\rho^2(A, (x, \omega))] \leq e^{(1-\lambda)^2/4} \int_x (\exp[\frac{1}{4}\rho^2(A_\omega, x)])^\lambda (\exp[\frac{1}{4}\rho^2(B, x)])^{1-\lambda}.$$

By Hölder's Inequality this is at most

$$e^{(1-\lambda)^2/4} \left[\int_x \exp[\frac{1}{4}\rho^2(A_\omega, x)]\right]^\lambda \left[\int_x \exp[\frac{1}{4}\rho^2(B, x)]\right]^{1-\lambda}$$

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

$$\int_x \exp[\frac{1}{4}\rho^2(A,(x,\omega))] \leq \frac{1}{\Pr[B]}\left(2 - \frac{\Pr[A_\omega]}{\Pr[B]}\right).$$

We integrate over ω giving

$$\int_\omega\int_x \exp[\frac{1}{4}\rho^2(A,(x,\omega))] \leq \frac{1}{\Pr[B]}\left(2 - \frac{\Pr[A]}{\Pr[B]}\right) = \frac{1}{\Pr[A]}x(2-x)$$

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 2 *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.

Lets 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

$$\Pr[Y \leq m - tm^{1/2} \binom{k}{2}^{1/2}] \Pr[Y \geq m] < e^{-t^2/4}.$$

Set $t = \Theta(m^{1/2}/k)$ so that $m = tm^{1/2} \binom{k}{2}^{1/2}$. Then

$$\Pr[\omega(G) < k] = \Pr[Y \leq 0] < 2e^{-t^2/4} < \exp\left[-\Omega(\frac{n^2}{\ln^6 n})\right]$$

which improves the bound of Theorem 7.3.2. Still, we should note that application of the Extended Janson Inequality in §10.3 does even better.

7.8 KIM-VU POLYNOMIAL CONCENTRATION

A recent approach of Kim and Vu (to appear) 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

$$Prob(\chi(H) \leq 400) < 1/100.$$

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

$$P_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f(\frac{i}{n}).$$

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 $Var[X] \leq \mu + \Delta$ so that

$$\Pr[\wedge_{i \in I} \overline{B_i}] = \Pr[X = 0] \leq \frac{Var[X]}{E[X]^2} \leq \frac{\mu + \Delta}{\mu^2}.$$

Suppose $\mu \to \infty$, $\mu << \Delta$, 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

$$\Pr[B_i | \wedge_{j \in J} \overline{B_j}] \leq \Pr[B_i]$$

valid for all index sets $J \subset I, i \notin J$ and

$$\Pr[B_i | B_k \wedge \bigwedge_{j \in J} \overline{B_j}] \leq \Pr[B_i | B_k]$$

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

$$\begin{aligned}
&E\left[-\ln[\Pr[\wedge_{i\in S}\overline{B_i}]]\right] \\
&\geq E\left[\textstyle\sum_{i\in S}\Pr[B_i]\right] - \frac{1}{2}E\left[\textstyle\sum_{i,j\in S, i\sim j}\Pr[B_i \wedge B_j]\right] \\
&= p\mu - p^2\frac{\Delta}{2}.
\end{aligned}$$

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 one. 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 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

$$C_{x_1} \wedge \ldots \wedge C_{x_r} = \wedge\overline{B_{x_iyz}},$$

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

$$\sum \Pr[B_{x_iyz}] = p^3\left(r\binom{n-1}{2} - O(n)\right) = r\mu + O(n^{-1+o(1)}).$$

As before Δ is p^5 times the number of pairs $x_i yz \sim x_j y'z'$. 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

$$\begin{aligned}
\Pr[\text{there exists a disfam } J,\ |J| = s] &\leq \sum^{*} \Pr[\wedge_{j \in J} B_j] \\
= \sum^{*} \prod_{j \in J} \Pr[B_j] &= \tfrac{1}{s!} \textstyle\sum^o \Pr[B_{j_1}] \cdots \Pr[B_{j_s}] \\
\leq \frac{1}{s!} \sum^{a} \Pr[B_{j_1}] \cdots \Pr[B_{j_s}] &\leq \tfrac{1}{s!} [\textstyle\sum_{i \in I} \Pr[B_i]]^s = \mu^s/s!.
\end{aligned}$$

■

Lemma 8.4.1 gives an effective upper bound when $\mu^s << s!$ - 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

$$\begin{aligned} \Pr[\textit{there exists a maxdisfam } J,\ |J| = s] &\leq \frac{\mu_s^s}{s!} e^{-\mu_s} e^{\frac{\Delta}{2}} \\ &\leq \frac{\mu^s}{s!} e^{-\mu} e^{s\nu} e^{\frac{\Delta}{2}}. \end{aligned}$$

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 Inequality to 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 Section 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* $\mu >> \ln n$. *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* $\mu >> \ln n$. *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

$$\Pr[\text{there exists a maxdisfam } J, |J| \leq \mu(1-\epsilon)] \leq (1+o(1))\Pr[P \leq \mu(1-\epsilon)]$$

$$\Pr[\text{there exists a maxdisfam } J, \mu(1+\epsilon) \leq |J| \leq 3\mu]$$
$$\leq (1+o(1)) \Pr[\mu(1+\epsilon) \leq P \leq 3\mu]$$

$$\Pr[\text{there exists a maxdisfam } J, |J| \geq 3\mu]$$
$$\leq \Pr[\text{there exists a disfam } J, |J| \geq 3\mu] \quad \leq \sum_{s=3\mu}^{\infty} \frac{\mu^s}{s!} = O((1-c)^\mu)$$

where $c > 0$ is an absolute constant. Since $\mu >> \ln n$ 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 $n^6p^{11} = o(n^{-1})$. 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 9 and there is no set of 4 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

$$\Pr[\wedge_{i=1}^{\infty} \vee_{j=i}^{\infty} A_j] = 0.$$

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 $\epsilon = .9, \delta = .1$ and

$$\Pr[|f(n) - \mu| > .9\mu] < 2e^{-5\pi \ln n} < n^{-1.1}$$

for n sufficiently large. Take $c_1 < .1(50\pi)$ 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 Section 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, Section 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

$$\Pr[B_i | B_k \wedge \bigwedge_{j \in J} \overline{B_j}] \leq \Pr[B_i | B_k]$$

valid for all index sets $J \subset I, i, k \notin J$. 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

$$\sum_{i \sim j} y(i,j) \le 2[\Delta + \sum_{i \sim j} \Pr[B_i]\Pr[B_j]].$$

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*

$$\Pr[X \le (1-\gamma)\mu] < e^{-\gamma^2 \mu / (2 + \frac{\Delta}{\mu})}.$$

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 $k \le (2-\epsilon)\log_2 n$ 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)

$$c > 2k^2 H(1/k) \ln 2$$

$$\epsilon < e^{-5} 3^3 c^{-3},$$

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 4 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 $t \le \epsilon n$ vertices have at least $3t/2$ edges is less than

$$\sum_{t \le \epsilon n} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} \left(\frac{c}{n}\right)^{3t/2}.$$

We outline the analysis. When $t = O(1)$ the terms are negligible. Otherwise we bound each term from above by

$$\left[\frac{ne}{t}\left(\frac{te}{3}\right)^{3/2}\left(\frac{c}{n}\right)^{3/2}\right]^t \le \left[e^{5/2}3^{-3/2}c^{3/2}\sqrt{t/n}\right]^t.$$

Now since $t \le \epsilon n$ 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.

9

Pseudo-Randomness

> '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 Caroll

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 explicitly such graphs? 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 pseudo-random 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 $p \equiv 3 (mod\ 4)$ 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 $p \equiv 3(mod\ 4)$, -1 is a quadratic non-residue 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. $\sum_i d_{ij}d_{il} = \sum_i \chi(i-j)\chi(i-l) = \sum_{i\neq j,l} \chi(i-j)\chi(i-l)$
$= \sum_{i\neq j,l} \chi((i-j)/(i-l)) = \sum_{i\neq j,l} \chi(1+(l-j)/(i-l))$.

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:

$$(\textstyle\sum_{i\in A}\sum_{j\in B} d_{ij})^2 \leq |A|(\sum_{i\in A}(\sum_{j\in B} d_{ij})^2)$$

$$\leq |A|(\textstyle\sum_{i\in GF(p)}(\sum_{j\in B} d_{ij})^2)$$

$$= |A|(\textstyle\sum_{i\in GF(p)}(|B| + 2\sum_{j<l\in B} d_{ij}d_{il}))$$

$$= |A||B|p + 2|A|\textstyle\sum_{j<l\in B}\sum_{i\in GF(P)} d_{ij}d_{il}$$

$$= |A||B|p - |A||B|(|B|-1) = |A||B|(p-|B|+1) \leq |A||B|p$$

,

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 anti-symmetric 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 pseudo-random 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:

$$\begin{aligned}((D - A)x, x) &= \textstyle\sum_{u \in V} (d(x(u))^2 - \sum_{v; uv \in E} x(v)x(u)) \\ &= d \textstyle\sum_{u \in V} (x(u))^2 - 2 \sum_{uv \in E} x(v)x(u) = \sum_{uv \in E} (x(v) - x(u))^2.\end{aligned}$$

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 non-bipartite 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 neighbours of v in G, and let $N_B(v) = N(v) \cap B$ denote the set of all neighbours 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 neighbours 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 by Ajtai, 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

$$c_1 = \textstyle\sum_{v\in V} f(v)/\sqrt{n} = \sum_{v\in V-C} f(v)/\sqrt{n}$$

$$\leq (\textstyle\sum_{v\in V-C} f(v)^2)^{1/2}((1-c)n/n)^{1/2} = \sqrt{1-c}\,,$$

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 $\gamma_1 \geq \gamma_2 \cdots \geq \gamma_m$ 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 neighbours 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 non-primality of q) if for the sequence $x_0, \ldots, x_a$ defined by $x_0 = x^b (mod\ q)$ and $x_i = x_{i-1}^2 (mod\ q)$ 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 non-prime 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 non-prime 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 less 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 non-prime, 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 $i(mod\ (q-1))$. 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 pseudo-random 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 neighbours 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 pseudo random 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 behaviour 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: $\sum_{u,v \in V} \big|\, |N(u) \cap N(v)| - \frac{n}{4} \big| = o(n^3)$.

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

$$P_1(4) \Longrightarrow P_2 \Longrightarrow P_3 \Longrightarrow P_4 \Longrightarrow P_5$$
$$\Longrightarrow P_6 \Longrightarrow P_1(s) \ for\ all\ s \geq 1\ \ (\Longrightarrow P_1(4)).$$

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

$$\sum_{i=1}^n \lambda_i^4 = d^4 + \sum_{i=2}^n \lambda_i^4$$

$$= (1+o(1))(n/2)^4 + \sum_{i=2}^n \lambda_i^4 = N_G(C(4)) + O(n^3)$$

$$= (1+o(1))(n/2)^4.$$

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

$$e(B,C) = e(B \cup C) - e(B) - e(C)$$

$$= \frac{1}{4}(|B|+|C|)^2 - \frac{1}{4}|B|^2 - \frac{1}{4}|C|^2 + o(n^2) = \frac{1}{2}|B||C| + o(n^2),$$

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:

$$e(B,C) = \frac{1}{2}(b-x)(c-x) + \frac{1}{2}x(c-x) + \frac{1}{2}x(b-x) + \frac{2}{4}x^2 + o(n^2)$$

$$= \frac{1}{2}bc + o(n^2) = \frac{1}{2}|B||C| + o(n^2),$$

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

$$\sum_{u \in V, u \neq v} | \, |N(u) \cap N(v)| - \frac{n}{4}|.$$

Define

$$B_1 = \{u \in V, u \neq v : \ |N(u) \cap N(v)| \geq \frac{n}{4}\},$$

and similarly

$$B_2 = \{u \in V, u \neq v : \ |N(u) \cap N(v)| < \frac{n}{4}\}.$$

Let C be the set of all neighbours of v in G. Observe that

$$\sum_{u \in B_1} | \, |N(u) \cap N(v)| - \frac{n}{4}|$$

$$= \sum_{u \in B_1} |N(u) \cap N(v)| - |B_1|\frac{n}{4}$$

$$= e(B_1, C) - |B_1|\frac{n}{4}.$$

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

$$\sum_{u \in B_2} | \, |N(u) \cap N(v)| - \frac{n}{4}| = o(n^2).$$

It follows that for every vertex v of G;

$$\sum_{u \in V, u \neq v} | \, |N(u) \cap N(v)| - \frac{n}{4}| = o(n^2),$$

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

$$\sum_{u,v \in V} | \, s(u, v) - \frac{n}{2}| = o(n^3). \tag{9.2}$$

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

$$f_r(\alpha, \epsilon) = |\{v \in V : v \neq \alpha_1, \ldots \alpha_r \textit{ and } a(v, \alpha_j) = \epsilon_j \textit{ for all } 1 \leq j \leq r\}|.$$

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

$$a(u, \alpha_k) = a(v, \alpha_k) = \epsilon_k \textit{ for all } 1 \leq k \leq r.$$

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;

$$\sum_{u \neq v} s(u,v)_{(r)} = \sum_{u \neq v} (\frac{n}{2} + \epsilon_{uv})_{(r)}$$

$$= \sum_{k=0}^{r} \sum_{u \neq v} c_k (\frac{n}{2})^k \epsilon_{uv}^{r-k} \quad (for\ appropriate\ constants\ c_k)$$

$$= (\frac{n}{2})^r n_{(2)} + \sum_{k=0}^{r-1} \sum_{u \neq v} c_k (\frac{n}{2})^k \epsilon_{uv}^{r-k}$$

$$\leq (\frac{n}{2})^r n_{(2)} + \sum_{k=0}^{r-1} \sum_{u \neq v} |c_k| n^k |\epsilon_{uv}|^{r-k}$$

$$\leq n^{r+2}2^{-r} + c \sum_{k=0}^{r-1} n^k \sum_{u \neq v} |\epsilon_{uv}|^{r-k} \qquad (for\ an\ appropriate\ constant\ c)$$

$$\leq n^{r+2}2^{-r} + c \sum_{k=0}^{r-1} n^k \cdot o(n^{r-k+2})$$

$$= n^{r+2}2^{-r}(1+o(1)),$$

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:

$$|N_{r+1}-N_r E(f_r)|^2=|\sum_{N_r\ terms}(f_r(\alpha,\epsilon)-E(f_r))|^2.$$

By Cauchy-Schwarz, the last expression is at most

$$N_r\sum_{N_r\ terms}(f_r(\alpha,\epsilon)-E(f_r))^2$$
$$\leq N_r\sum_{\alpha,\epsilon}(f_r(\alpha,\epsilon)-E(f_r))^2$$
$$=N_r\cdot o(n^{r+2})=o(n^{2r+2}).$$

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 Section 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 non-adjacent 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 pseudo-random 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 3-regular graph on $2n$ vertices obtained by picking 3 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 2 *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:

$$P^{2k}(u,v) = \sum_{w \in V} P^k(u,w)P^k(w,v) = \sum_{w \in V} P^k(u,w)P^k(v,w), \tag{9.6}$$

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 (9.6) is when $u = v$, completing the proof of the theorem. ■

Part II

TOPICS

10

Random Graphs

It is six in the morning. The house is asleep. Nice music is playing. I prove and conjecture.
– Paul Erdős, in a letter to Vera Sós

Let n be a positive integer, $0 \leq p \leq 1$. The random graph $G(n, p)$ is a probability space over the set of graphs on the vertex set $\{1, \ldots, n\}$ determined by

$$\Pr[\{i, j\} \in G] = p$$

with these events mutually independent. This model is often used in the probabilistic method for proving the existence of certain graphs. In this chapter we study the properties of $G(n, p)$ for their own sake.

Random Graphs is an active area of research which combines probability theory and graph theory. The subject began in 1960 with the monumental paper *On the Evolution of Random Graphs* by Paul Erdős and Alfred Rényi. The book *Random Graphs* by B. Bollobás (1985) is the standard source for the field. In this chapter we explore only a few of the many topics in this fascinating area.

There is a compelling dynamic model for random graphs. For all pairs i, j let $x_{i,j}$ be selected uniformly from $[0, 1]$, the choices mutually independent. Imagine p going from 0 to 1. Originally, all potential edges are "off". The edge from i to j (which we may imagine as a neon light) is turned on when p reaches $x_{i,j}$ and then stays on. At $p = 1$ all edges are "on". At time p the graph of all "on" edges has distribution $G(n, p)$. As p increases $G(n, p)$ *evolves* from empty to full.

In their original paper, Erdős and Rényi let $G(n, e)$ be the random graph with n vertices and precisely e edges. Again there is a dynamic model: Begin with no edges and add edges randomly one by one until the graph becomes full. Generally $G(n, e)$

will have very similar properties as $G(n,p)$ with $p \sim \frac{e}{\binom{n}{2}}$. We will work on the probability model exclusively.

10.1 SUBGRAPHS

The term "the random graph" is, strictly speaking, a misnomer. $G(n,p)$ is a probability space over graphs. Given any graph theoretic property A there will be a probability that $G(n,p)$ satisfies A, which we write $\Pr[G(n,p) \models A]$. When A is monotone $\Pr[G(n,p) \models A]$ is a monotone function of p. As an instructive example, let A be the event "G is triangle free". Let X be the number of triangles contained in $G(n,p)$. Linearity of expectation gives

$$E[X] = \binom{n}{3}p^3.$$

This suggests the parametrization $p = c/n$. Then

$$\lim_{n\to\infty} E[X] = \lim_{n\to\infty} \binom{n}{3}p^3 = c^3/6.$$

It turns out that the distribution of X is asymptotically Poisson. In particular

$$\lim_{n\to\infty} \Pr[G(n,p) \models A] = \lim_{n\to\infty} \Pr[X=0] = e^{-c^3/6}.$$

Note that

$$\lim_{c\to 0} e^{-c^3/6} = 1$$

$$\lim_{c\to\infty} e^{-c^3/6} = 0$$

When $p = 10^{-6}/n$, $G(n,p)$ is very unlikely to have triangles and when $p = 10^6/n$, $G(n,p)$ is very likely to have triangles. In the dynamic view the first triangles almost always appear at $p = \Theta(1/n)$. If we take a function such as $p(n) = n^{-.9}$ with $p(n) >> n^{-1}$ then $G(n,p)$ will almost always have triangles. Occasionally we will abuse notation and say, for example, that $G(n, n^{-.9})$ contains a triangle - this meaning that the probability that it contains a triangle approaches 1 as n approaches infinity. Similarly, when $p(n) << n^{-1}$, for example, $p(n) = 1/(n \ln n)$, then $G(n,p)$ will almost always not contain a triangle and we abuse notation and say that $G(n, 1/(n \ln n))$ is triangle free. It was a central observation of Erdős and Rényi that many natural graph theoretic properties become true in a very narrow range of p. They made the following key definition.

Definition. $r(n)$ is called a threshold function for a graph theoretic property A if
(i) When $p(n) << r(n)$, $\lim_{n\to\infty} \Pr[G(n,p) \models A] = 0$
(ii) When $p(n) >> r(n)$, $\lim_{n\to\infty} \Pr[G(n,p) \models A] = 1$
or visa versa.

In our example, $1/n$ is a threshold function for A. Note that the threshold function, when one exists, is not unique. We could equally have said that $10/n$ is a threshold function for A.

Let's approach the problem of $G(n, c/n)$ being triangle free once more. For every set S of three vertices let B_S be the event that S is a triangle. Then $\Pr[B_S] = p^3$. Then "triangle freeness" is precisely the conjunction $\wedge \overline{B}_S$ over all S. If the B_S were mutually independent then we *would* have

$$\Pr[\wedge \overline{B}_S] = \prod [\overline{B}_S] = (1-p^3)^{\binom{n}{3}} \sim e^{-\binom{n}{3}p^3} \to e^{-c^3/6}.$$

The reality is that the B_S are not mutually independent though when $|S \cap T| \leq 1$, B_S and B_T are mutually independent.

We apply Janson's Inequality, Theorem 8.1.1. In the notation of §8.1 $I = \{S \subset V(G) : |S| = 3\}$ and $S \sim T$ if and only if $|S \cap T| = 2$. Here $\epsilon = p^3 = o(1)$, $\mu = \binom{n}{3}p^3 \sim c^3/6$, and $M = e^{-\mu(1+o(1))} = e^{-c^3/6+o(1)}$. There are $6\binom{n}{4} = O(n^4)$ pairs S, T of triples with $S \sim T$. For each $\Pr[B_S \wedge B_T] = p^5$. Thus

$$\Delta = O(n^4)p^5 = n^{-1+o(1)} = o(1).$$

When $\Delta = o(1)$ Janson's Inequality sandwiches an asymptotic bound:

$$\lim_{n\to\infty} \Pr[\wedge \overline{B}_S] = \lim_{n\to\infty} M = e^{-c^3/6}.$$

Can we duplicate this success with the property A that G contains no (not necessarily induced) copy of a general given graph H? We use the definitions of balanced and strictly balanced of §4.4.

Theorem 10.1.1 *Let H be a* strictly *balanced graph with v vertices, e edges and a automorphisms. Let $c > 0$ be arbitrary. Let A be the property that G contains no copy of H. Then with $p = cn^{-v/e}$,*

$$\lim_{n\to\infty} \Pr[G(n,p) \models A] = exp[-c^e/a].$$

Proof. Let A_α, $1 \leq \alpha \leq \binom{n}{v} v!/a$, range over the edge sets of possible copies of H and let B_α be the event $G(n,p) \supseteq A_\alpha$. We apply Janson's Inequality. As

$$\lim_{n\to\infty} \mu = \lim_{n\to\infty} \binom{n}{v} v! p^e/a = c^e/a,$$

we find

$$\lim_{n\to\infty} M = exp[-c^e/a].$$

Now we examine (as in Theorem 4.4.2)

$$\Delta = \sum_{\alpha \sim \beta} \Pr[B_\alpha \wedge B_\beta].$$

We split the sum according to the number of *vertices* in the intersection of copies α and β. Suppose they intersect in j vertices. If $j = 0$ or $j = 1$ then $A_\alpha \cap A_\beta = \emptyset$ so that $\alpha \sim \beta$ cannot occur. For $2 \leq j \leq v$ let f_j be the maximal $|A_\alpha \cap A_\beta|$ where $\alpha \sim \beta$ and α, β intersect in j vertices. As $\alpha \neq \beta$, $f_v < e$. When $2 \leq j \leq v - 1$ the critical observation is that $A_\alpha \cap A_\beta$ is a subgraph of H and hence, as H is strictly balanced,

$$\frac{f_j}{j} < \frac{e}{v}.$$

There are $O(n^{2v-j})$ choices of α, β intersecting in j points, since α, β are determined, except for order, by $2v - j$ points. For each such α, β

$$\Pr[B_\alpha \wedge B_\beta] = p^{|A_\alpha \cup A_\beta|} = p^{2e - |A_\alpha \cap A_\beta|} \leq p^{2e - f_j}.$$

Thus

$$\Delta = \sum_{j=2}^{v} O(n^{2v-j}) O(n^{-\frac{v}{e}(2e - f_j)}).$$

But

$$2v - j - \frac{v}{e}(2e - f_j) = \frac{v f_j}{e} - j < 0,$$

so each term is $o(1)$ and hence $\Delta = o(1)$. By Janson's Inequality

$$\lim_{n \to \infty} \Pr[\wedge \overline{B}_\alpha] = \lim_{n \to \infty} M = exp[-c^e/a],$$

completing the proof. ■

10.2 CLIQUE NUMBER

In this section we fix $p = 1/2$, (other values yield similar results) and consider the clique number $\omega(G(n,p))$. For a fixed $c > 0$ let $n, k \to \infty$ so that

$$\binom{n}{k} 2^{-\binom{k}{2}} \to c.$$

As a first approximation

$$n \sim \frac{k}{e\sqrt{2}} \sqrt{2}^k$$

and

$$k \sim \frac{2 \ln n}{\ln 2}.$$

Here $\mu \to c$ so $M \to e^{-c}$. The Δ term was examined in §4.5. For this k, $\Delta = o(E[X]^2)$ and so $\Delta = o(1)$. Therefore

$$\lim_{n,k \to \infty} \Pr[\omega(G(n,p)) < k] = exp[-c].$$

Being more careful, let $n_0(k)$ be the minimum n for which

$$\binom{n}{k} 2^{-\binom{k}{2}} \geq 1.$$

Observe that for this n the left hand side is $1 + o(1)$. Note that $\binom{n}{k}$ grows, in n, like n^k. For any $\lambda \in (-\infty, +\infty)$ if

$$n = n_0(k)[1 + \frac{\lambda + o(1)}{k}]$$

then

$$\binom{n}{k} 2^{-\binom{k}{2}} = [1 + \frac{\lambda + o(1)}{k}]^k = e^{\lambda} + o(1),$$

and so

$$\Pr[\omega(G(n,p)) < k] = e^{-e^{\lambda}} + o(1).$$

As λ ranges from $-\infty$ to $+\infty$, $e^{-e^{\lambda}}$ ranges from 1 to 0. As $n_0(k+1) \sim \sqrt{2} n_0(k)$ the ranges will not "overlap" for different k. More precisely, let K be arbitrarily large and set

$$I_k = [n_0(k)[1 - \frac{K}{k}], n_0(k)[1 + \frac{K}{k}]].$$

For $k \geq k_0(K)$, $I_{k-1} \cap I_k = \emptyset$. Suppose $n \geq n_0(k_0(K))$. If n lies between the intervals (which occurs for "most" n), which we denote by $I_k < n < I_{k+1}$, then

$$\Pr[\omega(G(n,p)) < k] \leq e^{-e^{K}} + o(1),$$

nearly zero, and

$$\Pr[\omega(G(n,p)) < k+1] \geq e^{-e^{-K}} + o(1),$$

nearly one, so that

$$\Pr[\omega(G(n,p)) = k] \geq e^{-e^{-K}} - e^{-e^{K}} + o(1),$$

nearly one. When $n \in I_k$ we still have $I_{k-1} < n < I_{k+1}$ so that

$$\Pr[\omega(G(n,p)) = k \text{ or } k-1] \geq e^{-e^{-K}} - e^{-e^{K}} + o(1),$$

nearly one. As K may be made arbitrarily large this yields the celebrated two point concentration theorem on clique number, Corollary 4.5.2 in §4.5. Note, however, that for most n the concentration of $\omega(G(n, 1/2))$ is actually on a single value!

10.3 CHROMATIC NUMBER

In this section we fix $p = 1/2$ (there are similar results for other p) and let G be the random graph $G(n, 1/2)$. We shall find bounds on the chromatic number $\chi(G)$. A different derivation of the main result of this section is presented in Chapter 7, Section 7.3. Set

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

Let $k_0 = k_0(n)$ be that value for which

$$f(k_0 - 1) > 1 > f(k_0).$$

Then $n = \sqrt{2}^{k(1+o(1))}$ so for $k \sim k_0$,

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

Set

$$k = k(n) = k_0(n) - 4$$

so that

$$f(k) > n^{3+o(1)}.$$

Now we use the Extended Janson Inequality (Theorem 8.1.2) to estimate $\Pr[\omega(G) < k]$. Here $\mu = f(k)$. (Note that Janson's Inequality gives a lower bound of $2^{-f(k)} = 2^{-n^{3+o(1)}}$ to this probability but this is way off the mark since with probability $2^{-\binom{n}{2}}$ the random G is empty!) The value Δ was examined in §4.5 where

$$\frac{\Delta}{\mu^2} = \frac{\Delta^*}{\mu} = \sum_{i=2}^{k-1} g(i).$$

There $g(2) \sim k^4/n^2$ and $g(k-1) \sim 2kn2^{-k}/\mu$ were the dominating terms. In our instance $\mu > n^{3+o(1)}$ and $2^{-k} = n^{-2+o(1)}$ so $g(2)$ dominates and

$$\Delta \sim \frac{\mu^2 k^4}{n^2}.$$

Hence we bound the *clique* number probability

$$\Pr[\omega(G) < k] < e^{-\mu^2/2\Delta} = e^{-\Theta(n^2/(\ln n)^4)},$$

as $k = \Theta(\ln n)$. (The possibility that G is empty gives a lower bound so that we may say the probability is $e^{-n^{2+o(1)}}$, though a $o(1)$ in the hyperexponent leaves lots of room.)

Theorem 10.3.1 [B. Bollobás (1988)] *Almost always*

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

Proof. Let $\alpha(G) = \omega(\overline{G})$ denote, as usual, the independence number of G. The complement of G has the same distribution $G(n, 1/2)$. Hence $\alpha(G) \le (2 + o(1)) \log_2 n$ almost always. Thus

$$\chi(G) \ge \frac{n}{\alpha(G)} \ge \frac{n}{2\log_2 n}(1 + o(1))$$

almost always.

The reverse inequality was an open question for a full quarter century! Set $m = \lfloor n/\ln^2 n \rfloor$. For any set S of m vertices the restriction $G|_S$ has the distribution of $G(m, 1/2)$. Let $k = k(m) = k_0(m) - 4$ as above. Note

$$k \sim 2\log_2 m \sim 2\log_2 n.$$

Then

$$\Pr[\alpha[G|_S] < k] < e^{-m^{2+o(1)}}.$$

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

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

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

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

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

and this occurs for almost all G. ■

10.4 BRANCHING PROCESSES

Paul Erdős and Alfred Rényi, in their original 1960 paper, discovered that the random graph $G(n, p)$ undergoes a remarkable change at $p = 1/n$. Speaking roughly, let first $p = c/n$ with $c < 1$. Then $G(n, p)$ will consist of small components, the largest of which is of size $\Theta(\ln n)$. But now suppose $p = c/n$ with $c > 1$. In that short amount of "time" many of the components will have joined together to form a "giant component" of size $\Theta(n)$. The remaining vertices are still in small components, the largest of which has size $\Theta(\ln n)$. They dubbed this phenomenon the *Double Jump*. We prefer the descriptive term Phase Transition because of the connections to percolation (e.g., freezing) in mathematical physics.

To better understand the Phase Transition we make a lengthy detour into the subject of Branching Processes. Imagine that we are in a unisexual universe and we start with a single organism. Imagine that this organism has a number of children given by a given random variable Z. (For us, Z will be Poisson with mean c.) These children then themselves have children, the number again being determined by Z. These grandchildren then have children, etc. As $Z = 0$ will have nonzero probability there will be some chance that the line dies out entirely. We want to study the total number of organisms in this process, with particular eye to whether or not the process continues forever. (The original application of this model was to a study of the -gasp!- male line of British peerage.)

Now lets be more precise. Let $Z_1, Z_2, \ldots$ be independent random variables, each with distribution Z. Define $Y_0, Y_1, \ldots$ by the recursion

$$Y_0 = 1$$

$$Y_i = Y_{i-1} + Z_i - 1$$

and let T be the least t for which $Y_t = 0$. If no such t exists (the line continuing forever) we say $T = +\infty$. The Y_i and Z_i mirror the Branching Process as follows. We view all organisms as living or dead. Initially there is one live organism and no dead ones. At each time unit we select one of the live organisms, it has Z_i children, and then it dies. The number Y_i of live organisms at time i is then given by the recursion. The process stops when $Y_t = 0$ (extinction) but it is a convenient fiction to define the recursion for all t. Note that T is not affected by this fiction since once $Y_t = 0$, T has been defined. T (whether finite or infinite) is the total number of organisms, including the original, in this process. (A natural approach, found in many probability texts, is to have all organisms of a given generation have their children at once and study the number of children of each generation. While we may think of the organisms giving birth by generation it will not affect our model.)

A major result of Branching Processes is that when $E[Z] = c < 1$ with probability one the process dies out ($T < \infty$) but when $E[Z] = c > 1$ then there is a nonzero probability that the process goes on forever ($T = \infty$). Intuitively this makes sense: the values Y_i of the population form a Markov Chain. When $c < 1$ there is drift "to the left" so eventually $Y_i = 0$ whereas when $c > 1$ there is drift "to the right", $Y_i \to +\infty$, and with finite probability the population never hits zero. If the process doesn't die early the population is likely to just keep growing. We give a proof based on the Large Deviation bounds A.1.15 . Observe that $Z_1 + \ldots + Z_t$ has a Poisson distribution with mean ct. First suppose $c < 1$. For any t

$$\Pr[T > t] \le \Pr[Y_t > 0] = \Pr[Z_1 + \ldots + Z_t \ge t] \le (1 - \delta)^t$$

with $\delta > 0$. As $\lim_{t \to \infty} \Pr[T > t] = 0$, $\Pr[T = \infty] = 0$. Now suppose $c > 1$. Again using A.1.15

$$\Pr[Y_t \le 0] = \Pr[Z_1 + \ldots + Z_t \le t] \le (1 - \delta)^t,$$

with a different $\delta > 0$. As $\sum_{t=1}^{\infty} (1 - \delta)^t$ converges there is a t_0 with

$$\sum_{t=t_0}^{\infty} \Pr[Y_t \le 0] < 1$$

Then

$$\sum_{t=0}^{\infty} \Pr[Y_t + t_0 - 1 \le 0] < 1$$

since for $t < t_0$, $Y_t + t_0 - 1 \ge 1 - t + t_0 - 1 > 0$ always. We now condition on the first organism having t_0 children. The conditional distribution of

$$Y_t = t_0 + (Z_2 - 1) + \ldots + (Z_t - 1)$$

is the unconditional distribution of

$$Y_{t-1} + (t_0 - 1) = t_0 + (Z_1 - 1) + \ldots + (Z_{t-1} - 1),$$

and so

$$\sum_{t=0}^{\infty} \Pr[Y_t \le 0 | Z_1 = t_0] < 1.$$

With positive probability $Z_1 = t_0$ and then with positive probability all $Y_t > 0$ so that $\Pr[T = \infty] > 0$.

Generating functions give a more precise result. Let

$$p_i = \Pr[Z_j = i]$$

and define the generating function

$$p(x) = \sum_{i=0}^{\infty} p_i x^i.$$

In our case

$$p_i = e^{-c} c^i / i!,$$

so that

$$p(x) = \sum_{i=0}^{\infty} e^{-c} c^i x^i / i! = e^{c(x-1)}.$$

Let

$$q_i = \Pr[T = i]$$

and set

$$q(x) = \sum_{i=0}^{\infty} q_i x^i$$

(the sum not including the $i = \infty$ case). Conditioning on the first organism having s children the generating function for the total number of offspring is $x(q(x))^s$. Hence

$$q(x) = \sum_{s=0}^{\infty} p_s x q(x)^s = x p[q(x)].$$

That is, the generating function $y = q(x)/x$ satisfies the functional equality $y = p(xy)$, i.e.,

$$y = e^{c(xy-1)}.$$

The extinction probability

$$y = \Pr[T < \infty] = \sum_{i=0}^{\infty} q_i = q(1) = q(1)/1$$

must satisfy

$$y = e^{c(y-1)} \qquad (**)$$

For $c < 1$ this has the unique solution $y = 1$, corresponding to the certain extinction. (In fact, for $c = 1$ the only solution is also $y = 1$, proving extinction in the critical case as well). For $c > 1$ there are two solutions, $y = 1$ and some $y \in (0, 1)$. For $c > 1$ we let $f(c)$ denote that y satisfying (**), $0 < y < 1$. As $\Pr[T < \infty] < 1$ we know

$$\Pr[T < \infty] = f(c).$$

When a branching process dies, we call $H = (Z_1, \ldots, Z_T)$ the *history* of the process. A sequence $(z_1, \ldots, z_t)$ is a possible history if and only if the sequence y_i given by $y_0 = 1, y_i = y_{i-1} + z_i - 1$ has $y_i > 0$ for $0 \le i < t$ and $y_t = 0$. When Z is Poisson with mean λ

$$\Pr[H = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \frac{e^{-\lambda}\lambda^{z_i}}{z_i!} = \frac{e^{-\lambda}(\lambda e^{-\lambda})^{t-1}}{\prod_{i=1}^{t} z_i!},$$

since $z_1 + \ldots + z_t = t - 1$.

We call $d < 1 < c$ a conjugate pair if

$$de^{-d} = ce^{-c}.$$

The function $f(x) = xe^{-x}$ increases from 0 to e^{-1} in [0,1) and decreases back to 0 in $(1, \infty)$ so that each $c \neq 1$ has a unique conjugate. Let $c > 1$ and $y = \Pr[T < \infty]$ so that $y = e^{c(y-1)}$. Then $(cy)e^{-cy} = ce^{-c}$, so

$$d = cy.$$

Duality Principle. Let $d < 1 < c$ be conjugates. The Branching Process with mean c, conditional on extinction, has the same distribution as the Branching Process with mean d.

Proof. It suffices to show that for every history $H = (z_1, \ldots, z_t)$

$$\frac{e^{-c}(ce^{-c})^{t-1}}{y\prod_{i=1}^{t} z_i!} = \frac{e^{-d}(de^{-d})^{t-1}}{\prod_{i=1}^{t} z_i!}$$

This is immediate as $ce^{-c} = de^{-d}$ and $ye^{-d} = ye^{-cy} = e^{-c}$. ■

10.5 THE GIANT COMPONENT

Now let's return to random graphs. We define a procedure to find the component $C(v)$ containing a given vertex v in a given graph G. We are motivated by Karp (1990) , in which this approach is applied to random digraphs. In this procedure vertices will be live, dead or neutral. Originally v is live and all other vertices are neutral, time $t = 0$ and $Y_0 = 1$. Each time unit t we take a live vertex w and check all pairs $\{w, w'\}$, w' neutral, for membership in G. If $\{w, w'\} \in G$ we make w' live, otherwise it stays neutral. After searching all neutral w' we set w dead and let Y_t equal the new number of live vertices. When there are no live vertices the process terminates and $C(v)$ is the set of dead vertices. Let Z_t be the number of w' with $\{w, w'\} \in G$ so that

$$Y_0 = 1$$

$$Y_t = Y_{t-1} + Z_t - 1.$$

With $G = G(n,p)$ each neutral w' has independent probability p of becoming live. Here, critically, no pair $\{w, w'\}$ is ever examined twice, so that the conditional probability for $\{w, w'\} \in G$ is always p. As $t-1$ vertices are dead and Y_{t-1} are live

$$Z_t \sim B[n - (t-1) - Y_{t-1}, p].$$

Let T be the least t for which $Y_t = 0$. Then $T = |C(v)|$. As in Section 10.4 we continue the recursive definition of Y_t, this time for $0 \le t \le n$.

Claim 3 *For all t*

$$Y_t \sim B[n-1, 1-(1-p)^t] + 1 - t.$$

Proof. It is more convenient to deal with

$$N_t = n - t - Y_t$$

the number of neutral vertices at time t and show, equivalently,

$$N_t \sim B[n-1, (1-p)^t].$$

This is reasonable since each $w \neq v$ has independent probability $(1-p)^t$ of staying neutral t times. Formally, as $N_0 = n - 1$ and

$$\begin{aligned} N_t \quad = n - t - Y_t \quad &= n - t - B[n - (t-1) - Y_{t-1}, p] - Y_{t-1} + 1 \\ &= N_{t-1} - B[N_{t-1}, p] \\ &= B[N_{t-1}, 1-p], \end{aligned}$$

the result follows by induction. ■

We set $p = c/n$. When t and Y_{t-1} are small we may approximate Z_t by $B[n, c/n]$ which is approximately Poisson with mean c. Basically small components will have size distribution as in the Branching Process of Section 10.4. The analogy must break down for $c > 1$ as the Branching Process may have an infinite population whereas

$|C(v)|$ is surely at most n. Essentially, those v for which the Branching Process for $C(v)$ does not "die early" all join together to form the giant component.

Fix c. Let $Y_0^*, Y_1^*, \ldots, T^*, Z_1^*, Z_2^*, \ldots, H^*$ refer to the Branching Process of §5.4 and $Y_0, Y_1, \ldots, T, Z_1, Z_2, \ldots, H$ refer to the Random Graph process. For any possible history $(z_1, \ldots, z_t)$

$$\Pr[H^* = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \Pr[Z^* = z_i],$$

where Z^* is Poisson with mean c while

$$\Pr[H = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \Pr[Z_i = z_i],$$

where Z_i has Binomial Distribution $B[n - 1 - z_1 - \ldots - z_{i-1}, c/n]$. The Poisson distribution is the limiting distribution of Binomials. When $m = m(n) \sim n$ and c, i are fixed

$$\lim_{n\to\infty} \Pr[B[m, c/n] = z] = \lim_{n\to\infty} \binom{m}{z} (\frac{c}{n})^z (1 - \frac{c}{n})^{m-z} = e^{-c} c^z / z!.$$

Hence

$$\lim_{n\to\infty} \Pr[H = (z_1, \ldots, z_t)] = \Pr[H^* = (z_1, \ldots, z_t)].$$

Assume $c < 1$. For any fixed t, $\lim_{n\to\infty} \Pr[T = t] = \Pr[T^* = t]$. We now bound the size of the largest component. For any t

$$\Pr[T > t] \leq \Pr[Y_t > 0] = \Pr[B[n-1, 1-(1-p)^t] \geq t] \leq \Pr[B[n, tc/n] \geq t],$$

as $1 - (1-p)^t \leq tp$ and $n - 1 < n$. By Large Deviation Result A.1.14

$$\Pr[T > t] < e^{-\alpha t}$$

where $\alpha = \alpha(c) > 0$. Let $\beta = \beta(c)$ satisfy $\alpha\beta > 1$. Then

$$\Pr[T > \beta \ln n] < n^{-\alpha\beta} = o(n^{-1}).$$

There are n choices for initial vertex v. Thus almost always *all* components have size $O(\ln n)$.

Now assume $c > 1$. For any fixed t, $\lim_{n\to\infty} \Pr[T = t] = \Pr[T^* = t]$ but what corresponds to $T^* = \infty$? For $t = o(n)$ we may estimate $1 - (1-p)^t \sim pt$ and $n - 1 \sim n$ so that

$$\Pr[Y_t \leq 0] = \Pr[B[n-1, 1-(1-p)^t] \leq t - 1] \sim \Pr[B[n, tc/n] \leq t]$$

drops exponentially in t by Large Deviation results. When $t = \alpha n$ we estimate $1 - (1-p)^t$ by $1 - e^{-c\alpha}$. The equation $1 - e^{-c\alpha} = \alpha$ has solution $\alpha = 1 - y$ where y is the extinction probability of §10.4.

For $\alpha < 1 - y$, $1 - e^{-c\alpha} > \alpha$ and

$$\Pr[Y_t \leq 0] \sim \Pr[B[n, 1 - e^{-c\alpha}] \leq \alpha n]$$

is exponentially small, while for $\alpha > 1 - y$, $1 - e^{-c\alpha} < \alpha$ and $\Pr[Y_t \leq 0] \sim 1$. Thus almost always $Y_t = 0$ for some $t \sim (1-y)n$. Basically, $T^* = \infty$ corresponds to $T \sim (1-y)n$. Let $\epsilon, \delta > 0$ be arbitrarily small. With somewhat more care to the bounds we may show that there exists t_0 so that for n sufficiently large

$$\Pr[t_0 < T < (1-\delta)n(1-y) \text{ or } \mathrm{T} > (1+\delta)\mathrm{n}(1-\mathrm{y})] < \epsilon.$$

Pick t_0 sufficiently large so that

$$y - \epsilon \leq \Pr[T^* \leq t_0] \leq y.$$

Then as $\lim_{n\to\infty} \Pr[T \leq t_0] = \Pr[T^* \leq t_0]$ for n sufficiently large

$$y - 2\epsilon \leq \Pr[T \leq t_0] \leq y + \epsilon$$

$$1 - y - 2\epsilon \leq \Pr[(1-\delta)n(1-y) < T < (1+\delta)n(1-y)] < 1 - y + 3\epsilon.$$

Now we expand our procedure to find graph components. We start with $G \sim G(n,p)$, select $v = v_1 \in G$ and compute $C(v_1)$ as before. Then we delete $C(v_1)$, pick $v_2 \in G - C(v_1)$ and iterate. At each stage the remaining graph has distribution $G(m,p)$ where m is the number of vertices. (Note, critically, that no pairs $\{w, w'\}$ in the remaining graph have been examined and so it retains its distribution.) Call a component $C(v)$ small if $|C(v)| \leq t_0$, giant if $(1-\delta)n(1-y) < |C(v)| < (1+\delta)n(1-y)$ and otherwise failure. Pick $s = s(\epsilon)$ with $(y+\epsilon)^s < \epsilon$. (For ϵ small $s \sim K \ln \epsilon^{-1}$.) Begin this procedure with the full graph and terminate it when either a giant component or a failure component is found or when s small components are found. At each stage, as only small components have thus far been found, the number of remaining points is $m = n - O(1) \sim n$ so the conditional probabilities of small, giant and failure remain asymptotically the same. The chance of ever hitting a failure component is thus $\leq s\epsilon$ and the chance of hitting all small components is $\leq (y+\epsilon)^s \leq \epsilon$ so that with probability at least $1 - \epsilon'$, where $\epsilon' = (s+1)\epsilon$ may be made arbitrarily small, we find a series of less than s small components followed by a giant component. The remaining graph has $m \sim yn$ points and $pm \sim cy = d$, the conjugate of c as defined in §10.4. As $d < 1$ the previous analysis gives the maximal components. In summary: almost always $G(n, c/n)$ has a giant component of size $\sim (1-y)n$ and all other components of size $O(\ln n)$. Furthermore, the Duality Principle of §10.4 has a discrete analog.

Discrete Duality Principle. Let $d < 1 < c$ be conjugates. The structure of $G(n, c/n)$ with its giant component removed is basically that of $G(m, d/m)$ where m, the number of vertices not in the giant component, satisfies $m \sim ny$.

The small components of $G(n, c/n)$ can also be examined from a static view. For a fixed k let X be the number of tree components of size k. Then

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

Here we use the nontrivial fact, due to Cayley, that there are k^{k-2} possible trees on a given k-set. For c, k fixed

$$E[X] \sim n\frac{e^{-ck}k^{k-2}c^{k-1}}{k!}.$$

As trees are strictly balanced a second moment method gives $X \sim E[X]$ almost always. Thus $\sim p_k n$ points lie in tree components of size k where

$$p_k = \frac{e^{-ck}(ck)^{k-1}}{k!}$$

It can be shown analytically that $p_k = \Pr[T = k]$ in the Branching Process with mean c of §10.4. Let Y_k denote the number of cycles of size k and Y the total number of cycles. Then

$$E[Y_k] = \frac{(n)_k}{2k}(\frac{c}{n})^k \sim \frac{c^k}{2k}$$

for fixed k. For $c < 1$

$$E[Y] = \sum E[Y_k] \to \sum_{k=1}^{\infty} \frac{c^k}{2k}$$

has a finite limit, whereas for $c > 1$, $E[Y] \to \infty$. Even for $c > 1$ for any fixed k the number of k-cycles has a limiting expectation and so does not asymptotically affect the number of components of a given size.

10.6 INSIDE THE PHASE TRANSITION

In the evolution of the random graph $G(n,p)$ a crucial change takes place in the vicinity of $p = c/n$ with $c = 1$. The small components at that time are rapidly joining together to form a giant component. This corresponds to the Branching Process when births are Poisson with mean 1. There the number T of organisms will be finite almost always and yet have infinite expectation. No wonder that the situation for random graphs is extremely delicate. In recent years there has been much interest in looking "inside" the phase transition at the growth of the largest components. (See, e.g. Łuczak (1990) or the monumental Janson, Knuth, Łuczak and Pittel (1993)). The appropriate parametrization is, perhaps surprisingly,

$$p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}.$$

When $\lambda = \lambda(n) \to -\infty$ the phase transition has not yet started. The largest components are $o(n^{2/3})$ and there are many components of nearly the largest size. When $\lambda = \lambda(n) \to +\infty$ the phase transition is over - a largest component, of size $>> n^{2/3}$ has emerged and all other components are of size $o(n^{2/3})$. Let's fix λ, c and let X be the number of tree components of size $k = cn^{2/3}$. Then

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

Watch the terms cancel!

$$\binom{n}{k} = \frac{(n)_k}{k!} \sim \frac{n^k e^k}{k^k \sqrt{2\pi k}} \prod_{i=1}^{k-1} (1 - \frac{i}{n})$$

For $i < k$

$$-\ln(1 - \frac{i}{n}) = \frac{i}{n} + \frac{i^2}{2n^2} + O(\frac{i^3}{n^3}),$$

so that

$$\sum_{i=1}^{k-1} - \ln(1 - \frac{i}{n}) = \frac{k^2}{2n} + \frac{k^3}{6n^2} + o(1) = \frac{k^2}{2n} + \frac{c^3}{6} + o(1).$$

Also

$$p^{k-1} = n^{1-k}(1 + \frac{\lambda}{n^{1/3}})^{k-1},$$

$$(k-1)\ln(1 + \frac{\lambda}{n^{1/3}}) = (k-1)(\frac{\lambda}{n^{1/3}} - \frac{\lambda^2}{2n^{2/3}} + O(n^{-1})) = \frac{\lambda k}{n^{1/3}} - \frac{\lambda^2 c}{2} + o(1).$$

Also

$$\ln(1-p) = -p + O(n^{-2}) = -\frac{1}{n} - \frac{\lambda}{n^{4/3}} + O(n^{-2})$$

and

$$k(n-k) + \binom{k}{2} - (k-1) = kn - \frac{k^2}{2} + O(n^{2/3}),$$

so that

$$[k(n-k) + \binom{k}{2} - (k-1)]\ln(1-p) = -k + \frac{k^2}{2n} - \frac{\lambda k}{n^{1/3}} + \frac{\lambda c^2}{2} + o(1)$$

and

$$E[X] \sim \frac{n^k k^{k-2}}{k^k \sqrt{2\pi k} n^{k-1}} e^A,$$

where

$$\begin{aligned} A &= k - \tfrac{k^2}{2n} - \tfrac{c^3}{6} + \tfrac{\lambda k}{n^{1/3}} - \tfrac{\lambda^2 c}{2} - k + \tfrac{k^2}{2n} - \tfrac{\lambda k}{n^{1/3}} + \tfrac{\lambda c^2}{2} + o(1) \\ &= -\tfrac{c^3}{6} - \tfrac{\lambda^2 c}{2} + \tfrac{\lambda c^2}{2} + o(1), \end{aligned}$$

so that

$$E[X] \sim n^{-2/3} e^{-\frac{c^3}{6} - \frac{\lambda^2 c}{2} + \frac{\lambda c^2}{2}} c^{-5/2} (2\pi)^{-1/2}.$$

For any particular such k, $E[X] \to 0$ but if we sum k between $cn^{2/3}$ and $(c+dc)n^{2/3}$ we multiply by $n^{2/3}dc$. Going to the limit gives an integral: For any fixed a, b, λ let X be the number of tree components of size between $an^{2/3}$ and $bn^{2/3}$. Then

$$\lim_{n\to\infty} E[X] = \int_a^b e^{-\frac{c^3}{6} - \frac{\lambda^2 c}{2} + \frac{\lambda c^2}{2}} c^{-5/2} (2\pi)^{-1/2} dc.$$

The large components are not all trees. Wright (1977) proved that for fixed l there are asymptotically $c_l k^{k-2+\frac{3}{2}l}$ connected graphs on k points with $k-1+l$ edges, where c_l was given by a specific recurrence. Asymptotically in l, $c_l = l^{-l/2(1+o(1))}$. The calculation for $X^{(l)}$, the number of such components on k vertices, leads to extra factors of $c_l k^{\frac{3}{2}l}$ and n^{-l} which gives $c_l c^{\frac{3}{2}l}$. For fixed a, b, λ, l the number $X^{(l)}$ of components of size between $an^{2/3}$ and $bn^{2/3}$ with $l-1$ more edges than vertices satisfies

$$\lim_{n\to\infty} E[X^{(l)}] = \int_a^b e^{-\frac{c^3}{6}-\frac{\lambda^2 c}{2}+\frac{\lambda c^2}{2}} c^{-5/2}(2\pi)^{-1/2}(c_l c^{\frac{3}{2}l})dc,$$

and letting X^* be the total number of components of size between $an^{2/3}$ and $bn^{2/3}$

$$\lim_{n\to\infty} E[X^*] = \int_a^b e^{-\frac{c^3}{6}-\frac{\lambda^2 c}{2}+\frac{\lambda c^2}{2}} c^{-5/2}(2\pi)^{-1/2} g(c)dc,$$

where

$$g(c) = \sum_{l=0}^{\infty} c_l c^{\frac{3}{2}l},$$

a sum convergent for all c, (here $c_0 = 1$). A component of size $\sim cn^{2/3}$ will have probability $c_l c^{\frac{3}{2}l}/g(c)$ of having $l-1$ more edges than vertices, independent of λ. As $\lim_{c\to 0} g(c) = 1$, most components of size $\epsilon n^{2/3}$, $\epsilon << 1$, are trees but as c gets bigger the distribution on l moves inexoribly higher.

An Overview. For any fixed λ the sizes of the largest components are of the form $cn^{2/3}$ with a distribution over the constant. For $\lambda = -10^6$ there is some positive limiting probability that the largest component is bigger than $10^6 n^{2/3}$ and for $\lambda = +10^6$ there is some positive limiting probability that the largest component is smaller than $10^{-6}n^{2/3}$, though both these probabilities are miniscule. The functions integrated have a pole at $c = 0$, reflecting the notion that for any λ there should be many components of size near $\epsilon n^{2/3}$ for $\epsilon = \epsilon(\lambda)$ appropriately small. When λ is large negative (e.g., -10^6) the largest component is likely to be $\epsilon n^{2/3}$, ϵ small, and there will be many components of nearly that size. The nontree components will be a negligible fraction of the tree components.

Now consider the evolution of $G(n,p)$ in terms of λ. Suppose that at a given λ there are components of size $c_1 n^{2/3}$ and $c_2 n^{2/3}$. When we move from λ to $\lambda + d\lambda$ there is a probability $c_1 c_2 d\lambda$ that they will merge. Components have a peculiar gravitation in which the probability of merging is proportional to their sizes. With probability $(c_1^2/2)d\lambda$ there will be a new internal edge in a component of size $c_1 n^{2/3}$ so that large components rarely remain trees. Simultaneously, big components are eating up other vertices.

With $\lambda = -10^6$, say, we have feudalism. Many small components (castles) are each vying to be the largest. As λ increases the components increase in size and a few large components (nations) emerge. An already large France has much better chances of becoming larger than a smaller Andorra. The largest components tend strongly to merge and by $\lambda = +10^6$ it is very likely that a giant component, Roman

Empire, has emerged. With high probability this component is nevermore challenged for supremacy but continues absorbing smaller components until full connectivity - One World - is achieved.

10.7 ZERO-ONE LAWS

In this section we restrict our attention to graph theoretic properties expressible in the First Order theory of graphs. The language of this theory consists of variables $(x, y, z, \ldots)$, which always represent vertices of a graph, equality and adjacency ($x = y, x \sim y$), the usual Boolean connectives $(\wedge, \neg, \ldots)$ and universal and existential quanfication $(\forall_x, \exists_y)$. Sentences must be finite. As examples, one can express the property of containing a triangle

$$\exists_x \exists_y \exists_z [x \sim y \wedge x \sim z \wedge y \sim z].$$

having no isolated point

$$\forall_x \exists_y [x \sim y]$$

and having radius at most two

$$\exists_x \forall_y [\neg(y = x) \wedge \neg(y \sim x) \longrightarrow \exists_z [z \sim y \wedge y \sim x]].$$

For any property A and any n, p we consider the probability that the random graph $G(n, p)$ satisfies A, denoted

$$\Pr[G(n,p) \models A].$$

Our objects in this section will be the theorem of Glebskii, Kogan, Liagonkii and Talanov (1969) and independently Fagin (1976) (theorem 10.7.1), and that of Shelah and Spencer (1988) (theorem 10.7.2).

Theorem 10.7.1 *For any fixed p, $0 < p < 1$ and any First Order A*

$$\lim_{n \to \infty} \Pr[G(n,p) \models A] = 0 \textit{ or } 1.$$

Theorem 10.7.2 *For any* irrational α, $0 < \alpha < 1$, *setting $p = p(n) = n^{-\alpha}$, and for any First Order A,*

$$\lim_{n \to \infty} \Pr[G(n,p) \models A] = 0 \textit{ or } 1.$$

Both proofs are only outlined.

We shall say that a function $p = p(n)$ satisfies the Zero-One Law if the above equality holds for every First Order A.

The Glebskii/Fagin Theorem has a natural interpretation when $p = .5$ as then $G(n, p)$ gives equal weight to every (labelled) graph. It then says that any First Order property A holds for either almost all graphs or for almost no graphs. The Shelah/Spencer Theorem may be interpreted in terms of threshold functions. The general results of Section 10.1 give, as one example, that $p = n^{-2/3}$ is a threshold function for containment of a K_4. That is, when $p << n^{-2/3}$, $G(n,p)$ almost surely does not contain a K_4 whereas when $p >> n^{-2/3}$ it almost surely does contain a K_4. In between, say at $p = n^{-2/3}$, the probability is between 0 and 1, in this case $1 - e^{-1/24}$. The (admittedly rough) notion is that *at* a threshold function the Zero-One Law will not hold and so to say that $p(n)$ satisfies the Zero-One Law is to say that $p(n)$ is not a threshold function - that it is a boring place in the evolution of the random graph, at least through the spectacles of the First Order language. In stark terms: What happens in the evolution of $G(n,p)$ at $p = n^{-\pi/7}$? The answer: Nothing!

Our approach to Zero-One Laws will be through a variant of the Ehrenfeucht Game, which we now define. Let G, H be two vertex disjoint graphs and t a positive integer. We define a perfect information game, denoted $EHR[G,H,t]$, with two players, denoted Spoiler and Duplicator. The game has t rounds. Each round has two parts. First the Spoiler selects either a vertex $x \in V(G)$ or a vertex $y \in V(H)$. He chooses which graph to select the vertex from. Then the Duplicator must select a vertex in the other graph. At the end of the t rounds t vertices have been selected from each graph. Let $x_1, \ldots, x_t$ be the vertices selected from $V(G)$ and $y_1, \ldots, y_t$ be the vertices selected from $V(H)$ where x_i, y_i are the vertices selected in the i-th round. Then Duplicator wins if and only if the induced graphs on the selected vertices are order-isomorphic: i.e., if for all $1 \le i < j \le t$

$$\{x_i, x_j\} \in E(G) \longleftrightarrow \{y_i, y_j\} \in E(H).$$

As there are no hidden moves and no draws one of the players must have a winning strategy and we will say that that player wins $EHR[G,H,t]$.

Lemma 10.7.3 *For every First Order A there is a $t = t(A)$ so that if G, H are any graphs with $G \models A$ and $H \models \neg A$ then Spoiler wins $EHR[G,H,t]$.*

A detailed proof would require a formal analysis of the First Order language so we give only an example. Let A be the property $\forall_x \exists_y [x \sim y]$ of not containing an isolated point and set $t = 2$. Spoiler begins by selecting an isolated point $y_1 \in V(H)$ which he can do as $H \models \neg A$. Duplicator must pick $x_1 \in V(G)$. As $G \models A$, x_1 is not isolated so Spoiler may pick $x_2 \in V(G)$ with $x_1 \sim x_2$ and now Duplicator cannot pick a "duplicating" y_2.

Theorem 10.7.4 *A function $p = p(n)$ satisfies the Zero-One Law if and only if for every t, letting $G(n, p(n)), H(m, p(m))$ be independently chosen random graphs on disjoint vertex sets*

$$\lim_{m,n\to\infty} \Pr[\, \textit{Duplicator wins} EHR[G(n,p(n)), H(m,p(m)), t]] = 1.$$

Remark. For any given choice of G, H somebody must win $EHR[G,H,t]$. (That is, there is no random play, the play is perfect.) Given this probability distribution over

(G, H) there will be a probability that $EHR[G, H, t]$ will be a win for Duplicator, and this must approach one.

Proof. We prove only the "if" part. Suppose $p = p(n)$ did not satisfy the Zero-One Law. Let A satisfy

$$\lim_{n\to\infty} \Pr[G(n, p(n)) \models A] = c,$$

with $0 < c < 1$. Let $t = t(A)$ be as given by the Lemma. With limiting probability $2c(1-c) > 0$ exactly one of $G(n, p(n)), H(n, p(n))$ would satisfy A and thus Spoiler would win, contradicting the assumption. This is not a full proof since when the Zero-one Law is not satisfied $\lim_{n\to\infty} \Pr[G(n, p(n)) \models A]$ might not exist. If there is a subsequence n_i on which the limit is $c \in (0, 1)$ we may use the same argument. Otherwise there will be two subsequences n_i, m_i on which the limit is zero and one respectively. Then letting $n, m \to \infty$ through n_i, m_i respectively, Spoiler will win $EHR[G, H, t]$ with probability approaching one. ■

Theorem 10.7.4 provides a bridge from Logic to Random Graphs. To prove that $p = p(n)$ satisfies the Zero-One Law we now no longer need to know anything about Logic - we just have to find a good strategy for the Duplicator.

We say that a graph G has the full level s extension property if for every distinct $u_1, \ldots, u_a, v_1, \ldots, v_b \in G$ with $a + b \le s$ there is an $x \in V(G)$ with $\{x, u_i\} \in E(G)$, $1 \le i \le a$ and $\{x, v_j\} \notin E(G)$, $1 \le j \le b$. Suppose that G, H both have the full level $s - 1$ extension property. Then Duplicator wins $EHR[G, H, s]$ by the following simple strategy. On the i-th round, with $x_1, \ldots, x_{i-1}, y_1, \ldots, y_{i-1}$ already selected, and Spoiler picking, say, x_i, Duplicator simply picks y_i having the same adjacencies to the $y_j, j < i$ as x_i has to the $x_j, j < i$. The full extension property says that such a y_i will surely exist.

Theorem 10.7.5 *For any fixed p, $0 < p < 1$, and any s, $G(n, p)$ almost always has the full level s extension property.*

Proof. For every distinct $u_1, \ldots, u_a, v_1, \ldots, v_b, x \in G$ with $a + b \le s$ we define $E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}$ to be the event that $\{x, u_i\} \in E(G)$, $1 \le i \le a$ and $\{x, v_j\} \notin E(G)$, $1 \le j \le b$. Then

$$\Pr[E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}] = p^a(1-p)^b.$$

Now define

$$E_{u_1,\ldots,u_a,v_1,\ldots,v_b} = \wedge_x \overline{E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}$$

the conjunction over $x \ne u_1, \ldots, u_a, v_1, \ldots, v_b$. These events are mutually independent over x since they involve different edges. Thus

$$\Pr[\wedge_x \overline{E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}] = [1 - p^a(1-p)^b]^{n-a-b}.$$

Set $\epsilon = \min(p, 1-p)^s$ so that

$$\Pr[\wedge_x \overline{E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}] \le (1-\epsilon)^{n-s}.$$

The key here is that ϵ is a fixed (dependent on p, s) positive number. Set

$$E = \vee E_{u_1,\ldots,u_a,v_1,\ldots,v_b},$$

the disjunction over all distinct $u_1, \ldots, u_a, v_1, \ldots, v_b \in G$ with $a + b \leq s$. There are less than $s^2 n^s = O(n^s)$ such choices as we can choose a, b and then the vertices. Thus

$$\Pr[E] \leq s^2 n^s (1-\epsilon)^{n-s}.$$

But

$$\lim_{n\to\infty} s^2 n^s (1-\epsilon)^{n-s} = 0$$

and so E holds almost never. Thus $\neg E$, which is precisely the statement that $G(n,p)$ has the full level s extension property, holds almost always. ■

But now we have proven Theorem 10.7.1. For any $p \in (0,1)$ and any fixed s as $m, n \to \infty$ with probability approaching one both $G(n,p)$ and $H(m,p)$ will have the full level s extension property and so Duplicator will win $EHR[G(n,p), H(m,p), s]$.

Why can't Duplicator use this strategy when $p = n^{-\alpha}$? We illustrate the difficulty with a simple example. Let $.5 < \alpha < 1$ and let Spoiler and Duplicator play a three move game on G, H. Spoiler thinks of a point $z \in G$ but doesn't tell Duplicator about it. Instead he picks $x_1, x_2 \in G$, both adjacent to z. Duplicator simply picks $y_1, y_2 \in H$, either adjacent or not adjacent dependent on whether $x_1 \sim x_2$. But now wily Spoiler picks $x_3 = z$. $H \sim H(m, m^{-\alpha})$ does not have the full level 2 extension property. In particular, most pairs y_1, y_2 do not have a common neighbor. Unless Duplicator was lucky, or shrewd, he then cannot find $y_3 \sim y_1, y_2$ and so he loses. This example does not say that Duplicator will lose with perfect play - indeed, we will show that he almost always wins with perfect play - it only indicates that the strategy used need be more complex.

We begin our proof of the Zero-One Law Theorem 10.7.2. Let $\alpha \in (0,1)$, α irrational, be fixed. A *rooted graph* is a pair (R, H) where H is a graph on vertex set, say, $V(H) = \{X_1, \ldots, X_r, Y_1, \ldots, Y_v\}$ and $R = \{X_1, \ldots, X_r\}$ is a specified subset of $V(H)$, called the roots. For example, (R, H) might consist of one vertex Y_1 adjacent to the two roots X_1, X_2. Let $v = v(R, H)$ denote the number of vertices which are not roots and let $e = e(R, H)$ denote the number of edges, excluding those edges between two roots. We say (R, H) is *dense* if $v - e\alpha < 0$ and *sparse* if $v - e\alpha > 0$. The irrationality of α assures us that all (R, H) are in one of these categories. We call (R, H) *rigid* if for all S with $R \subseteq S \subset V(H)$, (S, H) is dense. We call (R, H) *safe* if for all S with $R \subset S \subseteq V(H)$, $(R, H|_S)$ is sparse. Several elementary properties of these concepts are given as Exercise 4. We sometimes write (R, S) for $(R, H|_S)$ when the graph H is understood.

We think of rooted graphs as on abstract points. In a graph G we say that vertices $y_1, \ldots, y_v$ form an (R, H) extension of $x_1, \ldots, x_r$ if whenever X_i is adjacent to Y_j in H, x_i is adjacent to y_j in G and also whenver Y_i and Y_j are adjacent in H, y_i and y_j are adjacent in G. Note that we allow G to have more edges than H and that the edges between the roots "don't count".

Lemma 10.7.6 [Generic Extension] *Let (R, H), as given above, be safe. Let $t \geq 0$ be an arbitrary, but fixed, integer. Then in $G \sim G(n, n^{-\alpha})$ almost surely for all $x_1, \ldots, x_r$ there exist $y_1, \ldots, y_v$ such that*

1. *$y_1, \ldots, y_v$ form an (R, H) extension of $x_1, \ldots, x_r$.*

2. *x_i, y_j are adjacent in G if and only if X_i, Y_j are adjacent in H and y_i, y_j are adjacent in G if and only if Y_i, Y_j are adjacent in H.*

3. *(For $t > 0$) If $z_1, \ldots, z_u$ with $u \leq t$ form a rigid (R', H') extension over $x_1, \ldots, x_r, y_1, \ldots, y_v$ then there are no adjacencies between any z_k and any y_j.*

Example. Let $\alpha \in (\frac{1}{2}, 1)$, $t = 2$, and let (R, H) have root X_1, nonroot Y_1 and edge $\{X_1, Y_1\}$. Note that (R', H') consisting of two roots X_1, X_2 with a common neighbor Y_1 has $v = 1, e = 2$ and is rigid. Generic Extension in this instance says that every x_1 has a neighbor y_1 such that x_1, y_1 do not have a common neighbor z_1.

Proof. From Exercise 5 almost surely every $x_1, \ldots, x_r$ has $\Theta(n^v p^e)$ (R, H) extensions $y_1, \ldots, y_v$. Our rough notion will be that the number of these $y_1, \ldots, y_v$ that fail to be generic, in any of the bounded number of ways that could occur, would be bounded by a smaller power of n.

Call y special if $y \in cl_{t+v}(x_1, \ldots, x_r)$ (as defined below), otherwise nonspecial. Let K, from the Finite Closure Lemma 10.7.7 below, be an almost sure bound on the number of special y, uniform over all choices of the x's. Extend (R, H) to (R^+, H^+) by adding K new roots and no new edges. This is still safe and of the same type as (R, H) so again by Exercise 5 almost surely every $x_1, \ldots, x_r, z_1, \ldots, z_K$ has $\Theta(n^v p^e)$ (R^+, H^+) extensions $y_1, \ldots, y_v$. Letting the z's include all the special vertices we have that almost surely every $x_1, \ldots, x_r$ has $\Theta(n^v p^e)$ (R, H) extensions $y_1, \ldots, y_v$ with all y_i nonspecial. Now we bound from above the number of those nonspecial (R, H) extensions which fail condition 2 or 3.

Consider those extensions (R, H') with an additional edge y_i, y_j or x_i, y_j. This cannot contain a rigid subextension as that would make some y_i special. Hence by Exercise 4 it must be a safe extension. Applying Exercise 5 there are $\Theta(n^v p^{e+1}) = o(n^v p^e)$ such extensions..

Consider extensions by $y_1, \ldots, y_v$ and $z_1, \ldots, z_u$ as in condition 3 with some z_j, y_k adjacent. We can further assume the z's form a minimal rigid extension over the x's and y's. Let the z's have type (v_1, e_1) as an extension over the x's and y's so that $v_1 - e_1\alpha$ is negative. If the y's and z's together formed a safe extension over the x's there would be $\Theta(n^{v+v_1} p^{e+e_1}) = o(n^v p^e)$ such extensions and hence at most that many choices for the y's. Otherwise, by Exercise 4, there would be a rigid subextension. It could not overlap the nonspecial y's. From the minimality it must be precisely all of the z's. Given the x's from the Finite Closure Lemma 10.7.7 there are $O(1)$ choices for the z's. Then the y's form a (v, e') extension over the x's and y's with $e' > e$. This extension has no rigid subextensions (again as the y's are nonspecial) and hence is safe. Again applying Exercise 5 there are $\Theta(n^v p^{e'})$ such y's for each choice of the z's and so $O(n^v p^{e'}) = o(n^v p^e)$ total choices of such y's.

In all cases the number of y's that fail conditions 2 or 3 is $o(n^v p^e)$. Hence there exist y's, indeed most choices of nonspecial y's, which are (R, H) extensions and satisfy conditions 2 and 3. ■

A rigid t-chain in G is a sequence $X = X_0 \subset X_1 \subset \ldots \subset X_K$ with all (X_{i-1}, X_i) rigid and all $|X_{i+1} - X_i| \leq t$. The t-closure of X, denoted by $cl_t(X)$, is the maximal Y for which there exists a rigid t-chain (of arbitrary length) $X = X_0 \subset X_1 \subset \ldots \subset X_K = Y$. When there are no such rigid t-chains we define $cl_t(X) = X$. To see this is well defined we note [using Exercise 4] that if $X = X_0 \subset X_1 \subset \ldots \subset X_K = Z$ and $X = X_0 \subset Y_1 \subset \ldots \subset Y_L = Y$ are rigid t-chains then so is $X = X_0 \subset X_1 \subset \ldots \subset X_K \subset Z \cup Y_1 \subset \ldots \subset Z \cup Y_L = Z \cup Y$. Alternatively, the t-closure $cl_t(X)$ is the minimal set containing X which has no rigid extensions of $\leq t$ vertices. We say $x_1, \ldots, x_r \in G$, $y_1, \ldots, y_r \in H$ have the same t-type if their t-closures are isomorphic as graphs, the isomorphism sending each x_i to the corresponding y_i.

The t-closure is a critical definition, describing the possible special properties of the roots. Suppose, for example, $\alpha \in (\frac{1}{2}, 1)$ and consider $cl_1(x_1, x_2)$. The only rigid extension with $t = 1$ in this range is a nonroot adjacent to two (or more) roots. A sample 1-type would be: x_1, x_2 have common neighbors y_1, y_2 and then x_1, y_1 have common neighbor y_3 and there are no further edges amongst these vertices and no pairs have common neighbors other than those described. A randomly chosen x_1, x_2 would have type: x_1, x_2 have no common neighbors and are not adjacent.

We can already describe the nature of Duplicator's strategy. At the end of the r-th move, with $x_1, \ldots, x_r$ and $y_1, \ldots, y_r$ having been selected from the two graphs, Duplicator will assure that these sets have the same a_r-type. We shall call this the $(a_1, \ldots, a_t)$ *lookahead strategy*. Here a_r must depend only on t, the total number of moves in the game and α. We shall set $a_t = 0$ so that at the end of the game, if Duplicator can stick to the $(a_1, \ldots, a_t)$ lookahead strategy then he has won. If, however, Spoiler picks, say, x_{r+1} so that there is no corresponding y_{r+1} with $x_1, \ldots, x_{r+1}$ and $y_1, \ldots, y_{r+1}$ having the same a_{r+1}-type then the strategy fails and we say that Spoiler wins. The values a_r give the "lookahead" that Duplicator uses but before defining them we need some preliminary results.

Lemma 10.7.7 [Finite Closure] *Let α, $r > 0$ be fixed. Set ε equal to the minimal value of $\frac{e\alpha - v}{v}$ over all integers v, e with $1 \leq v \leq t$ and $e\alpha - v > 0$. Let K be such that $r - K\varepsilon < 0$. Then in $G(n, n^{-\alpha})$ almost surely*

$$|cl_t(X)| \leq K + r$$

for all $X \subset G$ *with* $|X| = r$.

Proof. If not there would be a rigid t-chain $X = X_0 \subset X_1 \subset \ldots \subset X_L = Y$ with $K + r < |Y| < K + r + t$. Letting (X_{i-1}, X_i) have type (v_i, e_i) the restriction of G to Y would have $r + \sum v_i$ vertices and at least $\sum e_i$ edges. But

$$(r + \sum v_i) - \alpha(\sum e_i) = r + \sum(v_i - \alpha e_i) \leq r - \varepsilon \sum v_i < r - K\varepsilon < 0$$

and G almost surely has no such subgraph. ■

Remark. The bound on $|cl_t(X)|$ given by this proof depends strongly on how close α may be approximated by rationals of denominator at most t. This is often the case. If, for example, $\frac{1}{2} + \frac{1}{s-1} > \alpha > \frac{1}{2} + \frac{1}{s}$ then a.s. there will be two points $x_1, x_2 \in G(n, n^{-\alpha})$ having s common neighbors so that $|cl_1(x_1, x_2)| \geq s + 2$.

Now we define the $a_1, \ldots, a_t$ of the lookahead strategy by reverse induction. We set $a_t = 0$. If at the end of the game Duplicator can assure that the 0-types of $x_1, \ldots, x_t$ and $y_1, \ldots, y_t$ are the same then they have the same induced subgraphs and he has won. Suppose, inductively, that $b = a_{r+1}$ has been defined. We define $a = a_r$ to be any integer satisfying

1. $a \geq b$
2. Almost surely $|cl_b(W)| - r \leq a$ for all sets W of size $r + 1$.

Now we need show that almost surely this strategy works. Let $G_1 \sim G(n, n^{-\alpha})$, $G_2 \sim G(m, m^{-\alpha})$ and Duplicator tries to play the $(a_1, \ldots, a_t)$ lookahead strategy on $EHR(G_1, G_2, t)$.

Consider the $(r+1)$-st move. We have $b = a_{r+1}$, $a = a_r$ as above. Points $x_1, \ldots, x_r \in G_1, y_1, \ldots, y_r \in G_2$ have already been selected. Set $X = \{x_1, \ldots, x_r\}$, $Y = \{y_1, \ldots, y_r\}$ for notational convenience. We assume Duplicator has survived thus far so that $cl_a(X) \cong cl_a(Y)$, the isomorphism sending each x_i to the corresponding y_i. Spoiler picks, say, $x = x_{r+1} \in G_1$. Set $X^+ = X \cup \{x\}$ and $Y^+ = Y \cup \{y\}$ where y is Duplicator's as yet undetermined countermove. We distinguish two cases.

We say Spoiler has moved Inside if $x \in cl_a(X)$. Then as $b \leq a$, $cl_b(X^+) \subseteq cl_a(X)$. Duplicator looks at the isomorphism $\Psi : cl_a(X) \to cl_a(Y)$ and selects $y = \Psi(x)$.

We say Spoiler has moved Outside if $x \notin cl_a(X)$. Let NEW be those vertices of $cl_b(X^+)$ that do not lie in $cl_a(X)$. $NEW \neq \emptyset$ as $x \in NEW$. $|NEW| \leq a$ as $NEW \subseteq cl_b(X^+) - X$. Consider NEW as an (R, H) extension of $cl_a(X)$. This extension must be safe as otherwise it would have a rigid subextension NEW^- but that subextension would then be in $cl_a(X)$. Duplicator now goes to G_2 and, applying the Generic Extension Lemma 10.7.6 with $t = b$, finds an (R, H) extension of $cl_a(Y)$. That is, he finds an edge preserving injection $\Psi : cl_a(X) \cup NEW \to H$ extending the isomorphism between $cl_a(X)$ and $cl_a(Y)$. Duplicator selects $y = \Psi(x)$.

Why does this work? Set $NEW' = \Psi(NEW)$ and $CORE = \Psi(cl_b(X^+))$. We can reach $cl_b(X^+)$ by a rigid b-chain from X^+ and the isomorphism gives the same chain from Y^+ to $CORE$ so that $cl_b(Y^+)$ contains $CORE$. But can it have additional vertices? We use the genericity to say no. Suppose there was a rigid extension $MORE$ over $CORE$ with at most b nonroots. We can't have $MORE$ entirely inside $\Psi[cl_a(X) \cup NEW]$ as then $\Psi^{-1}[MORE]$ would be in $cl_b(X^+)$ as well. Let $MORE^+$ be the vertices of $MORE$ lying outside $\Psi[cl_a(X) \cup NEW]$. $MORE^+$ is then a rigid extension of $\Psi[cl_a(X) \cup NEW]$. By the genericity $MORE^+$ would have no adjacencies to NEW' and so would be a rigid extension of $\Psi[cl_a(X)] = cl_a(Y)$. As $a \geq b$ the a-closure of a set cannot have rigid extensions with $\leq b$ vertices. Hence there is no $MORE$.

The first move follows the same pattern but is somewhat simpler. Set $b = a_1$ and let a satisfy $a \geq b$ and $a \geq |cl_b(x)|$ for any x. Spoiler plays $x \in G_1$. [Effectively, there is no Inside move as $X = \emptyset$ is the set of previous moves and $cl_a(\emptyset) = \emptyset$.] Duplicator calculates the graph $H = cl_b(x)$ which has, say, v vertices [including x] and e edges. Since H is a subgraph of G_1 the threshold function for the appearance of H must come before $n^{-\alpha}$. In particular, for every subgraph H' of H with v' vertices and e' edges we cannot have $v' - \alpha e' < 0$ and therefore must have $v' - \alpha e' > 0$. The conditions of Theorem 4.4.5 then apply and G_2 almost surely has $\Theta(m^{e-v\alpha})$ copies of H. Consider any graph H^+ consisting of H together with a rigid extension of H with at most b vertices. Such H^+ would have $v + v^+$ vertices and $e + e^+$ edges with $v^+ - \alpha e^+ < 0$. The expected number of copies of H^+ is then $\Theta(m^{e-v\alpha+(v^+-\alpha e^+)})$ which is $o(m^{e-v\alpha})$. Hence there will be in G_2 a copy of H which is not part of any such H^+. [Effectively, this is generic extension over the empty set.] Duplicator finds the edge preserving injection $\Psi : cl_b(x) \to G_2$ giving such a copy of H and selects $y = \Psi(x)$.

We have shown that the $(a_1, \ldots, a_t)$ lookahead strategy almost surely results in a win for Duplicator. By Theorem 10.7.4 this implies the Zero-One Law Theorem 10.7.2.

10.8 EXERCISES

1. Show that there is a graph on n vertices with minimum degree at least $n/2$ in which the size of every dominating set is at least $\Omega(\log n)$.

2. Find a threshold function for the property: $G(n, p)$ contains a copy of the graph consisting of a complete graph on 4 vertices plus an extra vertex joined to one of its vertices.

3. Let X be the number of cycles in the random graph $G(n, p)$ with $p = \frac{c}{n}$. Give an exact formula for $E[X]$. Find the asymptotics of $E[X]$ when $c < 1$. Find the asymptotics of $E[X]$ when $c = 1$.

4. Here we write (R, S) for $(R, H|_S)$ where H is some fixed graph.

   - Let $R \subset S \subset T$. Show that if $(R, S), (S, T)$ are both dense then so is (R, T). Show that if $(R, S), (S, T)$ are both sparse then so is (R, T)
   - Let $R \subset S$. Show that if (R, S) is rigid then $(X \cup R, X \cup S)$ is rigid for any X.
   - $R \subset U$ with (R, U) *not* sparse. Show there is a T with $R \subset T \subset U$ with (R, T) dense. Show further there is an S with $R \subset S \subset T$ with (R, S) rigid.
   - Show that *any* (R, T) is either rigid or sparse itself or there exists S with $R \subset S \subset T$ such that (R, S) is rigid and (S, T) is sparse.

5. We call (R, H) *hinged* if it is safe but there is no S with $R \subset S \subset V(H)$ such that (S, H) is safe. For $x_1, \ldots, x_r \in G$ let $N(x_1, \ldots, x_r)$ denote the number of (R, H) extensions. Set $\mu = E[N] \sim n^v p^e$.

   - Let (R, H) be hinged and fix $x_1, \ldots, x_r \in G$. Following the model of §8.8.5, especially Theorem 8.5.4, show that

   $$\Pr[|N(x_1, \ldots, x_r) - \mu| > \epsilon\mu] = o(n^{-r}).$$

   - Deduce that almost surely all $N(x_1, \ldots, x_r) \sim \mu$.
   - Show that $N(x_1, \ldots, x_r) \sim \mu$ holds for any safe (R, H), by decomposing (R, H) into hinged extensions.

THE PROBABILISTIC LENS: *Counting Subgraphs*

A graph $G = (V, E)$ on n vertices has 2^n induced subgraphs but some will surely be isomorphic. How many different subgraphs can G have? Here we show that there are graphs G with $2^n(1 - o(1))$ different subgraphs. The argument we give is fairly coarse. It is typical of those situations where a probabilistic approach gives fairly quick answers to questions otherwise difficult to approach.

Let G be a random graph on n vertices with edge probability $1/2$. Let $S \subseteq V$, $|S| = t$ be fixed. For any one to one $\rho : S \to V$, $\rho \neq id$, let A_ρ be the event that ρ gives a graph isomorphism - i.e., for $x, y \in S$, $\{x, y\} \in E \Leftrightarrow \{\rho x, \rho y\} \in E$. Set $M_\rho = \{x \in S : \rho x \neq x\}$. We split the set of ρ by $g = g(\rho) = |M_\rho|$.

Consider the $g(t - g) + \binom{g}{2}$ pairs x, y with $x, y \in S$ and at least one of x, y in M. For all but at most $g/2$ of these pairs $\{x, y\} \neq \{\rho x, \rho y\}$. (The exceptions are when $\rho x = y, \rho y = x$.) Let E_ρ be the set of pairs $\{x, y\}$ with $\{x, y\} \neq \{\rho x, \rho y\}$. Define a graph H_ρ with vertices E_ρ and vertex $\{x, y\}$ adjacent to $\{\rho x, \rho y\}$. In H_ρ each vertex has degree at most two ($\{x, y\}$ may also be adjacent to $\{\rho^{-1}x, \rho^{-1}y\}$) and so it decomposes into isolated vertices, paths and circuits. On each such component there is an independent set of size at least one-third the number of elements, the extreme case being a triangle. Thus there is a set $I_\rho \subseteq E_\rho$ with

$$|I_\rho| \geq |E_\rho| \geq \frac{g(t - g) + \binom{g}{2} - g/2}{3},$$

so that the pairs $\{x, y\}$, $\{\rho x, \rho y\}$ with $\{x, y\} \in I_\rho$ are all distinct.

For each $\{x, y\} \in I_\rho$ the event $\{x, y\} \in E \Leftrightarrow \{\rho x, \rho y\} \in E$ has probability $1/2$. Moreover these events are mutually independent over $\{x, y\} \in I_\rho$ since they involve distinct pairs. Thus we bound

$$\Pr[A_\rho] \leq 2^{-|I_\rho|} \leq 2^{-(g(t-g)+\binom{g}{2}-g/2)/3}.$$

For a given g the function ρ is determined by $\{x : \rho x \neq x\}$ and the values ρx for those x so that there are less than n^{2g} such ρ. We bound

$$\sum_{\rho \neq id} \Pr[A_\rho] = \sum_{g=1}^{t} \sum_{g(\rho)=g} \Pr[A_\rho] \leq \sum_{g=1}^{t} n^{2g} 2^{-(g(t-g)+\binom{g}{2}-g/2)/3}.$$

We make the rough bound

$$g(t-g) + \binom{g}{2} - g/2 = g(t - \frac{g}{2} - 1) \geq g(\frac{t}{2} - 1),$$

since $g \leq t$. Then

$$\sum_{\rho \neq id} \Pr[A_\rho] \leq \sum_{g=1}^{t} \left[n^2 2^{(-\frac{t}{2}+1)/3} \right]^g .$$

For, again being rough, $t > 50 \ln n$, $2^{\frac{1}{3}-\frac{t}{6}} < n^{-3}$ and $\sum_{\rho \neq id} \Pr[A_\rho] = o(1)$. That is, almost surely there is no isomorphic copy of $G|_S$.

For all $S \subseteq V$ with $|S| > 50 \ln n$ let I_S be the indicator random variable for there being no other subgraph isomorphic to $G|_S$. Set $X = \sum I_S$. Then $E[I_S] = 1 - o(1)$ so, by linearity of expectation - there being $2^n(1 - o(1))$ such S -

$$E[X] = 2^n(1 - o(1)).$$

Hence there is a specific G with $X > 2^n(1 - o(1))$.

11

Circuit Complexity

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again; the never-satisfied man is so strange - if he has completed a structure then it is not in order to dwell in it peacefully, but in order to begin another. I imagine the world conqueror must feel thus, who, after one kingdom is scarely conquered, stretches out his arms for another.
– Karl Friedrich Gauss

11.1 PRELIMINARIES

A *Boolean function* $f = f(x_1, \ldots, x_n)$ on the n variables $x_1, x_2, \ldots, x_n$ is simply a function $f : \{0,1\}^n \rightarrow \{0,1\}$. In particular $0, 1, x_1 \wedge \cdots \wedge x_n,\ x_1 \vee \cdots \vee x_n,\ x_1 \oplus \cdots \oplus x_n$ denote, as usual, the two constant functions, the *And* function (whose value is 1 iff $x_i = 1$ for all i), the *Or* function (whose value is 0 iff $x_i = 0$ for all i, and the *Parity* function (whose value is 0 iff an even number of variables x_i is 1), respectively. For a function f, we let $\overline{f} = f \oplus 1$ denote its complement *Not* f. The functions x_i and $\overline{x}_i$ are called *atoms*. In this section we consider the problem of computing various Boolean functions efficiently. A *circuit* is a directed, acyclic graph, with a special vertex with no outgoing edges called the Output vertex. Every vertex is labelled by a Boolean function of its immediate parents, and the vertices with no parents (i.e., those with no ingoing edges) are labelled either by one of the variables x_i or by a constant 0 or 1. For every assignment of binary values to each variable x_i one can compute, recursively, the corresponding value of each vertex of the circuit

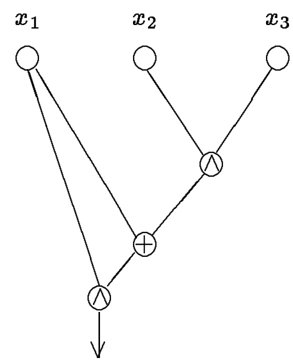


Fig. 11.1

by applying the corresponding function labelling it to the already computed values of its parents. We say that the circuit *computes* the function $f = f(x_1, \ldots, x_n)$ if for each $x_i \in \{0, 1\}$, the corresponding value of the output vertex of the circuit equals $f(x_1, \ldots, x_n)$. For example Figure 11.1 below presents a circuit computing $f(x_1, x_2, x_3) = (x_1 \oplus (x_2 \wedge x_3)) \wedge x_1$.

If every fanout in a circuit is at most one (i.e., the corresponding graph is a tree) the circuit is called a *formula*. If every fanin in a circuit is at most two the circuit is called a *binary circuit* . Therefore the circuit in Figure 11.1 is binary, but it is not a formula. The *size* of a circuit is the number of vertices in it, and its *depth* is the maximum length of a directed path in it. The *binary circuit complexity* of a Boolean function, is the size of the smallest binary circuit computing it. An easy counting argument shows that for large n the binary circuit complexity of almost all the functions of n variables is at least $(1 + o(1))2^n/n$. This is because the number of binary circuits of size s can be easily shown to be less than $(c_1 s)^s$, whereas the total number of Boolean functions on n variables is 2^{2^n}. On the other hand, there is no known non-linear, not to mention exponential (in n), lower bound for the binary circuit complexity of any "explicit" function. By "explicit" here we mean an NP-function, that is, one of a family $\{f_{n_i}\}_{i \geq 1}$ of Boolean functions, where f_{n_i} has n_i variables, $n_i \to \infty$, and there is a non-deterministic Turing machine which, given n_i and $x_1, \ldots, x_{n_i}$ can decide in (non deterministic) polynomial time (in n_i) if $f_{n_i}(x_1, \ldots, x_{n_i}) = 1$. (An example for such a family is the $\frac{n}{2}$ - clique function; here $n_i = \binom{i}{2}$, the n_i variables $x_1, \ldots, x_{n_i}$ represent the edges of a graph on i vertices and $f_{n_i}(x_1, \ldots, x_{n_i}) = 1$ iff the corresponding graph contains a clique on at least $i/2$ vertices). Any non-polynomial lower bound for the binary circuit complexityof an explicit function would imply (among other things) that $P \neq NP$ and thus solve the arguably most important open problem in Theoretical Computer Science. Unfortunately, at the moment, the best known lower bound for the binary circuit complexity of an explicit function of n variables is only $3n$, (see Blum (1984) , Paul (1977)). However, several non-trivial lower bounds are known when we impose certain restrictions on the structure of the circuits. Most of the known proofs of these bounds rely heavily on probabilistic methods. In this chapter we describe some of these results. We note that there are many additional beautiful known results about circuit complexity , see, e.g. Wegener (1987) and Karchmer and Wigderson (1990) , but those included here are not only among the crucial ones, but also represent the elegant methods used in this field. Since most results in this chapter are asymptotic we assume, throughout the chapter, whenever it is needed, that the number of variables we have is sufficiently large.

11.2 RANDOM RESTRICTIONS AND BOUNDED DEPTH CIRCUITS

Let us call a Boolean function G a t-And-Or if it can be written as an And of an arbitrary number of functions, each being an Or of at most t atoms, i.e., $G = G_1 \wedge \cdots \wedge G_w$, where $G_i = y_{i1} \vee \cdots \vee y_{ia_i}$, $a_i \leq t$ and each y_j is an atom. Similarly, we call a Boolean function an s-Or-And, if it can be written as an Or of And gates each containing at most s atoms. A *minterm* of a function is a minimal assignment of values to some of the variables that forces the function to be 1. Its *size* is the number of variables whose values are set. Notice that a function is an s-Or-And if and only if each of its minterms is of size at most s. A *restriction* is a map ρ of the set of indices $\{1, \ldots, n\}$ to the set $\{0, 1, *\}$. The restriction of the function $G = G(x_1, \ldots, x_n)$ by ρ, denoted by $G|\rho$, is the Boolean function obtained from G by setting the value of each x_i for $i \in \rho^{-1}\{0, 1\}$ to $\rho(i)$, and leaving each x_j for $j \in \rho^{-1}(*)$ as a variable. Thus, for example, if $G(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee x_3$ and $\rho(1) = 0$ $\rho(2) = \rho(3) = *$ then $G|_\rho = x_3$. For $0 \leq p \leq 1$, a *random p-restriction* is a random restriction ρ defined by choosing, for each $1 \leq i \leq n$ independently the value of $\rho(i)$ according to the following distribution:

$$Pr[\rho(i) = *] = p\,, \quad Pr[\rho(i) = 0] = Pr[\rho(i) = 1] = (1-p)/2\,. \tag{11.1}$$

Improving results of Furst, Saxe and Sipser (1984) , Ajtai (1983) and Yao (1985) , J. Håstad (1988) proved the following result, which is very useful in establishing lower bounds for bounded-depth circuits.

Lemma 11.2.1 **[***(The switching lemma)***]**
Let $G = G(x_1, \ldots, x_n)$ be a t-And-Or, i.e., $G = G_1 \wedge G_2 \wedge \cdots \wedge G_w$, where each G_i is an Or of at most t atoms. Let ρ be the random restriction defined by (11.1). Then

$$\begin{aligned} & Pr[G|\rho \text{ is not an } (s-1)\text{-Or-And}] \\ = \quad & Pr[G|\rho \text{ has a minterm of size } \geq s] \leq (5pt)^s\,. \end{aligned}$$

Proof. Let E_s be the event that $G|_\rho$ has a minterm of size at least s. To bound $Pr(E_s)$, we prove a stronger result; for any Boolean function F,

$$Pr[E_s | F|_\rho \equiv 1] \leq (5\,pt)^s\,. \tag{11.2}$$

Here we agree that if the condition is unsatisfied then the conditional probability is 0. Lemma 11.2.1 is obtained from (11.2) by taking $F \equiv 1$. We prove (11.2) by induction on w. For $w = 0$, $G \equiv 1$ and there is nothing to prove. Assuming (11.2) holds whenever the number of G_i is less than w, we prove it for w. Put $G = G_1 \wedge G^*$, where $G^* = G_2 \wedge \cdots \wedge G_w$, and let E_s^* be the event that $G^*|_\rho$ has a minterm of size at least s. By interchanging, if necessary, some of the variables with their complements we may assume, for convenience, that $G_1 = \bigvee_{i \in T} x_i$, where $|T| \leq t$. Either $G_1|_\rho \equiv 1$ or $G_1|_\rho \not\equiv 1$. In the former case, E_s holds if and only if E_s^* holds and hence, by induction

$$Pr[E_s | F|_\rho \equiv 1,\ G_1|_\rho \equiv 1] = Pr[E_s^* | (F \wedge G_1)|_\rho \equiv 1] \leq (5\,pt)^s\ . \tag{11.3}$$

The case $G_1|_\rho \not\equiv 1$ requires more work. In this case, any minterm of $G|_\rho$ must assign a value 1 to at least one x_i, for $i \in T$. For a nonempty $Y \subseteq T$ and for a function $\sigma : Y \to \{0, 1\}$ which is not identically 0, let $E_s(Y, \sigma)$ be the event that $G|_\rho$ has a minterm of size at least s which assigns the value $\sigma(i)$ to x_i for each $i \in Y$ and does not assign any additional values to variables x_j with $j \in T$. By the preceding remark

$$Pr[E_s | F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1] \leq \sum_{Y, \sigma} Pr[E_s(Y, \sigma) | F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1]\ . \tag{11.4}$$

Observe that the condition $G_1|_\rho \not\equiv 1$ means precisely that $\rho(i) \in \{0, *\}$ for all $i \in T$ and hence, for each $i \in T$:

$$Pr[\rho(i) = * | G_1|_\rho \not\equiv 1] = \frac{p}{p + (1 - p)/2} = 2p/(1 + p)\ .$$

Thus, if $|Y| = y$

$$Pr[\rho(Y) = * | G_1|_\rho \not\equiv 1] \leq (\frac{2p}{1 + p})^y\ .$$

The further condition $F|_\rho \equiv 1$ can only decrease this probability. This can be shown using the FKG inequality (see Chapter 6). It can also be shown directly as follows. For any fixed $\rho' : N - Y \to \{0, 1, *\}$, where $N = \{1, \ldots, n\}$, we claim that

$$Pr[\rho(Y) = * | F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1,\ \rho|_{N-Y} = \rho'] \leq (\frac{2p}{1 + p})^y\ .$$

Indeed, the given ρ' has a unique extension ρ with $\rho(Y) = *$. If that ρ does not satisfy the above conditions, then the conditional probability is zero. If it does, then so do all extensions ρ with $\rho(i) \in \{0, *\}$ for $i \in Y$, and so the inequality holds in this case too. As this holds for all fixed ρ' we conclude that indeed

$$Pr[\rho(Y) = * | F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1] \leq (\frac{2p}{1 + p})^y \leq (2p)^y\ . \tag{11.5}$$

Let $\rho' : T \to \{0, *\}$ satisfy $\rho(Y) = *$ and consider all possible restrictions ρ satisfying $\rho|_T = \rho'$. Under this condition, ρ may be considered as a random restriction on $N-T$. The event $F|_\rho \equiv 1$ reduces to the event $\tilde{F}|_{\rho|_{N-T}} \equiv 1$, where $\tilde{F}$ is the And of all functions obtained from F by substituting the values of x_i according to ρ' for those $i \in T$ with $\rho'(i) = 0$, and by taking all possibilities for all the other variables x_j for $j \in T$. If the event $E_s(Y, \sigma)$ occurs then $G^*|_{\rho\sigma}$ has a minterm of size at least $s-y$ that does not contain any variable x_i with $i \in T - Y$. But this happens if and only if $\tilde{G}|_{\rho|_{N-T}}$ has a minterm of size at least $s-y$, where $\tilde{G}$ is the function obtained from G^* by substituting the values of x_j for $j \in Y$ according to σ, the values of x_i for $i \in T - Y$ and $\rho'(i) = 0$ according to ρ' and by removing all the variables x_k with $k \in T - Y$ and $\rho'(k) = *$. Denoting this event by $\tilde{E}_{s-y}$ we can apply induction and obtain

$$Pr[E_s(Y,\sigma)|F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1,\ \rho|_T = \rho'] \le Pr[\tilde{E}_{s-y}|\tilde{F}|_\rho \equiv 1] \le (5pt)^{s-y} .$$

Since any ρ with $F|_\rho \equiv 1,\ G_1|_\rho \equiv 1,\ \rho(Y) = *$ must have $\rho|_T = \rho'$ for some ρ' of this form, and since the event $E_s(Y, \sigma)$ may occur only if $\rho(Y) = *$ we conclude that

$$Pr[E_s(Y,\sigma)|F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1,\ \rho(Y) = *] \le (5pt)^{s-y} ,$$

and, by (11.5),

$$\begin{aligned} & Pr[E_s(Y,\sigma)|F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1,] \\ = \quad & Pr[\rho(Y) = *|F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1]\cdot \\ & \cdot Pr[E_s(Y,\sigma)|F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1,\ \rho(Y) = *] \\ & \le (2p)^y(5pt)^{s-y} . \end{aligned}$$

Substituting in (11.4) and using the fact that $|T| \le t$ and that

$$\sum_{y=1}^{t}(2^y-1)2^y/(5^y y!) \le \frac{2}{5} + \sum_{y=2}^{\infty}\frac{(4/5)^y}{y!} = \frac{2}{5} + e^{4/5} - 1 - \frac{4}{5} < 1$$

we obtain,

$$\begin{aligned} & Pr[E_s|F|_\rho \equiv 1,\ G_1|_\rho \not\equiv 1] \\ & \le \textstyle\sum_{y=1}^{|T|}\binom{|T|}{y}(2^y-1)(2p)^y(5pt)^{s-y} \le (5pt)^s \sum_{y=1}^{t}\frac{t^y}{y!}(2^y-1)(\frac{2}{5t})^y \\ & = (5pt)^s \textstyle\sum_{y=1}^{t}(2^y-1)\cdot\frac{2^y}{5^y\cdot y!} \le (5pt)^s . \end{aligned}$$

This, together with (11.3) gives

$$Pr[E_s|F|_\rho \equiv 1] \le (5pt)^s$$

completing the induction and the proof. ■

By taking the complement of the function G in Lemma 11.2.1 and applying De Morgan's rules one clearly obtains its dual form; If G is a t-Or-And and ρ is the random restriction given by (11.1) then $Pr[G|_\rho$ is not an $(s-1)$-And-Or$] \le (5pt)^s$.

We now describe an application of the switching lemma that supplies a lower bound to the size of circuits of small depth that compute the parity function $x_1 \oplus \cdots \oplus x_n$. We consider circuits in which the vertices are arranged in levels, those in the first level are atoms (i.e., variables or their complements), and each other gate is either an Or or an And of an arbitrary number of vertices from the previous level. We assume that the gates in each level are either all And gates or all Or gates, and that the levels alternate between And levels and Or levels. A circuit of this form is called a $C(s, s', d, t)$-circuit if it contains at most s gates, at most s' of which are above the second level, its depth is at most d and the fanin of each gate in its second level is at most t. Thus, for example, the circuit that computes the parity function by computing an Or of the 2^{n-1} terms $x_1^{\varepsilon_1} \wedge \cdots \wedge x_n^{\varepsilon_n}$, where $(\varepsilon_1, \ldots, \varepsilon_n)$ ranges over all even binary vectors and $x_i^{\varepsilon_i} = x_i \oplus \varepsilon_i$, is a $C(2^{n-1}+1, 1, 2, n)$-circuit.

Theorem 11.2.2 *Let $f = f(x_1, \ldots, x_n)$ be a function and let C be a $C(\infty, s, d, t)$-circuit computing f, where $s \cdot (\frac{1}{2})^t \leq 0.5$. Then either f or its complement $\overline{f}$ has a minterm of size at most $n - \frac{n}{2\cdot(10t)^{d-2}} + t$.*

Proof. Let us apply to C, repeatedly, $d-2$ times a random $1/(10t)$- restriction. Each of these random restrictions, when applied to any bottom subcircuit of depth 2, transforms it by Lemma 11.2.1 with probability at least $1-(\frac{1}{2})^t$ from a t-Or-And to a t-And-Or (or conversely). If all these transformations succeed we can merge the new And gates with these from the level above them and obtain a circuit with a smaller depth. As the total size of the circuit is at most s and $s(\frac{1}{2})^t \leq 0.5$, we conclude that with probability at least half all transformations succeed and C is transformed into a $C(\infty, 1, 2, t)$-circuit. Each variable x_i, independently, is still a variable (i.e., has not been assigned a value) with probability $\frac{1}{(10t)^{d-2}}$. Thus, the number of remaining variables is a binomial random variable with expectation $\frac{n}{(10t)^{d-2}}$ and a little smaller variance . By the standard estimates for binomial distributions (see Appendix A) the probability that at least $\frac{n}{2\cdot(10t)^{d-2}}$ variables are still variables is more than a half. Therefore, with positive probability, at most $n - \frac{n}{2\cdot(10t)^{d-2}}$ of the variables have been fixed and the resulting restriction of f has a $C(\infty, 1, 2, t)$-circuit, i.e., its value can be fixed by assigning values to at most t additional variables. This completes the proof. ■

Corollary 11.2.3 *For any $d \geq 2$, there is no*
$C\left(\infty, \frac{1}{2} \cdot 2^{\frac{1}{10}n^{1/(d-1)}}, d, \frac{1}{10}n^{1/(d-1)}\right)$ *– circuit that computes the parity function* $f(x_1, \ldots, x_n) = x_1 \oplus \cdots \oplus x_n$.

Proof. Assuming there is such a circuit we obtain, by Theorem 11.2.2 , that the value of f can be fixed by assigning values to at most $n - \frac{1}{2}n^{1/(d-1)} + \frac{1}{10}n^{1/(d-1)} < n$ variables. This is false, and hence there is no such circuit.

■

The estimate in Corollary 11.2.3 is, in fact, nearly best possible. Since every $C(s, s', d, t)$ - circuit can be transformed into a $C(ts, s, d+1, 2)$ - circuit (by replacing each atom by an Or or And of two copies of itself), Corollary 11.2.3 easily implies that the depth d of any $C(s, s', d, t)$ - circuit of polynomial size that computes the parity of n bits is at least $\Omega(\log n/\log\log n)$. This lower bound is also optimal.

11.3 MORE ON BOUNDED-DEPTH CIRCUITS

In the previous section we saw that the parity function is hard to compute in small depth using And, Or and Not gates. It turns out that even if we allow the use of parity gates (in addition to the And, Or and Not gates) there are still some relatively simple functions that are hard to compute. Such a result was first proved by Razborov (1987) . His method was modified and strengthened by Smolensky (1987) . For an integer $k \geq 2$, let $Mod_k(x_1, x_2, \ldots, x_n)$ be the Boolean function whose value is 1 iff $\Sigma x_i \not\equiv 0(mod\ k)$. Smolensky showed that for every two powers p and q of distinct primes, the function Mod_p cannot be computed in a bounded-depth polynomial-size circuit which uses And, Or, Not and Mod_q gates. Here we present the special case of this result in which $q = 3$ and $p = 2$.

Let C be an arbitrary circuit of depth d and size s consisting of And, Or, Not and Mod_3 gates. A crucial fact, due to Razborov, is the assertion that the output of C can be approximated quite well (depending on d and s) by a polynomial of relatively small degree over $GF(3)$. This is proved by applying the probabilistic method as follows. Let us replace each gate of the circuit C by an approximate polynomial operation, according to the following rules which guarantee that in each vertex in the new circuit we compute a polynomial over $GF(3)$, whose values are all 0 or 1 (whenever the input is a 0-1 input).

(i) Each Not-gate $\overline{y}$ is replaced by the polynomial gate $(1 - y)$.

(ii) Each Mod_3 gate $Mod_3(y_1, \ldots, y_m)$ is replaced by the polynomial gate $(y_1 + y_2 + \cdots + y_m)^2$.

The rule for replacement of Or and And gates is a little more complicated. Observe that in the two previous cases (i) and (ii) there was no approximation; the new gates compute precisely what the old ones did, for all possible Boolean values of the variables. This can, in principal, be done here too. An And gate $y_1 \wedge \cdots \wedge y_m$ should simply be replaced by the product $y_1 \cdots y_m$. An Or gate $y_i \vee \cdots \vee y_m$ can then be computed by de-Morgan's rules. Since $y_1 \vee \cdots \vee y_m = \overline{(\overline{y}_1 \wedge \cdots \wedge \overline{y}_m)}$ and $\overline{y}$ is realized by $(1 - y)$, this would give

$$1 - (1 - y_1)(1 - y_2) \cdots (1 - y_m) \; . \tag{11.6}$$

The trouble is that this procedure would increase the degree of our polynomials too much. Hence, we need to be a little more tricky. Let ℓ be an integer, to be chosen later. Given a Or gate $y_1 \vee \cdots \vee y_m$, we choose ℓ random subsets $I_1, \ldots, I_\ell$ of $\{1, \ldots, m\}$, where for each $1 \leq i \leq \ell$ and for each $1 \leq j \leq m$ independently $Pr(j \in I_i) = 1/2$. Observe that for each fixed i, $1 \leq i \leq \ell$, the sum $\left(\sum_{j \in I_i} y_j\right)^2$ over $GF(3)$ is certainly 0 if $y_1 \vee \cdots \vee y_m = 0$, and is 1 with probability at least a half if $y_1 \vee \cdots \vee y_m = 1$. Hence, if we compute the Or function of the ℓ expressions $\left(\sum_{j\in I_i} y_j\right)^2$ $(1 \leq i \leq \ell)$, this function is 0 if $y_1 \vee \cdots \vee y_m = 0$ and is 1 with probability at least $1 - (1/2)^\ell$ if $y_1 \vee \cdots \vee y_m = 1$. We thus compute the Or and write it as a polynomial, in the way explained in equation (11.6). This gives

$$1 - \prod_{i=1}^{\ell} \left(1 - (\sum_{j\in I_i} y_j)^2 \right) . \tag{11.7}$$

Therefore, in our new circuit we replace each Or gate by an approximation polynomial gate of the form described in (11.7). Once we have an approximation to an Or - gate we can obtain the corresponding one for an And - gate by applying De Morgan rules. Since $y_1 \wedge \cdots \wedge y_m = \overline{(\overline{y}_1 \vee \cdots \vee \overline{y}_m)}$ we replace each And gate of the form $y_1 \wedge \cdots \wedge y_m$ by

$$\prod_{i=1}^{\ell} \left(1 - [\sum_{j\in I_i} (1 - y_j)]^2 \right) . \tag{11.8}$$

Observe that the polynomials in (11.7) and (11.8) are both of degree at most 2ℓ.

Given the original circuit C of depth d and size s, we can now replace all its gates by our approximative polynomial gates and get a new circuit CP, which depends on all the random choices made in each replacement of each of the And/Or gates. The new circuit CP computes a polynomial $P(x_1, \ldots, x_n)$ of degree at most $(2\ell)^d$. Moreover, for each fixed Boolean values of $x_1, x_2, \ldots, x_n$, the probability that all the new gates compute exactly what the corresponding gates in C computed is at least $1 - s/2^\ell$. Therefore, the expected number of inputs on which $P(x_1, \ldots, x_n)$ is equal to the output of C is at least $2^n(1 - s/2^\ell)$. We have thus proved the following.

Lemma 11.3.1 *For any circuit C of depth d and size s on n Boolean variables that uses Not, Or, And and Mod_3 - gates and for any integer ℓ, there is a polynomial $P = P(x_1, \ldots, x_n)$ of degree at most $(2\ell)^d$ over $GF(3)$ whose value is equal to the output of C on at least $2^n(1 - s/2^\ell)$ inputs.*

In order to apply this lemma for obtaining lower bounds for the size of any circuit of the above type that computes the parity function we need the following additional combinatorial result.

Lemma 11.3.2 *There is no polynomial $P(x_1, \ldots, x_n)$ over $GF(3)$ of degree at most $\sqrt{n}$ which is equal to the parity of $x_1, \ldots, x_n$ for a set S of at least $0.9 \cdot 2^n$ distinct binary vectors $(x_1, \ldots, x_n)$.*

Proof. Suppose this is false, and suppose $S \subset \{0,1\}^n$, $|S| \geq 0.9 \cdot 2^n$ and $P(x_1, \ldots, x_n) = x_1 \oplus \cdots \oplus x_n$ for all $(x_1, \ldots, x_n) \in S$. Define a polynomial $Q = Q(y_1, \ldots, y_n)$ by $Q = Q(y_1, \ldots, y_n) = P(y_1 + 2, \ldots, y_n + 2) - 2$, and

$T = \{(y_1, \ldots, y_n) \in \{1, -1\}^n : (y_1 + 2, \ldots, y_n + 2) \in S\}$, where all additions are in $GF(3)$. Clearly Q has degree at most $\sqrt{n}$ and $Q(y_1, \ldots, y_n) = \prod_{i=1}^{n} y_i$ for all $(y_1, \ldots, y_n) \in T$. Let now $G = G(y_1, \ldots, y_n) : T \to GF(3)$ be an arbitrary function. Extend it in an arbitrary way to a function from $(GF(3))^n \to GF(3)$, and write this function as a polynomial in n variables. (Trivially, any function from $(GF(3))^n \to GF(3)$ is a polynomial. This follows from the fact that it is a linear combination of functions of the form $\prod_{i=1}^{n}(y_i - \varepsilon_i)(y_i - \varepsilon_i - 1)$, where $\varepsilon_i \in GF(3)$). Replace each occurance of y_i^2 in this polynomial by 1 to obtain a multilinear polynomial $\tilde{G}$ which agrees with G on T. Now replace each monomial $\prod_{i \in U} y_i$, where $|U| > \frac{n}{2} + \frac{\sqrt{n}}{2}$ by $\prod_{i \notin U} y_i \cdot Q(y_1, \ldots, y_n)$, and replace this new polynomial by a multilinear one, $\approx{G}$, again by replacing each y_i^2 by 1. Since for $y_i \in \{\pm 1\}$, $\quad \prod_{i \notin U} y_i \cdot \prod_{i=1}^{n} y_i = \prod_{i \in U} y_i$, $\quad \overset{\approx}{G}$ is equal to G on T and its degree is at most $\frac{n}{2} + \frac{\sqrt{n}}{2}$. However, the number of possible $\overset{\approx}{G}$ is $3^{\sum_{i=0}^{\frac{n}{2}+\frac{\sqrt{n}}{2}} \binom{n}{i}} < 3^{0.88 \cdot 2^n}$, whereas the number of possible G is $3^{|T|} \geq 3^{0.9 \cdot 2^n}$. This is impossible, and hence the assertion of the lemma holds. ■

Corollary 11.3.3 *There is no circuit of depth d and size $s \leq \frac{1}{10} 2^{\frac{1}{2} n^{1/2d}}$ computing the parity of $x_1, x_2, \ldots, x_n$ using Not, And, Or and Mod_3 gates.*

Proof. Suppose this is false and let C be such a circuit. Put $\ell = \frac{1}{2} \cdot n^{1/2d}$. By Lemma 11.3.1 there is a polynomial $P = P(x_1, \ldots, x_n)$ over $GF(3)$, whose degree is at most $(2\ell)^d = \sqrt{n}$, which is equal to the parity of $x_1, \ldots, x_n$ on at least $2^n \left(1 - \frac{s}{2^{\frac{1}{2} n^{1/2d}}}\right) \geq 0.9 \cdot 2^n$ inputs. This contradicts lemma 11.3.2, and hence completes the proof.

■

11.4 MONOTONE CIRCUITS

A Boolean function $f = f(x_1, \ldots, x_n)$ is *monotone* if $f(x_1, \ldots, x_n) = 1$ and $x_i \leq y_i$ imply $f(y_1, \ldots, y_n) = 1$. A *binary monotone circuit* is a binary circuit that contains only binary And and Or gates. It is easy to see that a function is monotone if and only if there is a binary montone circuit that computes it. The *monotone complexity* of a monotone function is the smallest size of a binary monotone circuit that computes it. Until 1985, the largest known lower bound for the monotone complexity of a monotone NP function of n variables was $4n$. This was considerably improved in the fundamental paper of Razborov (1985) , where a bound of $n^{\Omega(\log n)}$ to the Clique$_k$-function (which is 1 iff a given graph contains a clique of size k) is established. Shortly afterwards, Andreev (1985) used similar methods to obtain an exponential lower bound to a somewhat unnatural NP-function. Alon and Boppana (1987) strengthened the combinatorial arguments of Razborov and proved an exponential lower bound for the monotone circuit complexity of the clique function. In this section we describe a special case of this bound by showing that there are no linear size monotone circuits that decide if a given graph contains a triangle. Although this result is much weaker than the ones stated above, it illustrates nicely all the probabilistic considerations in the more complicated proofs and avoids some of the combinatorial subtleties, whose detailed proofs can be found in the above mentioned papers.

Put $n = \binom{m}{2}$, and let $x_1, x_2, \ldots, x_n$ be n Boolean variables representing the edges of a graph on the set of vertices $\{1, 2, \ldots, m\}$. Let $T = T(x_1, \ldots, x_n)$ be the monotone Boolean function whose value is 1 if the corresponding graph contains a triangle. Clearly, there is a binary monotone circuit of size $O(m^3)$ computing T. Thus, the following theorem is tight, up to a poly-logarithmic factor.

Theorem 11.4.1 *The monotone circuit complexity of T is at least* $\Omega(m^3/\log^4 m)$.

Before we present the proof of this Theorem we introduce some notation and prove a simple lemma. For any Boolean function $f = f(x_1, \ldots, x_n)$ define $A(f) = \{(x_1, \ldots, x_n) \in \{0,1\}^n : f(x_1, \ldots, x_n) = 1\}$. Clearly $A(f \vee g) = A(f) \cup A(g)$ and $A(f \wedge g) = A(f) \cap A(g)$. Let C be a monotone circuit of size s computing the function $f = f(x_1, \ldots, x_n)$. Clearly, C supplies a monotone straight-line program of length s computing f, i.e., a sequence of functions $x_1,\ x_2, \ldots, x_n,\ f_1, \ldots, f_s$, where $f_s = f$ and each f_i, for $1 \leq i \leq s$, is either an Or or an And of two of the previous functions. By applying the operation A we obtain a sequence $A(C)$ of subsets of $(0,1)^n$: $A_{-n} = A_{x_n}, \ldots, A_{-1} = A_{x_1}, A_1, \ldots, A_s$ where $A_{x_i} = A(x_i)$, $A_s = A(f)$ and each A_i, for $1 \leq i \leq s$ is either a union or an intersection of two of the previous subsets. Let us replace the sequence $A(C)$ by an *approximating sequence* $M(C) : M_{-n} = M_{x_n} = A_{x_n}, \ldots, M_{-1} = M_{x_1} = A_{x_1}, M_1, \ldots, M_s$ defined by replacing the union and intersection operations in $A(C)$ by the approximating operations $\sqcup$ and $\sqcap$, respectively. The exact definition of these two operations will be given later, in such a way that for all admissible M and L the inclusions

$$M \sqcup L \supseteq M \cup L \quad \text{and} \quad M \sqcap L \subseteq M \cap L \tag{11.9}$$

will hold. Thus $M_{x_i} = A_{x_i}$ for all $1 \leq i \leq n$, and if for some $1 \leq j \leq s$ we have $A_j = A_\ell \cup A_k$ then $M_j = M_\ell \sqcup M_k$, whereas if $A_j = A_\ell \cap A_k$ then $M_j = M_\ell \sqcap M_k$. In the former case put $\delta_\sqcup^j = M_j - (M_\ell \cup M_k)$ and $\delta_\sqcap^j = \phi$, and in the latter case put $\delta_\sqcap^j = (M_\ell \cap M_k) - M_j$ and $\delta_\sqcup^j = \phi$.

Lemma 11.4.2 *For all members M_i of $M(C)$*

$$A_i - \left(\bigcup_{j \leq i} \delta_\sqcap^j \right) \subseteq M_i \subseteq A_i \cup \bigcup_{j \leq i} \delta_\sqcup^j \,. \tag{11.10}$$

Proof. We apply induction on i. For $i < 0$ $M_i = A_i$ and thus (11.10) holds. Assuming (11.10) holds for all M_j with $j < i$ we prove it for i. If $A_i = A_\ell \cup A_k$, then, by the induction hypothesis

$$M_i = M_\ell \cup M_k \cup \delta^i_\sqcup \subseteq A_\ell \cup A_k \cup \bigcup_{j\le i} \delta^j_\sqcup = A_i \cup \bigcup_{j\le i} \delta^j_\sqcup$$

and

$$M_i = M_\ell \sqcup M_k \supseteq M_\ell \cup M_k \supseteq \left(A_\ell - (\bigcup_{j\le \ell} \delta^j_\sqcap)\right) \cup \left(A_k - (\bigcup_{j\le k} \delta^j_\sqcap)\right)$$
$$\supseteq A_i - (\bigcup_{j\le i} \delta^j_\sqcap)\,, \tag{11.11}$$

as needed. If $A_i = A_\ell \cap A_k$ the proof is similar. ■

Lemma 11.4.2 holds for any choice of the operations $\sqcup$ and $\sqcap$ which satisfies (11.9). In order to prove Theorem 11.4.1 we define these operations as follows. Put $r = 100\ \log^2 m$. For any set R of at most r edges on $V = \{1, 2, \ldots, m\}$, let $\lceil R \rceil$ denote the set of all graphs on V containing at least one edge of R. In particular $\lceil \phi \rceil$ is the empty set. We also let $\lceil * \rceil$ denote the set of all graphs. The elements of $M(C)$ will all have the form $\lceil R \rceil$ or $\lceil * \rceil$. Note that $A_{x_i} = M_{x_i}$ is simply the set $\lceil R \rceil$ where R is a singleton containing the appropriate single edge. For two sets R_1 and R_2 of at most r edges each, we define $\lceil R_1 \rceil \sqcap \lceil R_2 \rceil = \lceil R_1 \cap R_2 \rceil$, $\quad \lceil R_1 \rceil \sqcap \lceil * \rceil = \lceil R_1 \rceil$ and $\lceil * \rceil \sqcap \lceil * \rceil = \lceil * \rceil$. Similarly, if $|R_1 \cup R_2| \le r$ we define $\lceil R_1 \rceil \sqcup \lceil R_2 \rceil = \lceil R_1 \cup R_2 \rceil$ whereas if $|R_1 \cup R_2| > r$ then $\lceil R_1 \rceil \sqcup \lceil R_2 \rceil = \lceil * \rceil$. Finally $\lceil * \rceil \sqcup \lceil R_1 \rceil = \lceil * \rceil \sqcup \lceil * \rceil = \lceil * \rceil$.

Proof.[theorem 11.4.1] We now prove Theorem 11.4.1 by showing that there is no monotone circuit of size $s < \binom{m}{3}/2r^2$ computing the function T. Indeed, suppose this is false and let C be such a circuit. Let $M(C) = M_{x_n}, \ldots, M_{x_1},\ M_1, \ldots, M_s$ be an approximating sequence of length s obtained from C as described above. By Lemma 11.4.2

$$A(T) - \left(\bigcup_{j\le s} \delta^j_\sqcap\right) \subseteq M_s \subseteq A(T) \cup \bigcup_{j\le s} \delta^j_\sqcup\ . \tag{11.12}$$

We consider two possible cases.

Case 1 $\quad M_s = \lceil R \rceil$, where $|R| \le r$.

Let us choose a random triangle Δ on $\{1, 2, \ldots, m\}$. Clearly

$$\Pr(\Delta \in M_s) \le \frac{r \cdot (m-2)}{\binom{m}{3}} < \frac{1}{2}\ .$$

Moreover, for each fixed $j,\ j \le s$

$$\Pr(\Delta \in \delta^j_\sqcap) \le \frac{r^2}{\binom{m}{3}}\ .$$

This is because if $\delta^j_\sqcap \ne \phi$, then $\delta^j_\sqcap = (\lceil R_1 \rceil \cap \lceil R_2 \rceil) - \lceil R_1 \cap R_2 \rceil$ for some two sets of edges $R_1,\ R_2$, each of cardinality at most r. The only triangles in this difference are those containing an edge from R_1 and another edge from R_2 (and no edge of both). Since there are at most r^2 such triangles the last inequality follows. Since $s < \binom{m}{3}/2r^2$ the last two inequalities imply that $Pr(\Delta \notin M_s \text{ and } \Delta \notin \bigcup_{j \le s} \delta_{\sqcap}^j) > 0$ and thus there is such a triangle Δ. Since this triangle belongs to $A(T)$ this contradicts (11.12), showing that Case 1 is impossible.

Case 2 $M_s = \lceil * \rceil$.

Let B be a random spanning complete bipartite graph on $V = \{1, 2, \ldots, m\}$ obtained by coloring each vertex in V randomly and independently by 0 or 1 and taking all edges connecting vertices with distinct colors. Since M_s is the set of all graphs, $B \in M_s$. Also $B \notin A(T)$, as it contains no triangle. We claim that for every fixed j, $j \le s$,

$$\Pr(B \in \delta_{\sqcup}^j) \le 2^{-\sqrt{r}/2} < \frac{1}{m^5}\,. \tag{11.13}$$

Indeed, if $\delta_{\sqcup}^j \ne \phi$, then $\delta_{\sqcup}^j = \lceil * \rceil - (\lceil R_1 \rceil \cup \lceil R_2 \rceil)$, where $|R_1 \cup R_2| > r$. Consider the graph whose set of edges is $R_1 \cup R_2$. Let d be its maximum degree. By Vizing's theorem the set of its edges can be partitioned into at most $d + 1$ matchings. Thus either $d > \frac{\sqrt{r}}{2}$ or the size of the maximum matching in this graph is at least $\sqrt{r}/2$. It follows that our graph contains a set of $k = \sqrt{r}/2$ edges $e_1, \ldots, e_k$ which form either a star or a matching. In each of these two cases $Pr(e_i \in B) = \frac{1}{2}$ and these events are mutually independent. Hence

$$Pr(B \notin \lceil R_1 \rceil \cup \lceil R_2 \rceil) \le 2^{-\sqrt{r}/2}\,,$$

implying (11.13). Note that a similar estimate can be established without Vizing's theorem by observing that B does not belong to $(\lceil R_1 \rceil \cup \lceil R_2 \rceil)$ if and only if the vertices in any connected component of the graph whose edges are $R_1 \cup R_2$ belong to the same color class of B.

Since $s < \binom{m}{3}/2r^2 < m^5$, inequality (11.13) implies that there is a bipartite B such that $B \in M_s$, $B \notin A(T)$ and $B \notin \bigcup_{j \le s} \delta_{\sqcup}^j$. This contradicts (11.12), shows that Case 2 is impossible and hence completes the proof of Theorem 11.4.1

■

11.5 FORMULAE

Recall that a formula is a circuit in which every fanout is at most 1. Unlike in the case of circuits, there are known superlinear lower bounds for the minimum size of formulae computing various explicit NP-functions over the full binary basis. For a Boolean function $f = f(x_1, \ldots, x_n)$, let us denote by $L(f)$ the minimum number of And and Or gates in a formula that uses And, Or and Not gates and computes f. By De Morgan rules we may assume that all Not gates appear in the first level of this formula. We conclude this chapter with a simple result of Subbotovskaya (1961) , which implies that for the parity function $f = x_1 \oplus \cdots \oplus$

x_n, $L(f) \geq \Omega(n^{3/2})$. This bound has been improved later by Khrapchenko (1971) to $L(f) = n^2 - 1$. However, we present here only the weaker $\Omega(n^{3/2})$ lower bound, not only because it demonstrates, once more, the power of relatively simple probabilistic arguments, but also because a modification of this proof enabled Andreev (1987) to obtain an $\Omega(n^{5/2}/(\log n)^{O(1)})$ lower bound for $L(g)$ for another NP-function $g = g(x_1, \ldots, x_n)$. J. Håstad (1998) , later improved this lower bound to $\Omega(n^{3-o(1)})$. This is at present the largest known lower bound for the formula - complexity of an NP-function of n variables over a complete basis.

The method of Subbotovskaya is based on random restrictions similar to the ones used in Section 11.2 . The main lemma is the following.

Lemma 11.5.1 *Let $f = f(x_1, \ldots, x_n)$ be a non-atom Boolean function of n variables. Then there is an i, $1 \leq i \leq n$ and an $\varepsilon \in \{0, 1\}$ such that for the function $g = f(x_1, \ldots, x_{i-1}, \varepsilon, x_{i+1}, \ldots, x_n)$ of $n-1$ variables obtained from f by substituting $x_i = \varepsilon$, the following inequality holds;*

$$(L(g) + 1) \leq (1 - \frac{3}{2n})(L(f) + 1) \leq (1 - \frac{1}{n})^{3/2}(L(f) + 1) .$$

Proof. Fix a formula F computing f with $l = L(f)$ And and Or gates. F can be represented by a binary tree each of whose $l + 1$ leaves is labeled by an atom x_i or $\overline{x}_i$. Let us choose, randomly, a variable x_i, $1 \leq i \leq n$ according to a uniform distribution, and assign to it a random binary value $\varepsilon \in \{0, 1\}$. When we substitute the values ε and $1 - \varepsilon$ to x_i and $\overline{x}_i$, respectively, the number of leaves in F is reduced; the expected number of leaves omitted in this manner is $(l + 1)/n$. However, further reduction may occur. Indeed, suppose a leaf is labeled x_i and it feeds, say, an And gate $x_i \wedge H$ in F. Observe that we may assume that the variable x_i does not appear in the subformula H, as otherwise F can be simplified by substituting $x_i = 1$ in H. If $x_i = \varepsilon = 0$, then H can be deleted once we substitute the value for x_i, thus further decreasing the number of leaves. Since the behaviour of this effect is similar for an Or-gate (and also for $\overline{x}_i$ instead of x_i), it follows that the expected number of additional leaves omitted is at least $(l + 1)/2n$. Hence the expected number of remaining leaves in the simplified formula is at most $(l + 1)[1 - \frac{3}{2n}]$, as claimed. ■

By repeatedly applying Lemma 11.5.1 we obtain;

Corollary 11.5.2 *If $f = f(x_1, \ldots, x_n)$ and $L(f) \leq (\frac{n}{k})^{3/2} - 1$, then one can assign values to $n - k$ variables so that the resulting function g is an atom.*

Proof. Repeated application of Lemma 11.5.1 $n - k$ times yields a g with

$$(L(g) + 1) \leq \prod_{i=k+1}^{n} (1 - \frac{1}{i})^{3/2}(L(f) + 1) = (k/n)^{3/2}(L(f) + 1) \leq 1 .$$

Hence g is either x_i or $\overline{x}_i$ for some i. ■

Corollary 11.5.3 *For the parity function $f = x_1 \oplus \cdots \oplus x_n$,*

$$L(f) > (\frac{n}{2})^{3/2} - 1 \ .$$

11.6 EXERCISES

1. Show that there exists a constant c such that the number of binary Boolean circuits of size s is at most $(cs)^s$.

2. Let f be a Boolean formula in the n variables $x_1, x_2, \ldots, x_n$, where f is an AND of an arbitrary (finite) number of clauses, each clause is an OR of 10 literals, where each literal is either a variable or its negation, and suppose each variable appears (negated or un-negated) in at most 10 clauses. Prove that f is satisfiable.

3. (*) Prove that there is a bounded-depth, polynomial size, monotone circuit of n Boolean inputs $x_1, x_2, \ldots, x_n$, computing a function f whose value is 1 if $\sum_{i=1}^{n} x_i \geq n/2 + n/\log n$, and is 0 if $\sum_{i=1}^{n} x_i \leq n/2 - n/\log n$.

THE PROBABILISTIC LENS: *Maximal Antichains*

A family $\mathcal{F}$ of subsets of $\{1, \ldots, n\}$ is called an *antichain* if no set of $\mathcal{F}$ is contained in another.

Theorem 1 *Let $\mathcal{F}$ be an antichain. Then*

$$\sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \leq 1.$$

Proof. Let σ be a uniformly chosen permutation of $\{1, \ldots, n\}$ and set

$$\mathcal{C}_\sigma = \{\{\sigma(j) : 1 \leq j \leq i\} : 0 \leq i \leq n\}$$

(The cases $i = 0, n$ give $\emptyset, \{1, \ldots, n\} \in \mathcal{C}$ respectively.) Define a random variable

$$X = |\mathcal{F} \cap \mathcal{C}_\sigma|.$$

We decompose

$$X = \sum_{A \in \mathcal{F}} X_A,$$

where X_A is the indicator random variable for $A \in \mathcal{C}$. Then

$$E[X_A] = \Pr[A \in \mathcal{C}_\sigma] = \frac{1}{\binom{n}{|A|}},$$

since $\mathcal{C}_\sigma$ contains precisely one set of size $|A|$, which is distributed uniformly among the $|A|$-sets. By linearity of expectation

$$E[X] = \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}}.$$

For *any* σ, $\mathcal{C}_\sigma$ forms a chain - every pair of sets is comparable. Since $\mathcal{F}$ is an antichain we *must* have $X = |\mathcal{F} \cap \mathcal{C}_\sigma| \leq 1$. Thus $E[X] \leq 1$. ■

Corollary 2 [Sperner's Theorem] *Let $\mathcal{F}$ be an antichain. Then*

$$|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Proof. The function $\binom{n}{x}$ is maximized at $x = \lfloor n/2 \rfloor$ so that

$$1 \geq \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \geq \frac{|\mathcal{F}|}{\binom{n}{\lfloor n/2 \rfloor}}.$$

■

12

Discrepancy

The mystery, as well as the glory, of mathematics lies not so much in the fact that abstract theories do turn out to be useful in solving problems but in that wonder of wonders, the fact that a theory meant for solving one type of problem is often the only way of solving problems of entirely different kinds, problems for which the theory was not intended. These coincidences occur so frequently that they must belong to the essence of mathematics.
– Gian-Carlo Rota

12.1 BASICS

Suppose we are given a finite family of finite sets. Our object is to color the underlying points Red and Blue so that all of the sets have nearly the same number of Red and Blue points. It may be that our cause is hopeless - if the family consists of all subsets of a given set Ω then regardless of the coloring some set, either the Red or the Blue points, will have size at least half that of Ω and be monochromatic. In the other extreme, should the sets of the family be disjoint then it is trivial to color so that all sets have the same number of Red and Blue points or, at worst if the cardinality is odd, the number of Red and Blue points differing by only one. The discrepancy will measure how good a coloring we may find.

To be formal, let a family $\mathcal{A}$ of subsets of Ω be given. Rather than using Red and Blue we consider colorings as maps

$$\chi : \Omega \longrightarrow \{-1, +1\}.$$

For any $A \subset \Omega$ we set

$$\chi(A) = \sum_{a \in A} \chi(a).$$

Define the discrepancy of $\mathcal{A}$ with respect to χ by

$$disc(\mathcal{A}, \chi) = \max_{A \in \mathcal{A}} |\chi(A)|$$

and the discrepancy of $\mathcal{A}$ by

$$disc(\mathcal{A}) = \min_{\chi:\Omega \longrightarrow \{-1,+1\}} disc(\mathcal{A}, \chi).$$

Other equivalent definitions of discrepancy reveal its geometric aspects. Let $\mathcal{A} = \{S_1, \ldots, S_m\}$, $\Omega = \{1, \ldots, n\}$ and let $B = [b_{ij}]$ be the $m \times n$ incidence matrix: $b_{ij} = 1$ if $j \in S_i$, otherwise $b_{ij} = 0$. A coloring χ may be associated with the vector $u = (\chi(1), \ldots, \chi(n)) \in \{-1, +1\}^n$ so that $Bu^T = (\chi(S_1), \ldots, \chi(S_m))$ and

$$disc(\mathcal{A}) = \min_{u \in \{-1,+1\}^n} |Bu^T|_\infty$$

where $|v|_\infty$ is the L^∞-norm, the maximal absolute value of the coordinates. Similarly, letting v_j denote the j-th column vector of B (the profile of point j)

$$disc(\mathcal{A}) = \min |\pm v_1 \pm \ldots \pm v_n|_\infty$$

where the minimum ranges over all 2^n choices of sign.

We will generally be concerned with upper bounds to the discrepancy. Unravelling the definitions, $disc(\mathcal{A}) \leq K$ if and only if there *exists* a coloring χ for which $|\chi(A)| \leq K$ for all $A \in \mathcal{A}$. Naturally, we try the random coloring.

Theorem 12.1.1 *Let $\mathcal{A}$ be a family of n subsets of an m-set Ω. Then*

$$disc(\mathcal{A}) \leq \sqrt{2m \ln(2n)}.$$

Proof. Let $\chi : \Omega \longrightarrow \{-1, +1\}$ be random. For $A \subset \Omega$ let X_A be the indicator random variable for $|\chi(A)| > \alpha$ where we set $\alpha = \sqrt{2m \ln(2n)}$. If $|A| = a$ then $\chi(A)$ has distribution S_a so by Theorem A.1.1

$$E[X_A] = \Pr[|\chi(A)| > \alpha] < 2e^{-\alpha^2/2a} \leq 2e^{-\alpha^2/2m} = 1/n$$

by our propitious choice of α. Let X be the number of $A \in \mathcal{A}$ with $|\chi(A)| > \alpha$ so that

$$X = \sum_{A \in \mathcal{A}} X_A$$

and linearity of expectation gives

$$E[X] = \sum_{A \in \mathcal{A}} E[X_A] < |\mathcal{A}|(1/n) = 1.$$

Thus for some χ we must have $X = 0$. This means $disc(\mathcal{A}, \chi) \leq \alpha$ and therefore $disc(\mathcal{A}) \leq \alpha$. ■

12.2 SIX STANDARD DEVIATIONS SUFFICE

When $\mathcal{A}$ has both n sets and n points Theorem 12.1.1 gives

$$disc(\mathcal{A}) = O(\sqrt{n\ln(n)}).$$

This is improved by the following result. Its proof resembles that of the main result of Beck (1981) . The approach via entropy was suggested by R. Boppana.

Theorem 12.2.1 [Spencer (1985a)] *Let $\mathcal{A}$ be a family of n subsets of an n-element set Ω. Then*

$$disc(\mathcal{A}) < 6\sqrt{n}.$$

With $\chi : \Omega \to \{-1, +1\}$ random, $A \in \mathcal{A}$, $\chi(A)$ has zero mean and variance at most $\sqrt{n}$. If $|\chi(A)| > 6\sqrt{n}$ then $\chi(A)$ is at least six standard deviations off the mean. The probability of this occuring is very small but a fixed positive constant and the number of sets $A \in \mathcal{A}$ is going to infinity. In fact, a random χ almost always will *not* work. The specific constant 6 (actually 5.32) was the result of specific calculations which could certainly be further improved and will not concern us here. Rather, we show Theorem 12.2.1 with "6"=11. A map

$$\chi : \Omega \longrightarrow \{-1, 0, +1\}$$

will be called a *partial* coloring. When $\chi(a) = 0$ we say a is uncolored. We define $\chi(A)$ as before.

Lemma 12.2.2 *Let $\mathcal{A}$ be a family of n subsets of an n-set Ω. Then there is a partial coloring χ with at most $10^{-9}n$ points uncolored such that*

$$|\chi(A)| \leq 10\sqrt{n}$$

for all $A \in \mathcal{A}$.

Here the values 10 and 10^{-9} are not best possible. The significant point is that they are absolute constants. Label the sets of $\mathcal{A}$ by $A_1, \ldots, A_n$ for convenience. Let

$$\chi : \Omega \longrightarrow \{-1, +1\}$$

be random. For $1 \leq i \leq n$ define

$$b_i = \text{ nearest integer to } \chi(A_i)/(20\sqrt{n}).$$

For example, $b_i = 0$ when $-10\sqrt{n} < \chi(A_i) < 10\sqrt{n}$ and $b_i = -3$ when $-70\sqrt{n} < \chi(A_i) < -50\sqrt{n}$. From A.1.1 (as in Theorem 12.1.1)

$$\Pr[b_i = 0] > 1 - 2e^{-50}$$

$$\Pr[b_i = 1] = \Pr[b_i = -1] < e^{-50}$$

$$\Pr[b_i = 2] = \Pr[b_i = -2] < e^{-450},$$

and, in general,

$$\Pr[b_i = s] = \Pr[b_i = -s] < e^{-50(2s-1)^2}$$

Now we bound the *entropy* $H[b_i]$. This important concept is explored more fully in Section 14.6. Letting $p_j = \Pr[b_i = j]$,

$$\begin{aligned} H(b_i) &= \sum_{j=-\infty}^{+\infty} -p_j \log_2(p_j) \\ &\leq (1 - 2e^{-50})[-\log_2(1 - 2e^{-50})] \\ &+ 2e^{-50}[-\log_2 e^{-50}] \\ &+ 2e^{-450}[-\log_2 e^{-450}] + \ldots \end{aligned}$$

The infinite sum clearly converges and is strongly dominated by the second term. Calculation gives

$$H(b_i) \leq \epsilon = 3 \times 10^{-20}.$$

Now consider the n-tuple $(b_1, \ldots, b_n)$. Of course, there may be correlation among the b_i. Indeed, if S_i and S_j are nearly equal then b_i and b_j will usually be equal. But by proposition 14.6.2 entropy is subadditive. Hence

$$H((b_1, \ldots, b_n)) \leq \sum_{i=1}^{n} H(b_i) \leq \epsilon n.$$

If a random variable Z assumes no value with probability greater than 2^{-t} then $H(Z) \geq t$. In contrapositive form, there is a particular n-tuple $(s_1, \ldots, s_n)$ so that

$$\Pr[(b_1, \ldots, b_n) = (s_1, \ldots, s_n)] \geq 2^{-\epsilon n}.$$

Our probability space was composed of the 2^n possible colorings χ, all equally likely. Thus, shifting to counting, there is a set $\mathcal{C}'$ consisting of at least $2^{(1-\epsilon)n}$ colorings $\chi : \Omega \longrightarrow \{-1, +1\}$, all having the same value $(b_1, \ldots, b_n)$.

Let us think of the class $\mathcal{C}$ of all colorings $\chi : \Omega \longrightarrow \{-1, +1\}$ as the Hamming Cube $\{-1, +1\}^n$ endowed with the Hamming metric

$$\rho(\chi, \chi') = |\{a : \chi(a) \neq \chi'(a)\}|.$$

Kleitman (1966a) has proven that if $\mathcal{D} \subset \mathcal{C}$ and

$$|\mathcal{D}| \geq \sum_{i \leq r} \binom{n}{i}$$

with $r \leq \frac{n}{2}$ then $\mathcal{D}$ has diameter at least $2r$. That is, the set of a given size with minimal diameter is the ball. ($\mathcal{D}$ has diameter at least r trivially which would suffice to prove Lemma 12.2.2 and Theorem 12.2.1 with weaker values for the constants.)
Proof. In our case we may take $r = \alpha n$ as long as $\alpha < \frac{1}{2}$ and

$$2^{H(\alpha)} \le 2^{1-\epsilon}.$$

Calculation gives that we may take $\alpha = \frac{1}{2}(1 - 10^{-9})$ with room to spare. (Taylor Series gives

$$H(\frac{1}{2} - x) \sim 1 - \frac{2}{\ln 2}x^2$$

for x small.) Thus $\mathcal{C}'$ has diameter at least $n(1 - 10^{-9})$. Let $\chi_1, \chi_2 \in \mathcal{C}'$ be at maximal distance. We set

$$\chi = \frac{\chi_1 - \chi_2}{2}.$$

χ is a partial coloring of Ω. $\chi(a) = 0$ if and only if $\chi_1(a) = \chi_2(a)$ which occurs for $n - \rho(\chi_1, \chi_2) \le 10^{-9}n$ coordinates a. Finally, and crucially, for each $1 \le i \le n$ the colorings χ_1, χ_2 yield the same value b_i which means that $\chi_1(A_i)$ and $\chi_2(A_i)$ lie on a common interval of length $20\sqrt{n}$. Thus

$$|\chi(A_i)| = |\frac{\chi_1(A_i) - \chi_2(A_i)}{2}| \le 10\sqrt{n},$$

as desired. ■

Theorem 12.2.1 requires a coloring of all points whereas Lemma 12.2.2 leaves $10^{-9}n$ points uncolored. The idea, now, is to iterate the procedure of Lemma 12.2.2, coloring all but, say, $10^{-18}n$ of the uncolored points on the second coloration. We cannot apply Lemma 12.2.2 directly since we have an asymmetric situation with n sets and only $10^{-9}n$ points.

Lemma 12.2.3 *Let $\mathcal{A}$ be a family of n subsets of an r-set Ω with $r \le 10^{-9}n$. Then there is a partial coloring χ of Ω with at most $10^{-40}r$ points uncolored so that*

$$|\chi(A)| < 10\sqrt{r}\sqrt{\ln(n/r)}$$

for all $A \in \mathcal{A}$.

Proof.

We outline the argument which leaves room to spare. Let $A_1, \ldots, A_n$ denote the sets of $\mathcal{A}$. Let $\chi : \Omega \longrightarrow \{-1, +1\}$ be random. For $1 \le i \le n$ define

$$b_i = \text{ nearest integer to } \frac{\chi(A_i)}{20\sqrt{r}\sqrt{\ln(n/r)}}.$$

Now the probability that $b_i = 1$ is less than $(r/n)^{50}$. The entropy $H(b_i)$ is dominated by this term and is less than

$$3(\frac{r}{n})^{50}[-\log_2((\frac{r}{n})^{50})] < 10^{-100}\frac{r}{n}.$$

The entropy of $(b_1, \ldots, b_r)$ is then less than $10^{-100}r$, one finds nearly antipodal χ_1, χ_2 with the same b's and takes $\chi = (\chi_1 - \chi_2)/2$ as before. ■

Proof.[Theorem 12.2.1] Apply Lemma 12.2.2 to find a partial coloring χ^1 and then apply Lemma 12.2.3 repeatedly on the remaining uncolored points giving $\chi^2, \chi^3, \ldots$ until all points have been colored. Let χ denote the final coloring. For any $A \in \mathcal{A}$,

$$\chi(A) = \chi^1(A) + \chi^2(A) + \ldots$$

so that

$$\begin{aligned} |\chi(A)| \quad \leq \quad & 10\sqrt{n} + 10\sqrt{10^{-9}n}\sqrt{\ln 10^9} + \\ & 10\sqrt{10^{-49}n}\sqrt{\ln 10^{49}} + 10\sqrt{10^{-89}n}\sqrt{\ln 10^{89}} + \ldots \end{aligned}$$

Removing the common $\sqrt{n}$ term gives a clearly convergent infinite series, strongly dominated by the first term so that

$$|\chi(A)| \leq 11\sqrt{n}$$

with room to spare. ■

Suppose that $\mathcal{A}$ consists of n sets on r points and $r < n$. We can apply Lemma 12.2.3 repeatedly (first applying Lemma 12.2.2 if $r > 10^{-9}n$) to give a coloring χ with

$$disc(\mathcal{A}, \chi) < K\sqrt{r}\sqrt{\ln(n/r)}$$

where K is an absolute constant. As long as $r = n^{1-o(1)}$ this improves the random coloring result of Theorem 12.1.1.

12.3 LINEAR AND HEREDITARY DISCREPANCY

We now suppose that $\mathcal{A}$ has more points than sets. We write $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $\Omega = \{1, \ldots, m\}$ and assume $m > n$. Note that $disc(\mathcal{A}) \leq K$ is equivalent to the existence of a set S, namely $S = \{j : \chi(j) = +1\}$, with $|S \cap A|$ within $K/2$ of $|A|/2$ for all $A \in \mathcal{A}$. We define the *linear discrepancy* $lindisc(\mathcal{A})$ by

$$lindisc(\mathcal{A}) = \max_{p_1,\ldots,p_m \in [0,1]} \ \min_{\epsilon_1,\ldots,\epsilon_m \in \{0,1\}} \ \max_{A \in \mathcal{A}} \Big|\sum_{i \in A} (\epsilon_i - p_i)\Big|.$$

The upper bound $lindisc(\mathcal{A}) \leq K$ means that given any $p_1, \ldots, p_m$ there is a "simultaneous roundoff" $\epsilon_1, \ldots, \epsilon_m$ so that, with $S = \{j : \epsilon_j = 1\}$, $|S \cap A|$ is within K of the weighted sum $\sum_{j \in A} p_j$ for all $A \in \mathcal{A}$. Taking all $p_j = \frac{1}{2}$, the upper bound implies $disc(\mathcal{A}) \leq 2K$. But $lindisc(\mathcal{A}) \leq K$ is much stronger. It implies, taking all $p_j = \frac{1}{3}$, the existence of an S with all $|S \cap A|$ within K of $|A|/3$, and much more. Linear discrepancy and its companion hereditary discrepancy defined below have been developed in Lovász, Spencer and Vesztergombi (1986) . For $X \subset \Omega$ let $\mathcal{A}|_X$ denote the restriction of $\mathcal{A}$ to X, i.e., the family $\{A \cap X : A \in \mathcal{A}\}$. The next result "reduces" the bounding of $disc(\mathcal{A})$ when there are more points than sets to the bounding of $lindisc(\mathcal{A})$ when the points do not outnumber the sets.

Theorem 12.3.1 *Let $\mathcal{A}$ be a family of n sets on m points with $m \geq n$. Suppose that $lindisc(\mathcal{A}|_X) \leq K$ for every subset X of at most n points. Then $lindisc(\mathcal{A}) \leq K$.*

Proof. Let $p_1, \ldots, p_m \in [0,1]$ be given. We define a reduction process. Call index j fixed if $p_j \in \{0,1\}$, otherwise call it floating, and let F denote the set of floating indices. If $|F| \leq n$ then halt. Otherwise, let y_j, $j \in F$, be a nonzero solution to the homogeneous system

$$\sum_{j \in A \cap F} y_j = 0, \ A \in \mathcal{A}$$

Such a solution exists since there are more variables ($|F|$) than equations (n) and may be found by standard techniques of linear algebra. Now set

$$p'_j = p_j + \lambda y_j, \ j \in F$$

$$p'_j = p_j, \ j \notin F$$

where we let λ be the real number of least absolute value so that for some $j \in F$ the value p'_j becomes zero or one. Critically

$$\textstyle\sum_{j \in A} p'_j = \sum_{j \in A} p_j + \lambda \sum_{j \in A \cap F} y_j = \sum_{j \in A} p_j \quad (*)$$

for all $A \in \mathcal{A}$. Now iterate this process with the new p'_j. At each iteration at least one floating j becomes fixed and so the process eventually halts at some $p^*_1, \ldots, p^*_m$. Let X be the set of floating j at this point. Then $|X| \leq n$. By assumption there exist ϵ_j, $j \in X$ so that

$$|\sum_{j \in A \cap X} p^*_j - \epsilon_j| \leq K, \ A \in \mathcal{A}.$$

For $j \notin X$ set $\epsilon_j = p^*_j$. As (*) holds at each iteration

$$\sum_{j \in A} p^*_j = \sum_{j \in A} p_j$$

and hence

$$|\sum_{j \in A}(p_j - \epsilon_j)| = |\sum_{j \in A}(p_j - p^*_j) + \sum_{j \in A \cap X}(p^*_j - \epsilon_j)| \leq K$$

for all $A \in \mathcal{A}$. ■

We now define the *hereditary discrepancy* $herdisc(\mathcal{A})$ by

$$herdisc(\mathcal{A}) = \max_{X \subseteq \Omega} disc(\mathcal{A}|_X).$$

Example. Let $\Omega = \{1, \ldots, n\}$ and let $\mathcal{A}$ consist of all intervals $[i,j] = \{i, i+1, \ldots, j\}$ with $1 \leq i \leq j \leq n$. Then $disc(\mathcal{A}) = 1$ as we may color Ω alternately $+1$ and -1. But also $herdisc(\mathcal{A}) = 1$. For given any $X \subseteq \Omega$, say with elements

$x_1 < x_2 < \ldots < x_r$, we may color X alternately by $\chi(x_k) = (-1)^k$. For any set $[i, j] \in \mathcal{A}$ the elements of $[i, j] \cap X$ are alternately colored.

Theorem 12.3.2 $\mathit{lindisc}(\mathcal{A}) \leq \mathit{herdisc}(\mathcal{A})$.

Proof. Set $K = \mathit{herdisc}(\mathcal{A})$. Let $\mathcal{A}$ be defined on $\Omega = \{1, \ldots, m\}$ and let $p_1 \ldots, p_m \in [0, 1]$ be given. First let us assume that all p_i have finite expansions when written in base two. Let T be the minimal integer so that all $p_i 2^T \in Z$. Let J be the set of i for which p_i has a one in the T-th digit of its binary expansion, i.e., so that $p_i 2^{T-1} \notin Z$. As $\mathit{disc}(\mathcal{A}|_J) \leq K$ there exist $\epsilon_j \in \{-1, +1\}$, so that

$$\left| \sum_{j \in J \cap A} \epsilon_j \right| \leq K$$

for all $A \in \mathcal{A}$. Write $p_j = p_j^{(T)}$. Now set

$$p_j^{(T-1)} = \begin{cases} p_j^{(T)} & \text{if } j \notin J \\ p_j^{(T)} + \epsilon_j 2^{-T} & \text{if } j \in J \end{cases}$$

That is, the $p_j^{(T-1)}$ are the "roundoffs" of the $p_j^{(T)}$ in the T-th place. Note that all $p_j^{(T-1)} 2^{-(T-1)} \in Z$. For any $A \in \mathcal{A}$

$$\left| \sum_{j \in A} p_j^{(T-1)} - p_j^{(T)} \right| = \left| \sum_{j \in J \cap A} 2^{-T} \epsilon_j \right| \leq 2^{-T} K.$$

Iterate this procedure, finding $p_j^{(T-2)}, \ldots, p_j^{(1)}, p_j^{(0)}$. All $p_j^{(0)} 2^{-0} \in Z$ so all $p_j^{(0)} \in \{0, 1\}$ and

$$\left| \sum_{j \in A} p_j^{(0)} - p_j^{(T)} \right| \leq \sum_{i=1}^{T} \left| \sum_{j \in A} p_j^{(i-1)} - p_j^{(i)} \right| \leq \sum_{i=1}^{T} 2^{-i} K \leq K,$$

as desired.

What about general $p_1, \ldots, p_m \in [0, 1]$? We can be flip and say that, at least to a computer scientist, all real numbers have finite binary expansions. More rigorously, the function

$$f(p_1, \ldots, p_m) = \min_{\epsilon_1, \ldots, \epsilon_m \in \{0,1\}} \max_{A \in \mathcal{A}} \left| \sum_{i \in A} (\epsilon_i - p_i) \right|$$

is the finite minimum of finite maxima of continuous functions and thus is continuous. The set of $p_1, \ldots, p_m \in [0, 1]$ with all $p_i 2^T \in Z$ for some T is a dense subset of $[0, 1]$. As $f \leq K$ on this dense set, $f \leq K$ for all $p_1, \ldots, p_m \in [0, 1]$.

■

Corollary 12.3.3 *Let $\mathcal{A}$ be a family of n sets on m points. Suppose $disc(\mathcal{A}|_X) \leq K$ for every subset X with at most n points. Then $disc(\mathcal{A}) \leq 2K$.*

Proof. For every $X \subseteq \Omega$ with $|X| \leq n$, $herdisc(\mathcal{A}|_X) \leq K$ so by Theorem 12.3.2 $lindisc(\mathcal{A}|_X) \leq K$. By Theorem 12.3.1 $lindisc(\mathcal{A}) \leq K$. But

$$disc(\mathcal{A}) \leq 2lindisc(\mathcal{A}) \leq 2K.$$

■

Corollary 12.3.4 *For any family $\mathcal{A}$ of n sets of arbitrary size*

$$disc(\mathcal{A}) \leq 12\sqrt{n}.$$

Proof. Apply Theorem 12.2.1 and Corollary 12.3.3. ■

12.4 LOWER BOUNDS

We now give two quite different proofs that, up to a constant factor, Corollary 12.3.4 is best possible. A Hadamard Matrix is a square matrix $H = (h_{ij})$ with all $h_{ij} \in \{-1, +1\}$ and with row vectors mutually orthogonal (and hence with column vectors mutually orthogonal). Let H be a Hadamard matrix of order n and let $v = (v_1, \ldots, v_n)$, $v_i \in \{-1, +1\}$. Then

$$Hv = v_1c_1 + \ldots v_nc_n$$

where c_i denotes the i-th column vector of H. Writing $Hv = (L_1, \ldots, L_n)$ and letting $|c|$ denote the usual Euclidean norm

$$L_1^2 + \ldots + L_n^2 = |Hv|^2 = v_1^2|c_1|^2 + \ldots v_n^2|c_n|^2 = n + \ldots + n = n^2$$

since the c_i are mutually orthogonal. Hence some $L_i^2 \geq n$ and thus

$$|Hv|_\infty = \max(|L_1|, \ldots, |L_n|) \geq \sqrt{n}.$$

Now we transfer this result to one on families of sets. Let H be a Hadamard matrix of order n with first row and first column all ones. (Any Hadamard matrix can be so "normalized" by multiplying appropriate rows and columns by -1.) Let J denote the all ones matrix of order n. Let $v_1, \ldots, L_1, \ldots$ be as above. Then

$$L_1 + \ldots + L_n = \sum_{i,j=1}^{n} v_jh_{ij} = \sum_{j=1}^{n} v_j \sum_{i=1}^{n} h_{ij} = nv_1 = \pm n,$$

since the first column sums to n but the other columns, being orthogonal to it, sum to zero. Set $\lambda = v_1 + \ldots + v_n$ so that $Jv = (\lambda, \ldots, \lambda)$ and

$$(H + J)v = (L_1 + \lambda, \ldots, L_n + \lambda).$$

We calculate

$$|(H + J)v|^2 = \sum_{i=1}^{n}(L_i + \lambda)^2 = \sum_{i=1}^{n}(L_i^2 + 2\lambda L_i + \lambda^2) = n^2 \pm 2n\lambda + n\lambda^2.$$

Assume n is even. (Hadamard matrices don't exist for odd n, except $n = 1$.) Then λ is an even integer. The quadratic (in λ) $n^2 \pm 2n\lambda + n\lambda^2$ has a minimum at ∓ 1 and so under the restiction of being an even integer its minimum is at $\lambda = 0, \mp 2$ and so

$$|(H + J)v|^2 \geq n^2.$$

Again, some coordinate must be at least $\sqrt{n}$. Setting $H^* = \frac{H+J}{2}$

$$|H^* v|_\infty \geq \sqrt{n}/2.$$

Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be any family of subsets of $\Omega = \{1, \ldots, n\}$ and let M denote the corresponding $m \times n$ incidence matrix. A coloring $\chi : \Omega \longrightarrow \{-1, +1\}$ corresponds to a vector $v = (\chi(1), \ldots, \chi(n)) \in \{-1, +1\}^n$. Then

$$disc(\mathcal{A}, \chi) = |Mv|_\infty$$

and

$$disc(\mathcal{A}) = \min_{v \in \{-1,+1\}^n} |Mv|_\infty.$$

In our case H^* has entries $0, 1$. Thus we have:

Theorem 12.4.1 *If a Hadamard Matrix exists of order $n > 1$ then there exists a family $\mathcal{A}$ consisting of n subsets of an n-set with*

$$disc(\mathcal{A}) \geq \sqrt{n}/2.$$

Remark. While it is not known precisely for which n a Hadamard matrix exists (the Hadamard conjecture is that they exist for $n = 1, 2$ and all multiples of 4; see, e.g., Hall (1986)), it is known that the orders of Hadamard matrices are dense in the sense that for all ϵ if n is sufficiently large there will exist a Hadamard matrix of order between n and $n(1 - \epsilon)$. This result suffices to extend Theorem 12.4.1 to an asymptotic result on all n.

Our second argument for the existence of $\mathcal{A}$ with high discrepancy involves turning the probabilistic argument "on its head". Let M be a random $0, 1$ matrix of order n. Let $v = (v_1, \ldots, v_n), v_j = \pm 1$ be fixed and set $Mv = (L_1, \ldots, L_n)$. Suppose half of the $v_j = +1$ and half are -1. Then

$$L_i \sim B(\frac{n}{2}, \frac{1}{2}) - B(\frac{n}{2}, \frac{1}{2})$$

which has roughly the Normal Distribution $N(0, \sqrt{n}/2)$. Pick $\lambda > 0$ so that

$$\int_{-\lambda}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt < \frac{1}{2}.$$

Then

$$\Pr[|L_i| < \lambda\sqrt{n}/2] < \frac{1}{2}.$$

When v is imbalanced the same inequality holds, we omit the details. Now, crucially, the L_i are mutually independent as each entry of M was independently chosen. Thus

$$\Pr[|L_i| < \lambda\sqrt{n}/2 \text{ for all } 1 \le i \le n] < (\frac{1}{2})^n.$$

There are "only" 2^n possible v. Thus the expected number of v for which $|Mv|_\infty < \lambda\sqrt{n}/2$ is less than $2^n 2^{-n} = 1$. For some M this value must be zero, there are no such v. The corresponding family $\mathcal{A}$ thus has

$$disc(\mathcal{A}) > \lambda\sqrt{n}/2.$$

12.5 THE BECK-FIALA THEOREM

For any family $\mathcal{A}$ let $deg(\mathcal{A})$ denote the maximal number of sets containing any particular point. The following result due to Beck and Fiala (1981) uses only methods from linear algebra and thus is technically outside the scope we have set for this book. We include it both for the sheer beauty of the proof and because the result itself is very much in the spirit of this chapter.

Theorem 12.5.1 *Let $\mathcal{A}$ be a finite family of finite sets, no restriction on either the number of sets nor on the cardinality of the sets, with $deg(\mathcal{A}) \le t$. Then*

$$disc(\mathcal{A}) \le 2t - 1.$$

Proof. For convenience write $\mathcal{A} = \{A_1, \ldots, A_m\}$ with all $A_i \subseteq \Omega = \{1, \ldots, n\}$. To each $j \in \Omega$ there is assigned a value x_j which will change as the proof progresses. Initially all $x_j = 0$. At the end all $x_j = \pm 1$. We will have $-1 \le x_j \le +1$ at all times and once $x_j = \pm 1$ it "sticks" there and that becomes its final value. A set S_i has value $\sum_{j \in S_i} x_j$. At any time j is called fixed if $x_j = \pm 1$, otherwise it is *floating*. A set S_i is safe if it has fewer than t floating points, otherwise it is *active*. Note, crucially, that as points are in at most t sets and active sets contain more than t floating points there must be fewer active sets than floating points.

We insist at all times that all active sets have value zero. This holds initially since all sets have value zero. Suppose this condition holds at some stage. Consider x_j a variable for each floating j and a constant for each fixed j. The condition that S_i has value zero then becomes a linear equation in these variables. This is an underdetermined system, there are fewer linear conditions (active sets) than variables (floating points). Hence we may find a line, parametrized

$$x'_j = x_j + \lambda y_j, \; j \text{ floating}$$

on which the active sets retain value zero. Let λ be the smallest value for which some x'_j becomes ± 1 and replace each x_j by x'_j. (Geometrically, follow the line until reaching the boundary of the cube in the space over the floating variables.) This process has left fixed variables fixed and so safe sets stayed safe sets (though active sets may have become safe) and so the condition still holds. In addition, at least one previously floating j has become fixed.

We iterate the above procedure until all j have become fixed. (Towards the end we may have no active sets at which time we may simply set the floating x_j to ± 1 arbitrarily.) Now consider any set S_i. Initially it had value zero and it retained value zero while it contained at least t floating points. Consider the time when it first becomes safe, say $1, \ldots, l$ were its floating points. At this moment its value is zero. The variables $y_1, \ldots, y_l$ can now change less than two to their final value since all values are in $[-1, +1]$. Thus, in total, they may change less than $2t$. Hence the final value of S_i is less than $2t$ and, as it is an integer, it is at most $2t - 1$. ■

Conjecture 12.5.2 *If $deg(\mathcal{A}) \leq t$ then $disc(\mathcal{A}) \leq K\sqrt{t}$, K an absolute constant.*

This conjecture seems to call for a melding of probabilistic methods and linear algebra. The constructions of t sets on t points, described in Section 12.4, show that, if true, this conjecture would be best possible.

12.6 EXERCISES

1. Let $\mathcal{A}$ be a family of n subsets of $\Omega = \{1, \ldots, m\}$ with m even. Let $\chi(i)$, $1 \leq i \leq \frac{m}{2}$ be independent and uniform in $\{-1, +1\}$ and set $\chi(i + \frac{m}{2}) = -\chi(i)$ for $1 \leq i \leq \frac{m}{n}$. Using this notion of random coloring improve Theorem 12.1.1 by showing

$$disc(\mathcal{A}) \leq \sqrt{\frac{m}{2} \ln(2n)}.$$

2. Let $\vec{v}_1, \ldots, \vec{v}_s \in R^n$. Let $x_1, \ldots, x_s \in [-1, +1]$ such that $\sum_{i=1}^{s} x_i \vec{v}_i = \vec{0}$ and such that $x_i \in \{-1, +1\}$ for all but at most n values of i. Let $\vec{v}_{s+1} \in R^n$. Use the linear ideas of §12.5 to find $x'_1, \ldots x'_s, x'_{s+1}$ with the following properties:

   - $\sum_{i=1}^{s+1} x'_i \vec{v}_i = \vec{0}$
   - All $x'_i \in [-1, +1]$
   - $x'_i \in \{-1, +1\}$ for all but at most n values of i
   - $x'_i = x_i$ whenever $x_i \in \{-1, +1\}$

Use the above to prove the following result of Bárány and Greenberg: Let $|\cdot|$ be an arbitrary norm in R^n. Let $\vec{v}_1, \ldots, \vec{v}_s \in R^n$ with all $|v_i| \leq 1$. Then there exist $x_1, \ldots, x_s \in \{-1, +1\}$ such that

$$\left| \sum_{i=1}^{t} x_i \vec{v}_i \right| \leq 2n$$

for all $1 \leq t \leq s$.

3. Let $A_1, \ldots, A_n \subset \Omega = \{1, \ldots, m\}$ with $m \sim n \ln n$. Assume further that all $|A_i| \leq n$. Use the methods of Theorem 12.2.1, including Kleitman's Theorem, to prove that there exists $\chi : \{1, \ldots, m\} \to \{-1, 0, +1\}$ such that all $\chi(A_i) = O(\sqrt{n \ln \ln n})$ and $\chi(x) = 0$ for at most n vertices x. Use Theorem 12.2.1 to deduce the existence of $\chi : \{1, \ldots, m\} \to \{-1, +1\}$ such that all $\chi(A_i) = O(\sqrt{n \ln \ln n})$.

THE PROBABILISTIC LENS:
Unbalancing Lights

For any $m \times n$ matrix $B = (b_{ij})$ with coefficients $b_{ij} = \pm 1$ set

$$F[B] = \max_{x_i, y_j = \pm 1} \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j b_{ij}.$$

As in §2.5 we may interpret B as an $m \times n$ array of lights, each either on ($b_{ij} = +1$) or off ($b_{ij} = -1$). For each row and each column there is a switch which, when pulled, changes all lights in that line from on to off or from off to on. Then $F[B]$ gives the maximal achievable number of lights on minus lights off. In §2.5 we found a lower bound for $F[B]$ when $m = n$. Here we set $n = 2^m$ and find the precise best possible lower bound.

With $n = 2^m$ let A be an $m \times n$ matrix with coefficients ± 1 containing every possible column vector precisely once. We claim $F[A]$ is the minimal value of $F[B]$ over all $m \times n$ matrices B.

For any given B let $x_1, \ldots, x_m = \pm 1$ be independently and uniformly chosen and set

$$X_j = \sum_{i=1}^{m} x_i b_{ij}$$

$$X = |X_1| + \ldots + |X_m|,$$

so that

$$F[B] = \max_{y_j = \pm 1} \max_{x_i = \pm 1} \sum_{j=1}^{n} y_j X_j = \max_{x_i = \pm 1} \sum_{j=1}^{n} |X_j| = \max X.$$

Regardless of the b_{ij}, X_i has distribution S_m so that $E[|X_i|] = E[|S_m|]$ and, by linearity of expectation,

$$E[X] = nE[|S_m|].$$

With $B = A$, any choices of $x_1, \ldots, x_m = \pm 1$ have the effect of permuting the columns - the matrix $(x_i a_{ij})$ also has every column vector precisely once - so that $X = |X_1| + \ldots + |X_m|$ is a constant. Note that $E[X]$ is independent of B. In general, fixing $E[X] = \mu$, the minimal possible value for $\max X$ is achieved when X is the constant μ. Thus $F[B]$ is minimized with $B = A$.

13

Geometry

> Few people think more than two or three times a year. I have made an international reputation for myself by thinking once or twice a week.
> – George Bernard Shaw

Suppose we choose randomly n points $P_1, \ldots, P_n$ on the unit circle, according to a uniform distribution. What is the probability that the origin lies in the convex hull of these points? There is a surprisingly simple (yet clever) way to compute this probability. Let us first choose n random pairs of antipodal points $Q_1, Q_{n+1} = -Q_1$, $Q_2, Q_{n+2} = -Q_2, \ldots, Q_n, Q_{2n} = -Q_n$ according to a uniform distribution. Notice that with probability 1 these pairs are all distinct. Next we choose each P_i to be either Q_i or its antipodal $Q_{n+i} = -Q_i$, where each choice is equally likely. Clearly this corresponds to a random choice of the points P_i. The probability that the origin does *not* belong to the convex hull of the points P_i, given the (distinct) points Q_j, is precisely $\frac{x}{2^n}$, where x is the number of subsets of the points Q_j contained in an open half plane determined by a line through the origin, which does not pass through any of the points Q_j. It is easy to see that $x = 2n$. This is because if we renumber the points Q_j so that their cyclic order on the circle is $Q_1, \ldots, Q_n, Q_{n+1}, \ldots, Q_{2n}$, and $Q_{n+i} = -Q_i$ then the subsets contained in such half planes are precisely $\{Q_i, \ldots, Q_{n+i-1}\}$, where the indices are reduced modulo $2n$. Therefore, the probability that the origin is in the convex hull of n randomly chosen points on the unit circle is precisely $1 - \frac{2n}{2^n}$. Observe that the same result holds if we replace the unit circle by any centrally symmetric bounded planar domain with center 0 and that the argument can be easily generalized to higher dimensions.

This result, due to Wendel (1962) , shows how in some cases a clever idea can replace a tedious computation. It also demonstrates the connection between probability and geometry. The probabilistic method has been recently used extensively for deriving results in discrete and computational geometry. Some of these results are described in this chapter.

13.1 THE GREATEST ANGLE AMONG POINTS IN EUCLIDEAN SPACES

There are several striking examples, in different areas of Combinatorics, where the probabilistic method supplies very simple counter- examples to long-standing conjectures. Here is an example, due to Erdős and Z. Fűredi (1983) .

Theorem 13.1.1 *For every $d \geq 1$ there is a set of at least $\lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^d \rfloor$ points in the d-dimensional Euclidean space R^d, such that all angles determined by three points from the set are strictly less than $\pi/2$.*

This theorem disproves an old conjecture of Danzer and Grűnbaum (1962) , that the maximum cardinality of such a set is at most $2d-1$. We note that as proved by Danzer and Grűnbaum the maximum cardinality of a set of points in R^d in which all angles are at most $\pi/2$ is 2^d.

Proof.[Theorem 13.1.1] We select the points of a set X in R^d from the vertices of the d-dimensional cube. As usual, we view the vertices of the cube, which are $0,1$-vectors of length d, as the characteristic vectors of subsets of a d-element set; i.e., each $0,1$-vector a of length d is associated with the set $A = \{i : 1 \leq i \leq d, a_i = 1\}$. A simple consequence of Pythagoras' Theorem gives that the three vertices a, b and c of the d-cube, corresponding to the sets A, B and C, respectively, determine a right angle at c if and only if

$$A \cap B \subset C \subset A \cup B. \tag{13.1}$$

As the angles determined by triples of points of the d-cube are always at most $\pi/2$, it suffices to construct a set X of cardinality at least the one stated in the theorem no three distinct members of which satisfy (13.1).

Define $m = \lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^d \rfloor$, and choose, randomly and independently, $2m$ d-dimensional $0,1$-vectors $a_1, \ldots, a_{2m}$, where each coordinate of each of the vectors independently is chosen to be either 0 or 1 with equal probability. For every fixed triple a, b and c of the chosen points, the probability that the corresponding sets satisfy equation (13.1) is precisely $(3/4)^d$. This is because (13.1) simply means that for each i, $1 \leq i \leq d$, neither $a_i = b_i = 0, c_i = 1$ nor $a_i = b_i = 1, c_i = 0$ hold. Therefore, the probability that for three fixed indices i, j and k, our chosen points, a_i, a_j, a_k form a right angle at a_k is $(3/4)^d$. Since there are $\binom{2m}{3}3$ possible triples that can produce such angles, the expected number of right angles is

$$\binom{2m}{3} 3(3/4)^d \leq m,$$

where the last inequality follows from the choice of m. Thus there is a choice of a set X of $2m$ points in which the number of right angles is at most m. By deleting one point from each such angle we obtain a set of at least $2m - m = m$ points in which all angles are strictly less than $\pi/2$. Notice that the remaining points are all distinct since (13.1) is trivially satisfied if $A = C$. This completes the proof.

■

It is worth noting that, as observed by Erdős and Füredi, the proof above can be easily modified to give the following;

Theorem 13.1.2 *For every $\epsilon > 0$ there is a $\delta > 0$ such that for every $d \geq 1$ there is a set of at least $(1+\delta)^d$ points in R^d so that all the angles determined by three distinct points from the set are at most $\pi/3 + \epsilon$.*

We omit the detailed proof of this result.

13.2 EMPTY TRIANGLES DETERMINED BY POINTS IN THE PLANE

For a finite set X of points in general position in the plane, let $f(X)$ denote the number of *empty* triangles determined by triples of points of X, i.e., the number of triangles determined by points of X which contain no other point of X. Katchalski and Meir (1988) studied the minimum possible value of $f(X)$ for a set X of n points. Define $f(n) = min\{f(X)\}$, as X ranges over all planar sets of n points in general position (i.e., containing no three colinear points). They proved that

$$\binom{n-1}{2} \leq f(n) < 200n^2.$$

These bounds were improved by Bárány and Fűredi (1987) , who showed that as n grows

$$(1 + o(1))n^2 \leq f(n) \leq (1 + o(1))2n^2.$$

The construction that establishes the upper bound is probabilistic, and is given in the following theorem. See also Valtr (1995) for a slightly better result.

Theorem 13.2.1 *Let $I_1, I_2, \ldots, I_n$ be parallel unit intervals in the plane, where*

$$I_i = \{(x, y) : x = i, 0 \leq y \leq 1\}.$$

For each i let us choose a point p_i randomly and independently from I_i according to a uniform distribution. Let X be the set consisting of these n randomly chosen points. Then the expected number of empty triangles in X is at most $2n^2 + O(nlogn)$.

Clearly, with probability 1, X is a set of points in general position and hence the above theorem shows that $f(n) \leq 2n^2 + O(nlogn)$.

Proof. We first estimate the probability that the triangle determined by the points p_i, p_{i+a} and p_{i+k} is empty, for some fixed i, a and $k = a + b \geq 3$. Let $A = (i, x)$, $B = (i + a, y)$ and $C = (i + k, z)$ be the points p_i, p_{i+a} and p_{i+k}, respectively. Let m be the distance between B and the intersection point of the segment AC with the interval I_{i+a}. Since each of the points p_j for $i < j < i+k$ are chosen randomly according to a uniform distribution on I_j, it follows that the probability that the triangle determined by A, B and C is empty is precisely

$$(1-\frac{m}{a})(1-2\frac{m}{a})\dots(1-(a-1)\frac{m}{a})(1-(b-1)\frac{m}{b})\dots(1-\frac{m}{b})$$
$$\leq exp(-\frac{m}{a}-2\frac{m}{a}\dots-(a-1)\frac{m}{a}-(b-1)\frac{m}{b}\dots-\frac{m}{b})$$
$$= exp(-\binom{a}{2}\frac{m}{a}-\binom{b}{2}\frac{m}{b}) = exp(-(k-2)m/2).$$

For every fixed choice of A and C, when the point $p_{i+a} = B$ is chosen randomly, the probabilty that its distance m from the intersection of the segment AC with the interval I_{i+a} is at most d is clearly at most $2d$, for all $d \geq 0$. Therefore, the probability that the triangle determined by p_i, p_{i+a} and p_{i+k} is empty is at most

$$2\int_{m\geq 0} exp(-(k-2)m/2)dm = 4/(k-2).$$

It follows that the expected value of the total number of empty triangles is at most

$$n-2+\sum_{1\leq i\leq n-3}\ \sum_{3\leq k\leq n-i}\ \sum_{1\leq a\leq k-1} 4/(k-2)$$
$$= n-2+\sum_{3\leq k\leq n-1}(n-k)\frac{4(k-1)}{k-2}$$
$$= n-2+\sum_{3\leq k\leq n-1}(n-k)4/(k-2)+4\sum_{3\leq k\leq n-1}(n-k)$$
$$= 2n^2+O(nlogn).$$

This completes the proof. ■

The result above can be extended to higher dimensions by applying a similar probabilistic construction. A set X of n points in the d- dimensional Euclidean space is called *independent* if no $d+1$ of the points lie on a hyperplane. A simplex determined by $d+1$ of the points is called *empty* if it contains no other point of X. Let $f_d(X)$ denote the number of empty simplices of X, and define $f_d(n) = min f_d(X)$, as X ranges over all independent sets of n points in R^d. Katchalski and Meir (1988) showed that $f_d(n) \geq \binom{n-1}{d}$. The following theorem of Bárány and Fűredi shows that here again, a probabilistic construction gives a matching upper bound, up to a constant factor (which depends on the dimension). We omit the detailed proof.

Theorem 13.2.2 *There exists a constant $K = K(d)$, such that for every convex, bounded set $A \subset R^d$ with nonempty interior, if X is a random set of n points obtained by n random and independent choices of points of A picked with uniform distribution, then the expected number of empty simplices of X is at most $K\binom{n}{d}$.*

13.3 GEOMETRICAL REALIZATIONS OF SIGN MATRICES

Let $A = (a_{i,j})$ be an m by n matrix with $+1, -1$-entries. We say that A is *realizable* in R^d if there are m hyperplanes $H_1, \ldots, H_m$ in R^d passing through the origin and n points $P_1, \ldots, P_n$ in R^d, so that for all i and j, P_j lies in the positive side of H_i if $a_{i,j} = +1$, and in the negative side if $a_{i,j} = -1$. Let $d(A)$ denote the minimum dimension d such that A is realizable in R^d, and define $d(m,n) = max(d(A))$, as A ranges over all m by n matrices with $+1, -1$-entries. Since $d(m,n) = d(n,m)$ we can consider only the case $m \geq n$.

The problem of determining or estimating $d(m,n)$, and in particular $d(n,n)$, was raised by Paturi and Simon (1984) . This problem was motivated by an attempt to estimate the maximum possible "unbounded-error probabilistic communication complexity" of Boolean functions. Alon, Frankl and Rödl (1985) proved that as n grows $n/32 \leq d(n,n) \leq (\frac{1}{2} + o(1))n$. Both the upper and the lower bounds are proved by combining probabilistic arguments with certain other ideas. In the next theorem we prove the upper bound, which is probably closer to the truth.

Theorem 13.3.1 *For all $m \geq n$,*

$$d(m,n) \leq (n+1)/2 + \sqrt{\frac{n-1}{2} \log m}.$$

For the proof, we need a definition and two lemmas. For a vector $\mathbf{a} = (a_1, \ldots, a_n)$ of $+1, -1$-entries, the number of *sign-changes* in $\mathbf{a}$ is the number of indices i, $1 \leq i \leq n-1$ such that $a_i = -a_{i+1}$. For a matrix A of $+1, -1$-entries, denote by $s(A)$ the maximum number of sign-changes in a row of A.

Lemma 13.3.2 *For any matrix A of $+1, -1$-entries, $d(A) \leq s(A) + 1$.*

Proof. Let $A = (a_{i,j})$ be an m by n matrix of $+1, -1$ entries and suppose $s = s(A)$. Let $t_1 < t_2 < \ldots < t_n$ be arbitrary reals, and define n points $P_1, P_2, \ldots, P_n$ in R^{s+1} by: $P_j = (1, t_j, t_j^2, \ldots, t_j^s)$. These points, whose last s coordinates represent points on the d-dimensional moment-curve, will be the points used in the realization of A. To complete the proof we have to show that each row of A can be realized by a suitable hyperplane through the origin. This is proved by applying some of the known properties of the moment-curve as follows. Consider the sign-vector representing an arbitrary row of A. Suppose this vector has r sign changes, where, of course, $r \leq s$. Suppose the sign changes in this vector occur between the coordinates i_j and $i_j + 1$, for $1 \leq j \leq r$. Choose arbitrary reals $y_1, \ldots, y_r$, where $t_{i_j} < y_j < t_{i_j+1}$ for $1 \leq j \leq r$. Consider the polynomial $P(t) = \prod_{j=1}^{r}(t - y_j)$. Since its degree is at most s there are real numbers a_j such that $P(t) = \sum_{j=0}^{s} a_j t^j$. Let H be the hyperplane in R^{s+1} defined by $H = \{(x_0, x_1, \ldots, x_s) \in R^{s+1} : \sum_{j=0}^{s} a_j x_j = 0\}$. Clearly, the point $P_j = (1, t_j, \ldots, t_j^s)$ is on the positive side of this hyperplane if $P(t_j) > 0$, and is on its negative side if $P(t_j) < 0$. Since the polynomial P changes sign only in the values y_j, it follows that the hyperplane H separates the points

$P_1, \ldots, P_n$ according to the sign pattern of the corresponding row of A. Hence, by choosing the orientation of H appropriately, we conclude that A is realizable in R^{s+1}, completing the proof of the lemma. ■

Lemma 13.3.3 *For every m by n matrix A of $+1, -1$-entries there is a matrix B obtained from A by multiplying some of the columns of A by -1, such that $s(B) \leq (n-1)/2 + \sqrt{\frac{n-1}{2} \log m}$.*

Proof. For each column of A, randomly and independently, choose a number $\epsilon \in \{+1, -1\}$, where each of the two choices is equally likely, and multiply this column by ϵ. Let B be the random sign-matrix obtained in this way. Consider an arbitrary fixed row of B. One can easily check that the random variable describing the number of sign changes in this row is a binomial random variable with parameters $n-1$ and $p = 1/2$. This is because no matter what the entries of A in this row are, the row of B is a totally random row of $-1, 1$ entries. By the standard estimates for Binomial distributions, described in Appendix A, the probability that this number is greater than $(n-1)/2 + \sqrt{\frac{n-1}{2} logm}$ is smaller than $1/m$. Therefore, with positive probability the number of sign changes in each of the m rows is at most that large, completing the proof. ■

Proof.[Theorem 13.3.1] Let A be an arbitrary m by n matrix of $+1, -1$-entries. By Lemma 13.3.3 there is a matrix B obtained from A by replacing some of its columns by their inverses, such that $s(B) \leq (n-1)/2 + \sqrt{\frac{n-1}{2} \log m}$. Observe that $d(A) = d(B)$, since any realization of one of these matrices by points and hyperplanes through the origin gives a realization of the other one by replacing the points corresponding to the altered columns by their antipodal points. Therefore, by Lemma 13.3.2

$$d(A) = d(B) \leq s(B) + 1 \leq (n+1)/2 + \sqrt{\frac{n-1}{2} \log m}.$$

This completes the proof. ■

It is worth noting that by applying the (general) six standard deviations theorem stated in the end of Section 12.2, the estimate in Lemma 13.3.3 (and hence in Theorem 13.3.1) can be improved to $n/2 + O(\sqrt{nlog(m/n)})$. It can be also shown that if n and m grow so that m/n^2 tends to infinity and $(log_2 m)/n$ tends to 0 then for almost all m by n matrices A of $+1, -1$-entries $d(A) = (\frac{1}{2} + o(1))n$.

13.4 ϵ-NETS AND VC-DIMENSIONS OF RANGE SPACES

What is the minimum number $f = f(n, \epsilon)$ such that every set X of n points in the plane contains a subset S of at most f points such that every triangle containing at least ϵn points of X contains at least one point of S? As we shall see in this section, there is an absolute constant c such that $f(n, \epsilon) \leq \frac{c}{\epsilon} log(1/\epsilon)$, and this estimate holds for every n. This somewhat surprising result is a very special case of a general theorem of Vapnik and Chervonenkis (1971) , which has been extended by Haussler and Welzl (1987) , and which has many interesting applications in Computational Geometry and in Statistics. In order to describe this result we need a few definitions. A *range space* S is a pair (X, R), where X is a (finite or infinite) set and R is a (finite or infinite) family of subsets of X. The members of X are called *points* and those of R are called *ranges*. If A is a subset of X then $P_R(A) = \{r \cap A : r \in R\}$ is the *projection* of R on A. In case this projection contains all subsets of A we say that A is *shattered*. The *Vapnik-Chervonenkis dimension* (or *VC-dimension*) of S, denoted by $VC(S)$, is the maximum cardinality of a shattered subset of X. If there are arbitrarily large shattered subsets then $VC(S) = \infty$.

The number of ranges in any finite range space with a given number of points and a given VC-dimension cannot be too large. For integers $n \geq 0$ and $d \geq 0$, define a function $g(d, n)$ by

$$g(d, n) = \sum_{i=0}^{d} \binom{n}{i}.$$

Observe that for all $n, d \geq 1$, $g(d, n) = g(d, n-1) + g(d-1, n-1)$. The following combinatorial lemma was proved, independently, by Sauer (1972) , Perles and Shelah and, in a slightly weaker form by Vapnik and Chervonenkis.

Lemma 13.4.1 *If (X, R) is a range space of VC-dimension d with $|X| = n$ points then $|R| \leq g(d, n)$.*

Proof. We apply induction on $n + d$. The assertion is trivially true for $d = 0$ and $n = 0$. Assuming it holds for n and $d - 1$ and for $n - 1$ and $d - 1$ we prove it for n and d. Let $S = (X, R)$ be a range space of VC-dimension d on n points. Suppose $x \in X$, and consider the two range-spaces $S - x$ and $S \setminus x$ defined as follows. $S - x = (X - \{x\}, R - x)$, where $R - x = \{r - \{x\} : r \in R\}$. $S \setminus x = (X - \{x\}, R \setminus x)$, where $R \setminus x = \{r \in R : x \notin r, r \cup \{x\} \in R\}$. Clearly, the VC-dimension of $S - x$ is at most d. It is also easy to see that the VC-dimension of $S \setminus x$ is at most $d - 1$. Therefore, by the induction hypothesis,

$$|R| = |R - x| + |R \setminus x| \leq g(d, n - 1) + g(d - 1, n - 1) = g(d, n),$$

completing the proof. ∎

It is easy to check that the estimate given in the above lemma is sharp for all possible values of n and d. If (X, R) is a range space of VC-dimension d and $A \subset X$, then the VC-dimension of $(A, P_R(A))$ is clearly at most d. Therefore, the last lemma implies the following.

Corollary 13.4.2 *If (X, R) is a range space of VC-dimension d then for every finite subset A of X, $|P_R(A)| \leq g(d, |A|)$.*

There are many range spaces with finite VC-dimension that arise naturally in discrete and computational geometry. One such example is the space $S = (R^d, H)$, whose points are all the points in the d-dimensional Euclidean space, and whose set of ranges is the set of all (open) half-spaces. Any set of $d+1$ affinely independent points is shattered in this space, and, by Radon's Theorem, no set of $d+2$ points is shattered. Therefore $VC(S) = d+1$. As shown by Dudley (1978) , if (X, R) has finite VC-dimension, so does (X, R_k), where R_k is the set of all Boolean combinations formed from at most k ranges in R. In particular, the following statement is a simple consequence of Corollary 13.4.2.

Corollary 13.4.3 *Let (X, R) be a range space of VC-dimension $d \geq 2$, and let (X, R_h) be the range space on X in which $R_h = \{(r_1 \cap \ldots \cap r_h) : r_1, \ldots r_h \in R\}$. Then $VC(X, R_h) \leq 2dhlog(dh)$.*

Proof. Let A be an arbitrary subset of cardinality n of X. By Corollary 13.4.2 $|P_R(A)| \leq g(d,n) \leq n^d$. Since each member of $P_{R_h}(A)$ is an intersection of h members of $P_R(A)$ it follows that $|P_{R_h}(A)| \leq \binom{g(d,n)}{h} \leq n^{dh}$. Therefore, if $n^{dh} < 2^n$, then A cannot be shattered. But this inequality holds for $n \geq 2dhlog(dh)$, since $dh \geq 4$. ■

As shown above, the range space whose set of points is R^d, and whose set of ranges is the set of all half spaces has VC-dimension $d+1$. This and the last corollary imply that the range space (R^d, C_h), where C_h is the set of all convex d-polytopes with h facets has a VC-dimension which does not exceed $2(d+1)hlog((d+1)h)$.

An interesting property of range spaces with a finite VC-dimension is the fact that each finite subset of such a set contains relatively small good samples in the sense described below. Let (X, R) be a range space and let A be a finite subset of X. For $0 \leq \epsilon \leq 1$, a subet $B \subset A$ is an *ϵ-sample* for A if for any range $r \in R$ the inequality

$$||A \cap r|/|A| - |B \cap r|/|B|| \leq \epsilon$$

holds. Similarly, a subset $N \subset A$ is an *ϵ-net* for A if any range $r \in R$ satisfying $|r \cap A| > \epsilon A$ contains at least one point of N.

Notice that every ϵ-sample for A is also an ϵ-net, and that the converse is not true. However, both notions define subsets of A that represent approximately some of the behavior of A with respect to the ranges. Our objective is to show the existence of small ϵ-nets or ϵ-samples for finite sets in some range spaces. Observe that if (X, R) is a range space with an infinite VC-dimension then for every n there is a shattered subset A of X of cardinality n. It is obvious that any ϵ-net (and hence certainly any ϵ-sample) for such an A must contain at least $(1-\epsilon)n$ points, i.e., it must contain almost all points of A. Therefore, in infinite VC-dimension there are no small nets or samples. However, it turns out that in finite VC-dimension, there are always very small nets and samples. The following theorem was proved by Vapnik and Chervonenkis (1971) .

Theorem 13.4.4 *There is a positive constant c such that if (X, R) is any range-space of VC-dimension at most d, $A \subset X$ is a finite subset and $\epsilon, \delta > 0$, then a random subset B of cardinality s of A where s is at least the minimum between $|A|$ and*

$$\frac{c}{\epsilon^2}(dlog\frac{d}{\epsilon} + log\frac{1}{\delta})$$

is an ϵ-sample for A with probability at least $1 - \delta$.

Using similar ideas, Haussler and Welzl (1987) proved the following theorem.

Theorem 13.4.5 *Let (X, R) be a range space of VC-dimension d, let A be a finite subset of X and supppose $0 < \epsilon, \delta < 1$. Let N be a set obtained by m random independent draws from A, where*

$$m \geq max(\frac{4}{\epsilon}log\frac{2}{\delta}, \frac{8d}{\epsilon}log\frac{8d}{\epsilon}) \tag{13.2}$$

Then N is an ϵ-net for A with probability at least $1 - \delta$.

Therefore, if A is a finite subset of a range space of finite VC-dimension d, then for any $\epsilon > 0$, A contains ϵ-nets as well as ϵ-samples whose size is at most some function of ϵ and d, independent of the cardinality of A ! The result about the triangles mentioned in the first paragraph of this section thus follows from Theorem 13.4.5, together with the observation following Corollary 13.4.3 that implies that the range space whose ranges are all triangles in the plane has a finite VC-dimension. We note that as shown by Pach and Woeginger (1990) , there are cases in which for fixed δ, the dependence of m in $1/\epsilon$ cannot be linear, but there is no known natural geometric example demonstrating this phenomenon. See also Komlós, Pach and Woeginger (1992) for a tight form of the last theorem.

The proofs of Theorems 13.4.4 and 13.4.5 are very similar. Since the computation in the proof of Theorem 13.4.5 is simpler, we describe here only the proof of this theorem, and encourage the reader to try and make the required modifications that yield a proof for Theorem 13.4.4.

Proof.[Theorem 13.4.5] Let (X, R) be a range space with VC-dimension d, and let A be a subset of X of cardinality $|A| = n$. Suppose m satisfies (13.2), and let $N = (x_1, \ldots, x_m)$ be obtained by m independent random choices of elements of A. (The elements in N are not necessarily distinct, of course). Let E_1 be the following event:

$$E_1 = \{\exists r \in R : |r \cap A| > \epsilon n, r \cap N = \emptyset\}.$$

To complete the proof we must show that the probability of E_1 is at most δ. To this end , we make an additional random choice and define another event as follows. Independently of our previous choice, we let $T = (y_1, \ldots, y_m)$ be obtained by m independent random choices of elements of A. Let E_2 be the event defined by:

$$E_2 = \{\exists r \in R : |r \cap A| > \epsilon n, r \cap N = \emptyset, |r \cap T| \geq \frac{\epsilon m}{2}\}.$$

(Since the elements of T are not neccessarily distinct, the notation $|r \cap T|$ means here $|\{i : 1 \leq i \leq m, y_i \in r\}|$. The quantities $|r \cap N|$ and $|r \cap (N \cup T)|$ are similarly defined).

Claim 4 $Pr(E_2) \geq \frac{1}{2} Pr(E_1)$.

Proof. It suffices to prove that the conditional probability $Pr(E_2|E_1)$ is at least $1/2$. Suppose that the event E_1 occurs. Then there is an $r \in R$ such that $|r \cap A| > \epsilon n$ and $r \cap N = \emptyset$. The conditional probability above is clearly at least the probability that for this specific r, $|r \cap T| \geq \frac{\epsilon m}{2}$. However $|r \cap T|$ is a binomial random variable with expectation ϵm and variance $\epsilon(1-\epsilon)m \leq \epsilon m$, and hence, by Chebyschev's inequality

$$Pr(|r \cap T| < \frac{\epsilon m}{2}) \leq \frac{\epsilon m}{(\epsilon m/2)^2} = \frac{4}{\epsilon m} \leq 1/2,$$

where the last inequality follows from (13.2). Thus, the assertion of Claim 4 is correct. ■

Claim 5

$$Pr(E_2) \leq g(d, 2m)2^{-\frac{\epsilon m}{2}}.$$

Proof. The random choice of N and T can be described in the following way, which is equivalent to the previous one. First one chooses $N \cup T = (z_1, \ldots, z_{2m})$ by making $2m$ random independent choices of elements of A, and then one chooses randomly precisely m of the elements z_i to be the set N, (the remaining elements z_j form the set T, of course). For each range $r \in R$ satisfying $|r \cap A| > \epsilon n$, let E_r be the event that $|r \cap T| > \frac{\epsilon m}{2}$ and $r \cap N = \emptyset$. A crucial fact is that if $r, r' \in R$ are two ranges, $|r \cap A| > \epsilon n$ and $|r' \cap A| > \epsilon n$ and if $r \cap (N \cup T) = r' \cap (N \cup T)$, then the two events E_r and $E_{r'}$, when both are conditioned on the choice of $N \cup T$, are identical. This is because the occurence of E_r depends only on the intersection $r \cap (N \cup T)$. Therefore, for any fixed choice of $N \cup T$, the number of distinct events E_r does not exceed the number of different sets in the projection $P_{N \cup T}(R)$. Since the VC-dimension of X is d, Corollary 13.4.2 implies that this number does not exceed $g(d, 2m)$.

Let us now estimate the probabilty of a fixed event of the form E_r, given the choice of $N \cup T$. This probability is at most

$$Pr(r \cap N = \emptyset \,|\, |r \cap (N \cup T)| > \frac{\epsilon m}{2}).$$

Define $p = |r \cap (N \cup T)|$. Since the choice of N among the elements of $N \cup T$ is independent of the choice of $N \cup T$, the last conditional probability is precisely

$$\frac{(2m-p)(2m-p-1)\ldots(m-p+1)}{2m(2m-1)\ldots(m+1)}$$

$$= \frac{m(m-1)\ldots(m-p+1)}{2m(2m-1)\ldots(2m-p+1)} \leq 2^{-p} \leq 2^{-\frac{\epsilon m}{2}}.$$

Since there are at most $g(d, 2m)$ potential distinct events E_r, it follows that the probability that at least one of them occurs given the choice of $N \cup T$ is at most $g(d, 2m)2^{-\frac{\epsilon m}{2}}$. Since this estimate holds conditioned on every possible choice of $N \cup T$ it follows that the probability of the event E_2 is at most $g(d, 2m)2^{-\frac{\epsilon m}{2}}$. This establishes Claim 5. ■

By Claim 4 and Claim 5, $Pr(E_1) \leq 2g(d, 2m)2^{-\frac{\epsilon m}{2}}$. To complete the proof of the theorem it remains to show that if m satisfies inequality (13.2) then

$$2g(d, 2m)2^{-\frac{\epsilon m}{2}} \leq \delta.$$

We describe the proof for $d \geq 2$. The computation for $d = 1$ is easier. Since $g(d, 2m) \leq (2m)^d$ it suffices to show that

$$2(2m)^d \leq \delta 2^{\frac{\epsilon m}{2}},$$

i.e., that

$$\frac{\epsilon m}{2} \geq dlog(2m) + log\frac{2}{\delta}.$$

From (13.2) it follows that

$$\frac{\epsilon m}{4} \geq log\frac{2}{\delta},$$

and hence it suffices to show that

$$\frac{\epsilon m}{4} \geq dlog(2m).$$

The validity of the last inequality for some value of m implies its validity for any bigger m, and hence it suffices to check that it is satisfied for $m = \frac{8d}{\epsilon}log\frac{8d}{\epsilon}$, i.e., that

$$2dlog\frac{8d}{\epsilon} \geq dlog(\frac{16d}{\epsilon}log\frac{8d}{\epsilon}).$$

The last inequality is equivalent to $\frac{4d}{\epsilon} \geq log\frac{8d}{\epsilon}$, which is certainly true. This completes the proof of the theorem. ■

Theorems 13.4.4 and 13.4.5 have been used for constructing efficient data-structures for various problems in computational geometry. A trivial example is just the observation that Theorem 13.4.4 implies the following: for every $\epsilon > 0$ there is a constant $c = c(\epsilon)$ such that for every n and every set A of n points in the plane there is a data structure of size $c(\epsilon)$ that enables us to estimate , given any triangle in the plane, the number of points of A in this triangle up to an additive error of ϵn. This is done simply by storing the coordinates of a set of points that form an ϵ-sample for A considered as a subset of the range space whose ranges are all planar triangles. More sophisticated data structures whose construction relies on the above two theorems can be found in the paper of Haussler and Welzl (1987) .

13.5 DUAL SHATTER FUNCTIONS AND DISCREPANCY

The *dual shatter function* h of a range space $S = (X, R)$ is the function h mapping integers to integers, defined by letting $h(g)$ denote the maximum, over all possible choices of g members of R, of the number of atoms in the Venn diagram of these members. It is not too difficult to prove that if the VC-dimension of S is d, then

$h(g) \leq O(g^{2^{d+1}-1})$, but in geometric applications it is usually better to bound this function directly.

In Matoušek, Welzl and Wernisch (1993) it is proved that if the dual shatter function of a range space $S = (X, R)$ satisfies $h(g) \leq O(g^t)$, A is any set of n points in the range space, and $\mathcal{F}$ is the projection $P_R(A)$ of R on A, then, the discrepancy of $\mathcal{F}$ satisfies

$$disc(\mathcal{F}) \leq O(n^{\frac{1}{2}-\frac{1}{2t}}\sqrt{\log n}). \tag{13.3}$$

This supplies nontrivial estimates in various geometric situations, improving the trivial bound that follows from Theorem 12.1.1 of Chapter 12. In most of these geometric applications it is widely believed that the $\sqrt{\log n}$ factor can be omitted. In the abstract setting, however, this factor cannot be omitted, as proved in Matoušek (1997) (for $t = 2, 3$) and later in Alon, Rónyai and Szabó (1999) for all t.

The proof of (13.3) is based on a beautiful result of Chazelle and Welzl (1989) , and its improvement by Haussler (1995) . It is somewhat simpler to prove the result with an extra logarithmic factor, and this is the proof we present here. See Pach and Agarwal (1995) , for some additional information.

Let $\mathcal{F}$ be a family of subsets of a finite set A. In what follows we consider graphs whose edges are (unordered) pairs of points of A. For $F \in \mathcal{F}$ and $x, y \in A$, the edge xy *stabs* F if F contains exactly one of the two points x and y. The following theorem is proved in Chazelle and Welzl (1989) . An improvement by a logarithmic factor appears in Haussler (1995) .

Theorem 13.5.1 *Let $(A, \mathcal{F})$ be a finite range space, where $|A| = n$, and suppose that its dual shatter function h satisfies $h(g) \leq cg^t$ for some fixed $c, t > 0$. Then, there is a $C = C(c, t)$ and a Hamilton path on A, such that each member F of $\mathcal{F}$ is stabbed by at most $Cn^{1-1/t}\log n$ edges of the path.*

To prove the above theorem, we need the following lemma.

Lemma 13.5.2 *Let $(A, \mathcal{F}), n, h, t$ and c be as above, let B be a finite subset of $p > 1$ points of A, and let $\mathcal{G}$ be a collection of m (not necessarily distinct) members of $\mathcal{F}$. Then there are two distinct points x, y in B , such that the edge xy stabs at most $\frac{bm\log p}{p^{1/t}}$ members of $\mathcal{G}$, where $b = b(c)$.*

Proof. We may and will assume that p is larger than $c+1$. Let g be the largest integer such that $cg^t \leq p-1$, that is, $g = \lfloor (\frac{p-1}{c})^{1/t} \rfloor$. Let L be a random collection of g members of $\mathcal{G}$, each picked, randomly and independently (with possible repetitions), among all m members of $\mathcal{G}$ with uniform distribution. The Venn diagram of all members of L partition B into at most $h(g) \leq cg^t < p$ atoms, and hence there are two distinct points x, y of B that lie in the same atom. To complete the proof it suffices to show that with positive probability, for each pair of points of B that stabs more than $\frac{bm\log p}{p^{1/t}}$ members of $\mathcal{G}$, at least one of these members lies in L (and hence the pair does not lie in an atom of the corresponding Venn diagram.) There are $\binom{p}{2}$ such pairs, and for each of them, the probability that L contains no member of $\mathcal{G}$ it stabs is at most

$$(1 - \frac{b \log p}{p^{1/t}})^g \leq e^{-\frac{b \log p}{p^{1/t}} \lfloor (\frac{p-1}{c})^{1/t} \rfloor},$$

which is less than $1/p^2$ for an appropriately chosen constant $b = b(c)$. This completes the proof. ■

Proof.[Theorem 13.5.1] Note, first that if d is the VC-dimension of the given space then there is a shattered set D of size d. It is not difficult to see that there are $g = \lceil \log_2 d \rceil$ sets among those shattering D, so that no two points of D lie in the same atom of their Venn diagram. Therefore, $d \leq c(\lceil \log_2 d \rceil)^t$, implying that $d \leq 2^{c' t \log t}$, where $c' = c'(c)$. By Lemma 13.4.1 this implies that the total number of ranges in R is at most $n^{2^{c' t \log t}}$.

We next prove that there is a spanning tree of A satisfying the assertion of Theorem 13.5.1, and then show how to replace it by a Hamilton path. By Lemma 13.5.2 with $B_0 = A, p_0 = n$ and $\mathcal{G}_0 = \mathcal{F}, m_0 = |\mathcal{G}_0|$ ($\leq n^{2^{c' t \log t}}$), we conclude that there is a pair $x_0 y_0$ of points in A such that the edge $x_0 y_0$ does not stab more than $\frac{b \log n}{n^{1/t}} m_0$ members of $\mathcal{G}$. Let $\mathcal{G}_1$ be the collection obtained from $\mathcal{G}$ by duplicating all members of $\mathcal{G}$ that are stabbed by $x_0 y_0$, and define $B_1 = B - x_0, p_1 = n - 1$, $m_1 = |\mathcal{G}_1| \leq m_0(1 + \frac{b \log n}{n^{1/t}})$. Applying Lemma 13.5.2 again, this time to B_1 and $\mathcal{G}_1$, we obtain another pair $x_1 y_1$, define $B_2 = B_1 - x_1, p_2 = p_1 - 1 = n - 2$, and let $\mathcal{G}_2$ be the collection obtained from $\mathcal{G}_1$ by duplicating all members of $\mathcal{G}_1$ stabbed by $x_1 y_1$, $m_2 = |\mathcal{G}_2|$. By the assertion of the lemma, $m_2 \leq m_1(1 + \frac{b \log n}{(n-1)^{1/t}})$. Proceeding in this manner we get a sequence $x_0 y_0, x_1 y_1, \ldots, x_{n-1} y_{n-1}$ of edges of a graph on A, a sequence of subsets $B_0 = A, B_1, \ldots, B_{n-1}$, where each B_i is obtained from the previous one by omitting the point x_{i-1}, and a sequence of collections $\mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_{n-1}$, where

$$\begin{aligned} |\mathcal{G}_{n-1}| &\leq m_0 \textstyle\prod_{i=0}^{n-1}(1 + \frac{b \log n}{(n-i)^{1/t}}) \\ &\leq n^{2^{c' t \log t}} e^{b \log n \sum_{i=0}^{n-1} (n-i)^{-1/t}} \leq 2^{b' n^{1-1/t} \log n}, \end{aligned} \tag{13.4}$$

for an appropriate $b' = b'(c, t)$.

Note, now, that the edges $x_i y_i$ form a spanning tree on the set A. The crucial observation is the fact that if a member of $\mathcal{F}$ is stabbed by s of the edges, then it is being duplicated s times during the above process that generates $\mathcal{G}_{n-1}$, implying that $2^s \leq |\mathcal{G}_{n-1}|$ and hence that $s \leq b' n^{1-1/t} \log n$. It remains to replace the spanning tree by a Hamilton path. To do so, replace each edge of the tree by two parallel edges, and take an Euler tour in the resulting graph (in which all degrees are even). This is a sequence $x_0, x_1, x_2, \ldots, x_{2n-2} = x_0$ of points of A such that each adjacent pair of elements of the sequence is an edge of the tree, and each edge appears twice this way. The subsequence of the above one obtained by keeping only the first appearance of each point of A is a Hamilton path, and it is easy to check that each member of $\mathcal{F}$ is stabbed by at most $2b' n^{1-1/t} \log n$ of its edges, completing the proof. ■

The following result is a simple consequence of Theorem 13.5.1. As mentioned above, its assertion can be improved by a factor of $\sqrt{\log n}$.

Theorem 13.5.3 *Let $(A, \mathcal{F})$ be a finite range space, where $|A| = n$, and suppose that its dual shatter function h satisfies $h(g) \leq cg^t$ for some fixed $c, t > 0$. Then, there is a $C' = C'(c, t)$ such that the discrepancy of $\mathcal{F}$ satisfies*

$$disc(\mathcal{F}) \leq C' n^{\frac{1}{2} - \frac{1}{2t}} \log n.$$

Proof. Without loss of generality, assume that the number of points of A is even (otherwise, simply omit a point). By Theorem 13.5.1 there is a Hamilton path $x_1 x_2 \ldots x_n$ on these points such that each member of $\mathcal{F}$ is stabbed by at most $Cn^{1-1/t} \log n$ edges of the path. Let $f : A \mapsto \{-1, 1\}$ be a random coloring of A where for each i, $1 \leq i \leq n/2$, randomly and independently, either $f(x_{2i-1}) = 1, f(x_{2i}) = -1$ or $f(x_{2i-1}) = -1, f(x_{2i}) = 1$, the two choices being equally likely. Fix a member $F \in \mathcal{F}$, and note that the contribution of each pair $x_{2i-1}x_{2i}$ to the sum $\sum_{x_j \in F} f(x_j)$ is zero, if the edge $x_{2i-1}x_{2i}$ does not stab F, and is either $+1$ or -1 otherwise. It thus follows that this sum has, in the notation of Appendix A, Theorem A.1, the distribution S_r for some $r \leq Cn^{1-1/t} \log n$. Thus, the probability it is, in absolute value, at least α, can be bounded by $2e^{-\alpha^2/2r}$. As shown in the first paragraph of the proof of Theorem 13.5.1, the total number of members of $\mathcal{F}$ does not exceed $n^{2^{c' t \log t}}$, and thus the probability that there exists a member $F \in \mathcal{F}$ for which the sum $\sum_{x_j \in F} f(x_j)$ exceeds $C' n^{\frac{1}{2} - \frac{1}{2t}} \log n$ is less than 1 for an appropriately chosen constant $C' = C'(c, t)$. ■

The range space whose set of points is an arbitrary set of points in the plane, and whose ranges are all discs in the plane, has dual shatter function $O(g^2)$. The above theorem thus shows that it is possible to color any set of n points in the plane red and blue, such that the absolute value of the difference between the number of red points and the number of blue points inside any disc would not exceed $n^{1/4+o(1)}$. Similar results can be proved for many other geometric range spaces.

13.6 EXERCISES

1. Let A be a set of n points in the plane, and let $\mathcal{F}$ be the set of all intersections of A with an open triangle in the plane. Prove that the discrepancy of $\mathcal{F}$ does not exceed $n^{1/4+o(1)}$.

2. Prove that n distinct points in the plane determine at most $O(n^{4/3})$ unit distances.

THE PROBABILISTIC LENS:
Efficient Packing

Let $C \subset R^n$ be bounded with Riemann measure $\mu = \mu(C) > 0$. Let $N(C, x)$ denote the maximal number of disjoint translates of C that may be packed in a cube of side x and define the packing constant

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

the maximal proportion of space that may be packed by copies of C. The following result improves the one described in Chapter 3, Section 3.4.

Theorem 1 *Let C be bounded, convex and centrally symmetric about the origin. Then*

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

Proof. Fix $\epsilon > 0$. Normalize so $\mu = \mu(C) = 2 - \epsilon$. For any real z let C_z denote the "slab" of $(z_1, \ldots, z_{n-1}) \in R^{n-1}$ such that $(z_1, \ldots, z_{n-1}, z) \in C$ and let $\mu(C_z)$ be the usual $n-1$-dimensional measure of C_z. Riemann measurability implies

$$\lim_{\gamma \to 0} \sum_{m \in Z} \mu(C_{m\gamma})\gamma = \mu(C).$$

Let K be an integer sufficiently large so that

$$\sum_{m \in Z} \mu(C_{mK^{-(n-1)}}) K^{-(n-1)} < 2$$

and further that all points of C have all coordinates less than $K/2$.

For $1 \leq i \leq n-1$ let $v_i \in R^n$ be that vector with all coordinates zero except K as the i-th coordinate. Let

$$v = (z_1, \ldots, z_{n-1}, K^{-(n-1)}),$$

where $z_1, \ldots, z_{n-1}$ are chosen uniformly and independently from the real interval $[0, K)$. Let Λ_v denote the lattice generated by the v's - i.e.

$$\Lambda_v = \{m_1v_1 + \ldots + m_{n-1}v_{n-1} + mv : m_1, \ldots, m_{n-1}, m \in Z\}$$

$$= \{(mz_1 + m_1K, \ldots, mz_{n-1} + m_{n-1}K, mK^{-(n-1)} : m_1, \ldots, m_{n-1}, m \in Z\}.$$

Let $\theta(x)$ denote that unique $x' \in (-\frac{K}{2}, \frac{K}{2}]$ so that $x - mK = x'$ for some $m \in Z$. For $m \in Z$ let A_m be the event that some $m_1v_1 + \ldots + m_{n-1}v_{n-1} + mv \in C$. Since all coordinates of all points of C are less than $K/2$, A_m occurs if and only if

$$(\theta(mz_1), \ldots, \theta(mz_{n-1}), mK^{-(n-1)}) \in C,$$

which occurs if and only if $(\theta(mz_1), \ldots, \theta(mz_{n-1})) \in C_{mK^{-(n-1)}}$ The independence and uniformity of the z_i over $[0, K)$ implies the independence and uniformity of the $\theta(z_i)$ over $(-\frac{K}{2}, \frac{K}{2}]$ and so

$$\Pr[A_m] = K^{-(n-1)}\mu(C_{mK^{-(n-1)}}).$$

Summing over positive m, and employing the central symmetry,

$$\sum_{m>0} \Pr[A_m] < \frac{1}{2} \sum_{m \in Z} K^{-(n-1)}\mu(C_{mK^{-(n-1)}}) < \frac{1}{2}2 = 1.$$

Hence there *exists* v with all A_m, $m > 0$ not holding. By the central symmetry A_m and A_{-m} are the same event so no A_m, $m \neq 0$ holds. When $m = 0$ the points $m_1v_1 + \ldots + m_{n-1}v_{n-1} = K(m_1, \ldots, m_{n-1}, 0)$ all lie outside C except the origin. For this v

$$\Lambda_v \cap C = \{0\}.$$

Consider the set of translates $C + 2w, w \in \Lambda_v$. Suppose

$$z = c_1 + 2w_1 = c_2 + 2w_2 \text{ with } c_1, c_2 \in C, w_1, w_2 \in \Lambda_v.$$

Then $(c_1 - c_2)/2 = w_2 - w_1$. From convexity and central symmetry $(c_1 - c_2)/2 \in C$. As $w_2 - w_1 \in \Lambda_v$, it is zero and hence $c_1 = c_2$ and $w_1 = w_2$. That is, the translates form a packing of R^n. As $\det(2\Lambda_v) = 2^n \det(\Lambda_v) = 2^n$ this packing has density $2^{-n}\mu = 2^{-n}(2 - \epsilon)$. As $\epsilon > 0$ was arbitrary, $\delta(C) \geq 2^{-(n-1)}$. ■

14

Codes, Games and Entropy

Why did you come to Casablanca anyway, Rick?
  I came for the waters.
Waters, what waters? Casablanca is in the desert.
  I was misinformed.
Claude Rains to Humphrey Bogart in *Casablanca*

14.1 CODES

Suppose we want to send a message, here considered a string of bits, across a noisy channel. There is a probability p that any bit sent will be received incorrectly. The value p is a parameter of the channel and cannot be changed. We assume that p is both the probability that a sent zero is received as a one and that a sent one is received as a zero. Sent bits are always received, but perhaps incorrectly. We further assume that the events that the bits are received incorrectly are mutually independent. The case $p = .1$ will provide a typical example.

How can we improve the reliability of the system? One simple way is to send each bit three times. When the three bits are received we use majority rule to decode. The probability of incorrect decoding is then $3p^2 + p^3 = .031$ in our instance. We have sacrificed speed - the rate of transmission of this method is $1/3$ - and gained accuracy in return. If we send each bit five times and use majority rule to decode, the probability of incorrect decoding drops to .01051 but the rate of transmission also drops to $1/5$. Clearly we may make the probability of incorrect decoding as low as needed, but seemingly with the tradeoff that the rate of transmission tends to zero. It is the fundamental theorem of Information Theory - due to Claude Shannon - that this tradeoff is not necessary: there are codes with rate of transmission approaching a positive constant (dependent on p) with probability of incorrect transmission approaching zero.

A *Coding Scheme* consists of positive integers m, n, a function $f : \{0,1\}^m \to \{0,1\}^n$ called the encoding function and a function $g : \{0,1\}^n \to \{0,1\}^m$ called the decoding function. The notion is that a message (or segment of message) $x \in \{0,1\}^m$ will be encoded and sent as $f(x)$ and a received message $y \in \{0,1\}^n$ will be decoded as $g(y)$. The rate of transmission of such a scheme is defined as m/n. Let $E = (e_1, \ldots, e_n)$ be a random string defined by $\Pr[e_i = 1] = p$, $\Pr[e_i = 0] = 1-p$, the values e_i mutually independent. We define the probability of correct transmission as $\Pr[g(f(x)+E) = x]$. Here x is assumed to be uniformly distributed over $\{0,1\}^m$ and independent of E, $+$ is $\mathrm{mod}\, 2$ vector addition.

A crucial role is played by the *entropy function*

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

defined for $p \in (0,1)$. For any fixed p the entropy function appears in the asymptotic formula

$$\binom{n}{pn} = \frac{n^n e^{-n}}{(pn)^{pn} e^{-pn}((1-p)n)^{(1-p)n} e^{-(1-p)n}}(1+o(1))^n = 2^{n(H(p)+o(1))}.$$

For $p \in (0, .5)$ we further bound

$$\sum_{i \le pn} \binom{n}{i} \le (1+pn)\binom{n}{pn} = 2^{n(H(p)+o(1))}.$$

Theorem 14.1.1 [Shannon's Theorem] *Let $p \in (0, .5)$ be fixed. For $\epsilon > 0$ arbitrarily small there exists a Coding Scheme with rate of transmission greater than $1 - H(p) - \epsilon$ and probability of incorrect transmission less than ϵ.*

Proof. Let $\delta > 0$ be such that $p + \delta < .5$ and $H(p+\delta) < H(p) + \epsilon/2$. For n large set $m = n(1 - H(p) - \epsilon)$, guaranteeing the rate of transmission. Let $f : \{0,1\}^m \to \{0,1\}^n$ be a random function - each $f(x)$ uniformly and independently chosen. Given f define the decoding function $g : \{0,1\}^n \to \{0,1\}^m$ by setting $g(y) = x$ if x is the unique vector in $\{0,1\}^m$ within $n(p+\delta)$ of $f(x)$. We measure distance by the Hamming Metric ρ: $\rho(y, y')$ is the number of coordinates in which y, y' differ. If there is no such x, or more than one such x, then we shall consider decoding to be incorrect.

There are two ways decoding can be incorrect. Possibly $f(x) + E$ is not within $n(p+\delta)$ of $f(x)$. The distance from $f(x) + E$ to $f(x)$ is simply the number of ones in E which has Binomial Distribution $B(n, p)$ and so this occurs with probability $o(1)$, in fact, with exponentially small probability. The only other possibility is that there is some $x' \neq x$ with $f(x') \in S$ where S is the set of y' within $n(p+\delta)$ of $f(x)$. Conditioning on the values $f(x), E$, $f(x')$ is still uniformly distributed over

$\{0,1\}^n$ and hence this occurs with probability $|S|2^{-n}$ for any particular x' and thus with total probability at most

$$2^m|S|2^{-n} < 2^{-n(\frac{\epsilon}{2}+o(1))} = o(1).$$

The total probability for incorrect decoding from both sources is thus $o(1)$ and, in fact, exponentially small. For n sufficiently large this is less than ϵ.

The average over all choices of f, x of the probability of incorrect decoding is less than ϵ. Therefore there exists a specific f (hence a specific Coding Scheme) with probability of incorrect coding less than ϵ. ■

Shannon's Theorem, dealing with the intensely practical subject of Communications, puts the shortcomings of the probabilistic approach in sharp contrast. Where is the Coding Scheme? Supposing that a Coding Scheme may be found how can encoding and decoding be rapidly processed? A Group Code is a Coding Scheme in which the map $f : \{0,1\}^m \to \{0,1\}^n$ is linear, i.e. $f(0) = 0$ and $f(x+x') = f(x)+f(x')$, all calculations modulo 2. (Alternatively, the range of f is a subgroup of $\{0,1\}^n$.) Group Codes are of particular interest, in part because of the ease of encoding.

Theorem 14.1.2 *Let $p \in (0, .5)$ be fixed. For $\epsilon > 0$ arbitrarily small there exists a Group Code with rate of transmission greater than $1 - H(p) - \epsilon$ and probability of incorrect transmission less than ϵ.*

Proof. For $1 \le i \le m$ let $u_i \in \{0,1\}^m$ be that vector with a one in position i, all other entries zero. Let $f(u_1), \ldots, f(u_m)$ be chosen randomly and independently and then extend f by setting

$$f(\epsilon_1 u_1 + \ldots + \epsilon_m u_m) = \epsilon_1 f(u_1) + \ldots + \epsilon_m u_m.$$

We follow the proof of Shannon's Theorem until bounding the probability that $f(x)+E$ lies within $n(p+\delta)$ of $f(x)$. Set $z = x - x' = \epsilon_1 u_1 + \ldots + \epsilon_m u_m$, again all modulo 2. As $x \neq x'$, $z \neq 0$. Reorder for convenience so that $\epsilon_m = 1$. By linearity $f(z) = f(x) - f(x')$ so we bound $\Pr[f(z) \in S]$ where S is the set of vectors within $n(p+\delta)$ of E. Fixing E and the $f(u_i)$, $i < m$, $f(z)$ still has an additive term $f(u_m)$ which is uniform and independent. Hence $f(z)$ is distributed uniformly. Thus $\Pr[f(z) \in S] = |S|2^{-n}$ and the remainder of the proof is as in Shannon's Theorem. ■

14.2 LIAR GAME

Paul is trying to find a number $x \in \{1, \ldots, n\}$ from a recalcitrant and mendacious Carole. He may ask q questions of the form "Is $x \in S$?," where S can be any subset of the possibilities. The questions are asked sequentially and Paul's choice of his i-th question can depend on previous responses. Carole is allowed to lie - but she can lie at most k times. For which n, q, k can Paul determine the number?

When $k = 0$ Paul can win exactly when $n \leq 2^q$. The values $n = 100$, $q = 10$, $k = 1$ make for an amusing parlor game. Carole is hardly a passive observer, she may play an adversary strategy. By that we mean that she does not select an x in advance but answers consistently with at least one x. At the end of the game if her answers were consistent with more than one x then she has won. The game, called the (n, q, k)-Liar Game, is now a perfect information game with no hidden move and no draws. Hence either Paul or Carole has a perfect winning strategy. But who?

We describe an equivalent game, the ChipLiar game. There is a board with positions $0, 1, \ldots, k$. There are n chips labeled $1, \ldots, n$ which are initially at position k. There are q rounds. On each round Paul selects a set S of the chips. Carole can *either* move every chip not in S one position to the left *or* move every chip in S one position to the left. (Here position $i - 1$ is one position to the left of position i. Chips moved one position to the left from position 0 are removed from the board.) At the end of the q rounds Carole wins if there is more than one chip remaining on the board and Paul wins if there is one or zero chip remaining on the board. Basically, chip i at position j represents that the answer $x = i$ has already received $k - j$ lies; Paul selecting S represents his asking if $x \in S$; Carole moving the chips not in S represents a Yes answer, moving the chips in S represents a No answer. (In the ChipLiar game Carole can remove all chips from the board while in the Liar game Carole must play consistently with at least one x. But when Carole removes all chips from the board she automatically has lost and hence this difference does not affect the determination of the winner.)

In the ChipLiar game there is no reason to place all chips at position k at the start. More generally, for $x_0, \ldots, x_k \geq 0$, we define the $(x_0, \ldots, x_k), q$-ChipLiar Game to be the above q round game with initial position consisting of x_i chips at position i. This, in turn, corresponds to a Liar Game in which there are x_i possibilities for which Carole is constrained to lie at most i times.

Let us define $B(q, j)$ as the probability that in q flips of a fair coin there are at most j heads. Of course, we have the exact formula:

$$B(q, j) = 2^{-q} \sum_{i=0}^{j} \binom{q}{i}.$$

Theorem 14.2.1 *If*

$$\sum_{i=0}^{k} x_i B(q, i) > 1$$

then Carole wins the $(x_0, \ldots, x_k), q$-ChipLiar Game.

Corollary 14.2.2 *If*

$$n > \frac{2^q}{\sum_{i=0}^{k} \binom{q}{i}}$$

then Carole wins the (n, q, k)-Liar Game.

Proof.[Theorem 14.2.1] Fix a strategy for Paul. Now Carole plays randomly! That is, at each round, after Paul has selected a set S of chips Carole flips a coin - if it comes up heads she moves every chip not in S one position to the left and if it comes up tails she moves every chip in S one position to the left. For each chip c let I_c be the indicator random variable for c remaining on the board at the end of the game. Set $X = \sum I_c$, the number of chips remaining on the board at the end of the game. Consider a single chip c. Each round Paul may have chosen $c \in S$ or $c \notin S$ but in either case c is moved to the left with probability one half. Suppose c starts at position i. It remains on the board at the end of the game if and only if in the q rounds it has been moved to the left at most i times. Then $E[I_c]$, the probability of this occuring, is precisely $B(q, i)$. By linearity of expectation $E[X] = \sum_{i=0}^{k} x_i B(q, i)$. The assumption of the theorem gives $E[X] > 1$. But then $X > 1$ must occur with positive probability. That is, Carole must win with positive probability.

No strategy of Paul allows him to always win. But this is a perfect information game with no draws so someone has a perfect strategy that always wins. That someone isn't Paul, so it must be Carole. ■

The above proof certainly illustrated the magical element of the probabilistic method. Carole has a winning strategy but what is it? The general notion of moving from a probabilistic existence proof to an explicit construction is called derandomization and will be dealt with in detail in the next chapter. Here we can give an explicit strategy. With l moves remaining in the game and y_i chips on position i define the weight of the position as $\sum_i y_i B(l, i)$ – note this is $E[Y]$ where Y is the number of chips that would remain on the board should Carole play the rest of the game randomly. Carole's explicit strategy is to always move so as to maximize the weight.

Consider any position with weight W and any move S by Paul. Let W^y, W^n be the new weights should Carole move all chips not in S or all chips in S respectively. We claim $W = \frac{1}{2}(W^y + W^n)$. One argument is that by linearity this identity reduces to the case of one chip and it then follows from the identity $B(l, j) = \frac{1}{2}(B(l-1, j) + B(l-1, j-1))$. But we needn't actually do any calculation. Carole's playing randomly can be thought of as first flipping a coin to decide on her first move and then playing randomly so that $E[Y]$ is the average of the two conditional expectations.

At the start of the game, by assumption, the weight is bigger than one. Carole's explicit strategy assures that the weight does not decrease so at the end of the game the weight is bigger than one. But at the end of the game the weight is the number of chips remaining. Being bigger than one Carole has won the game.

The converse of the Theorem, and even the Corollary, is false. Consider the LiarGame with $n = 5$, $q = 5$ questions and $k = 1$ possible lie. In the ChipLiar version this is the $(0, 5)$, 5-ChipLiar game. Here $B(5, 1) = 6/32$ and $5(6/32) \leq 1$. Still, Carole wins with perfect play. The problem is that Paul has no good first move. Suppose he selects two chips as S (asks "Is $x \leq 2$?" in the LiarGame). Then Carole moves the two chips one to the left (responds Yes) leaving the position $(2, 3)$ with four questions remaining. As $2B(4, 0) + 3B(4, 1) = 17/16 > 1$ Carole will now win. It is easy to check that all other moves of Paul fail. The difficulty here is that Paul was in a position with weight $W \leq 1$ but was unable to find a move such that $W^y \leq 1$ and $W^n \leq 1$.

14.3 TENURE GAME

Paul, Chair of Department, is trying to promote one of his faculty to tenure but standing in his way is a recalcitrant and meanspirited Carole, the Provost. There are k pre-tenure levels, labeled $1, \ldots, k$, level 1 the highest, and a level 0, representing tenure. For our purposes each faculty member is represented by a chip. The $(x_1, \ldots, x_k)$-Tenure Game begins with x_i chips at level i for $1 \leq i \leq k$ and no chips on level zero. Each year Paul presents a set S of chips to Carole. Carole may either

• Promote all chips in S and fire the others *or*

• Promote all chips not in S and fire those in S.

Promote, as used above, means to move from level i to level $i - 1$. Fired means just that, removing the chip from the game. If a chip reaches level 0 then Paul is the winner. The draconian promotion or perish provision insures that the game will end within k years with either Paul winning or Carole having successfully eliminated all chips.

Theorem 14.3.1 *If $\sum_i x_i 2^{-i} < 1$ then Carole wins the $(x_1, \ldots, x_k)$-Tenure Game.*

Proof. Fix a strategy for Paul. Now Carole plays randomly! That is, at each round, after Paul has selected a set S of chips Carole flips a coin - if it comes up heads she moves every chip not in S one position to the left and if it comes up tails she moves every chip in S one position to the left. For each chip c let I_c be the indicator random variable for c reaching level 0. Set $X = \sum I_c$, the number of chips reaching level 0 at the end of the game. Consider a single chip c. Each round Paul may have chosen $c \in S$ or $c \notin S$ but in either case c is moved to the left with probability one half. Suppose c starts at position i. It remains on the board at the end of the game if and only if the first i coinflips of Carole led to promotions for c. Then $E[I_c]$, the probability of this occuring, is precisely 2^{-i}. By linearity of expectation $E[X] = \sum_{i=1}^{k} x_i 2^{-i}$. The assumption of the theorem gives $E[X] < 1$. But then $X < 1$ must occur with positive probability. That is, Carole must win with positive probability.

No strategy of Paul allows him to always win. But this is a perfect information game with no draws so someone has a perfect strategy that always wins. That someone isn't Paul, so it must be Carole. ■

As with the LiarGame we may derandomize the above argument to give an explicit strategy for Carole. With y_i chips on position i define the weight of the position as $\sum_i y_i 2^{-i}$ – note this is $E[Y]$ where Y is the number of chips that would reach level 0 should Carole play the rest of the game randomly. Carole's explicit strategy is to always move so as to minimize the weight. Consider any position with weight W and any move S by Paul. Let W^y, W^n be the new weights should Carole move all chips not in S or all chips in S respectively. As in the LiarGame $W = \frac{1}{2}(W^y + W^n)$. At the start of the game, by assumption, the weight is less than one. Carole's explicit strategy assures that the weight does not increase so at all times the weight is smaller than one. A chip at level 0 would add one to the weight by itself so that this never occurs and hence Carole wins.

In the LiarGame the sufficient condition for Carole winning was not neccessary because Paul did not always have an appropriately splitting move. Here, however, we have an amusing lemma.

Lemma 14.3.2 *If a set of chips has weight at least one it may be split into two parts each of weight at least one half.*

Proof. There must be two chips at some position i, otherwise the weight is less than one. If there are two chips at position 1 simply split them. If there are two chips at position $i > 1$ glue them together, and consider them as one superchip at position $i - 1$. Then the proof follows by induction on the number of chips. ■

Theorem 14.3.3 *If* $\sum x_i 2^{-i} \geq 1$ *then Paul wins the* $(x_1, \ldots, x_k)$*-Tenure Game.*

Proof. The initial weight is at least one. Applying the Lemma Paul splits the chips into two parts each of weight at least one half and sets S equal one of the parts. Carole moves all chips in one part one position to the left, doubling their weight, leaving a new position of weight at least one. Thus the weight never goes below one. Therefore the game cannot end with all chips having been removed (which would have weight zero) and so it must end with a win for Paul. ■

14.4 BALANCING VECTOR GAME

The balancing vector game is a perfect information game with two players, Pusher and Chooser. There is a parameter $n \geq 1$, and we shall be concerned with asymptotics in n. There are n rounds, each involving vectors in R^n. There is a position vector $P \in R^n$, initially set at 0. Each round has two parts. First Pusher picks $v \in \{-1, +1\}^n$. Then Chooser either resets P to $P + v$ or to $P - v$. At the end of the n-th round the payoff to Pusher is $|P|_\infty$, the maximal absolute value of the coordinates of P. Let $VAL(n)$ denote the value of this game to Pusher, and let S_n denote, as usual, the sum of n independent uniform $\{1, -1\}$ random variables.

Theorem 14.4.1 *If* $\Pr[|S_n| > \alpha] < n^{-1}$ *then* $VAL(n) \leq \alpha$.

Proof. Consider the game a win for Pusher if the final $|P|_\infty > \alpha$. Suppose Chooser announces that she will flip a fair coin each round to determine whether to reset P as $P + v$ or $P - v$. Let x_i be the i-th coordinate for the final value of the position vector P. Let W_i be the event $|x_i| > \alpha$ and $W = \vee W_i$ so that W is the event of Pusher winning. Regardless of Pusher's strategy x_i has distribution S_n so that

$$\Pr[W] \leq \sum_{i=1}^{n} \Pr[|S_n| > \alpha] < 1.$$

Pusher cannot always win so Chooser always wins. ■

Corollary 14.4.2 $VAL(n) = O(\sqrt{n \ln n})$.

To give a lower bound on $VAL(n)$ one wants to find a strategy for Pusher that wins against any Chooser. It is not sufficient to find a strategy that does well against a randomly playing Chooser - the Chooser is an adversary. Still, the notion of a randomly playing Chooser motivates the following result.

Theorem 14.4.3 *If* $\Pr[|S_n| > \alpha] > cn^{-1/2}$*, c an absolute constant, then* $VAL(n) > \alpha$.

Corollary 14.4.4 $VAL(n) = \Omega(\sqrt{n \ln n})$ *and hence* $VAL(n) = \Theta(\sqrt{n \ln n})$.

Proof.[Theorem 14.4.3] Define, for $x \in Z$, $0 \le i \le n$,

$$w_i(x) = \Pr[|x + S_{n-i}| > \alpha].$$

For $P = (x_1, \ldots, x_n)$ set $w_i(P) = \sum_{1 \le j \le n} w_i(x_j)$. When P is the position vector at the end of the i-th round, $w_i(P)$ may be interpreted as the expected number of coordinates with absolute value greater than α at the end of the game, assuming random play by Chooser. At the beginning of the game $w_0(P) = w_0(0) > c\sqrt{n}$ by assumption. Given position P at the end of round i Pusher's strategy will be to select $v \in \{-1, +1\}^n$ so that $w_{i+1}(P - v)$ and $w_{i+1}(P + v)$ are close together.

The distribution $x + S_{n-i}$ splits into $x + 1 + S_{n-i-1}$ and $x - 1 + S_{n-i-1}$ depending on the first coin flip so that for any i, x

$$w_i(x) = \frac{1}{2}[w_{i+1}(x+1) + w_{i+1}(x-1)].$$

Set $P = (x_1, \ldots, x_n)$, $v = (v_1, \ldots, v_n)$. For $1 \le j \le n$ set

$$\Delta_j = w_{i+1}(x_j + 1) - w_{i+1}(x_j - 1)$$

so that

$$w_{i+1}(P + v) - w_{i+1}(P - v) = \sum_{j=1}^{n} v_j \Delta_j,$$

and, for $\epsilon = \pm 1$

$$w_{i+1}(P + \epsilon v) = w_i(P) + \frac{1}{2}\epsilon \sum_{j=1}^{n} v_j \Delta_j.$$

Now we bound $|\Delta_j|$. Observe that

$$\Delta_j = \Pr[S_{n-i-1} = y] - \Pr[S_{n-i-1} = z],$$

where y is the unique integer of the same parity as $n - i - 1$ in the interval $(\alpha - (x_j + 1), \alpha - (x_j - 1)]$ and z the same in $[-\alpha - (x_j + 1), -\alpha - (x_j - 1))$. Let us set

$$g(m) = \max_s \Pr[S_m = s] = \binom{m}{\lfloor m/2 \rfloor} 2^{-m} \sim \sqrt{\frac{2}{\pi m}}$$

so that $|\Delta_j| \leq g(n-i-1)$ for all j.

A simple strategy for Pusher is then to reorder the coordinates so that $|\Delta_1| \geq \ldots \geq |\Delta_n|$ and then select $v_1, \ldots, v_n \in \{-1, +1\}$ sequentially, giving $v_i\Delta_i$ the opposite sign of $v_1\Delta_1 + \ldots + v_{i-1}\Delta_{i-1}$. (When $i = 1$ or the sum is zero choose v_i arbitrarily.) This assures

$$|v_1\Delta_1 + \ldots + v_n\Delta_n| \leq |\Delta_1| \leq g(n-i-1).$$

Let P^i denote the position vector at the end of the i-th round and v Pusher's choice for the $i+1$-st round. Then regardless of Chooser's choice of $\epsilon = \pm 1$

$$w_{i+1}(P^{i+1}) = w_{i+1}(P^i + \epsilon v) \geq w_i(P^i) - \frac{1}{2}|\sum_{j=1}^{n} v_j\Delta_j| \geq w_i(P^i) - \frac{1}{2}g(n-i-1).$$

Thus

$$w_n(P^n) \geq w_0(P^0) - \frac{1}{2}\sum_{i=0}^{n-1} g(n-i-1).$$

Simple asymptotics give that the above sum is asymptotic to $(8n/\pi)^{1/2}$. Choosing $c > (2/\pi)^{1/2}$, $w_n(P^n) > 0$. But $w_n(P^n)$ is simply the *number* of coordinates with absolute value greater than α in the final $P = P^n$. This Pusher strategy assures there is more than zero, hence at least one, such coordinate and therefore Pusher wins ■

14.5 NONADAPTIVE ALGORITHMS

Let us modify the balancing game of §14.4 by requiring the vectors selected by Pusher to have coordinates zero and one rather than plus and minus one. Let $VAL^*(n)$ denote the value of the modified game. One can use the bounds on $VAL(n)$ to show $VAL^*(n) = \Theta(\sqrt{n \ln n})$.

In Chapter 12 we showed that any family of n sets $S_1, \ldots, S_n$ on n points $1, \ldots, n$ has discrepancy $O(\sqrt{n})$, i.e., there is a coloring $\chi : \{1, \ldots, n\} \to \{-1, +1\}$ so that all $|\chi(S_i)| \leq c\sqrt{n}$. The proof of this result does not yield an effective algorithm for finding such a coloring and indeed it is not known if there is a polynomial time algorithm to do so. Suppose one asks for a *nonadaptive* or *on-line* algorithm in the following sense. Instead of being presented the entire data of $S_1, \ldots, S_n$ at once one is presented with the points sequentially. At the j-th "round" the algorithm looks at point j - more specifically, at which sets S_i contain j or, equivalently, at the j-th column of the incidence matrix. At that stage the algorithm must decide how to color j and, once colored, the coloring cannot be changed. How small can we assure $\max|\chi(S_i)|$ with such an algorithm. We may think of the points as being presented by an adversary. Thinking of the points as their associated column vectors, Pusher as the Worst Case adversary and Chooser as the algorithm, the best such an algorithm can do is precisely $VAL^*(n)$.

The requirement that an algorithm be nonadaptive is both stronger and weaker than the requirement that an algorithm takes polynomial time. Still, this lends support to the conjecture that there is no polynomial time algorithm for finding a coloring with all $|\chi(S_i)| \le c\sqrt{n}$.

14.6 ENTROPY

Let X be a random variable taking values in some range S, and let $P(X = x)$ denote the probability that the value of X is x. The *binary entropy* of X, denoted by $H(X)$ is defined by

$$H(X) = \sum_{x \in S} P(X = x) \log_2 \left(\frac{1}{P(X = x)}\right).$$

If Y is another random variable, taking values in T, and (X, Y) is the random variable taking values in $S \times T$ according to the joint distribution of X and Y, then the *conditional entropy* of X given Y is

$$H(X|Y) = H(X, Y) - H(Y).$$

In this section we prove some simple properties of the entropy function, and describe several surprising combinatorial and geometric applications. Intuitively, the entropy of a random variable measures the amount of information it encodes. This provides an intuitive explanation to the four parts of the next simple lemma. The formal proof, given below, uses the properties of the functions $\log z$ and $z \log z$, where here, and in the rest of this section, all logarithms are in base 2.

Lemma 14.6.1 *Let X, Y and Z be three random variables taking values in S, T and U, respectively. Then*

1. $H(X) \le \log_2 |S|$.
2. $H(X, Y) \ge H(X)$.
3. $H(X, Y) \le H(X) + H(Y)$.
4. $H(X|Y, Z) \le H(X|Y)$.

Proof.

1. Since the function $\log z$ is concave it follows, by Jensen's Inequality, that

$$H(X) = \textstyle\sum_{i \in S} P(X = i) \log\left(\frac{1}{P(X=i)}\right)$$

$$\le \log\left(\textstyle\sum_{i \in S} P(X = i) \frac{1}{P(X=i)}\right) = \log |S|.$$

2. By the monotonicity of $\log z$ for all $z > 0$,

$$H(X, Y) = \textstyle\sum_{i \in S} \sum_{j \in T} P(X = i, Y = j) \log\left(\frac{1}{P(X=i, Y=j)}\right)$$

$$\begin{aligned} &\geq \textstyle\sum_{i\in S}\sum_{j\in T} P(X=i,Y=j)\log(\frac{1}{P(X=i)}) \\ = &\textstyle\sum_{i\in S} P(X=i)\log(\frac{1}{P(X=i)}) = H(X). \end{aligned} \tag{14.1}$$

3. By definition

$$\begin{aligned} &H(X)+H(Y)-H(X,Y) = \\ &\textstyle\sum_{i\in S}\sum_{j\in T} P(X=i,Y=j)\log(\frac{P(X=i,Y=j)}{P(X=i)P(Y=j)}) \\ = &\textstyle\sum_{i\in S}\sum_{j\in T} P(X=i)P(Y=j)f(z_{ij}), \end{aligned} \tag{14.2}$$

where $f(z) = z\log z$ and $z_{ij} = \frac{P(X=i,Y=j)}{P(X=i)P(Y=j)}$. Since $f(z)$ is convex it follows, by Jensen's Inequality, that the last quantity is at least

$$f(\sum_{i\in S}\sum_{j\in T} P(X=i)P(Y=j)z_{ij}) = f(1) = 0.$$

4. Note that

$$H(X|Y) = H(X,Y) - H(Y) = \sum_{i\in S}\sum_{j\in T} P(X=i,Y=j)\log(\frac{P(Y=j)}{P(X=i,Y=j)}).$$

Similarly

$$H(X|Y,Z) = \sum_{i\in S}\sum_{j\in T}\sum_{k\in U} P(X=i,Y=j,Z=k)\log(\frac{P(Y=j,Z=k)}{P(X=i,Y=j,Z=k)}).$$

Therefore,

$$\begin{aligned} &H(X|Y) - H(X|Y,Z) \\ &= \sum_{i\in S}\sum_{j\in T}\sum_{k\in U} P(X=i,Y=j,Z=k)\log(\frac{P(Y=j)P(X=i,Y=j,Z=k)}{P(X=i,Y=j)P(Y=j,Z=k)}) \\ &= \sum_{i\in S}\sum_{j\in T}\sum_{k\in U} \frac{P(X=i,Y=j)P(Y=j,Z=k)}{P(Y=j)} f(z_{ijk}), \end{aligned}$$

where $f(z) = z\log z$ and

$$z_{ijk} = \frac{P(Y=j)P(X=i,Y=j,Z=k)}{P(X=i,Y=j)P(Y=j,Z=k)}.$$

By the convexity of $f(z)$ (and since

$$\sum_{i\in S}\sum_{j\in T}\sum_{k\in U} \frac{P(X=i,Y=j)P(Y=j,Z=k)}{P(Y=j)} = 1 \ \ ),$$

it follows that the above quantity is at least

$$f(\sum_{i\in S}\sum_{j\in T}\sum_{k\in U} \frac{P(X=i,Y=j)P(Y=j,Z=k)}{P(Y=j)} z_{ijk}) = f(1) = 0.$$

■

The following simple but useful fact that the entropy is sub-additive has already been applied in Chapter 12 , Section 12.2.

Proposition 14.6.2 *Let $X = (X_1, \dots, X_n)$ be a random variable taking values in the set $S = S_1 \times S_2 \times \dots \times S_n$, where each of the coordinates X_i of X is a random variable taking values in S_i. Then*

$$H(X) \leq \sum_{i=1}^{n} H(X_i).$$

Proof. This follows by induction from Lemma 14.6.1, part 3. ■

The above proposition is used in Kleitman, Shearer and Sturtevant (1981) to derive several interesting applications in Extremal Finite Set Theory, including an upper estimate for the maximum possible cardinality of a family of k-sets in which the intersection of no two is contained in a third. The basic idea in Kleitman et al. (1981) can be illustrated by the following very simple corollary of the last proposition.

Corollary 14.6.3 *Let $\mathcal{F}$ be a family of subsets of $\{1, 2, \dots, n\}$ and let p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then*

$$|\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)},$$

where $H(y) = -y \log_2 y - (1-y)\log_2(1-y)$.

Proof. Associate each set $F \in \mathcal{F}$ with its characteristic vector $v(F)$, which is a binary vector of length n. Let $X = (X_1, \dots, X_n)$ be the random variable taking values in $\{0,1\}^n$, where $P(X = v(F)) = 1/|\mathcal{F}|$ for all $F \in \mathcal{F}$. Clearly $H(X) = |\mathcal{F}|(\frac{1}{|\mathcal{F}|} \log |\mathcal{F}|) = \log |\mathcal{F}|$, and since here $H(X_i) = H(p_i)$ for all $1 \leq i \leq n$, the result follows from Proposition 14.6.2. ■

The following interesting extension of Proposition 14.6.2 has been proved by Shearer, see Chung, Frankl, Graham and Shearer (1986) . As in that proposition, let $X = (X_1, \dots, X_n)$ be a random variable taking values in the set $S = S_1 \times S_2 \times \dots \times S_n$, where each X_i is a random variable taking values in S_i. For a subset I of $\{1, 2, \dots, n\}$, let $X(I)$ denote the random variable $(X_i)_{i \in I}$.

Proposition 14.6.4 *Let $X = (X_1, \dots, X_n)$ and S be as above. If $\mathcal{G}$ is a family of subsets of $\{1, \dots, n\}$ and each $i \in \{1, \dots, n\}$ belongs to at least k members of $\mathcal{G}$ then*

$$kH(X) \leq \sum_{G \in \mathcal{G}} H(X(G)).$$

Proof. We apply induction on k. For $k = 1$, replace each set $G \in \mathcal{G}$ by a subset of it to obtain a family $\mathcal{G}'$ whose members form a partition of $\{1, \dots, n\}$. By Lemma

14.6.1, part 2, $\sum_{G \in \mathcal{G}} H(X(G)) \geq \sum_{G' \in \mathcal{G}'} H(X(G'))$, and by Lemma 14.6.1, part 3, $\sum_{G' \in \mathcal{G}'} H(X(G')) \geq H(X)$, supplying the desired result for $k = 1$.

Assuming the result holds for $k - 1$, we prove it for k (≥ 2). If there is a $G \in \mathcal{G}$ with $G = \{1, \ldots, n\}$, the result follows from the induction hypothesis. Otherwise, let G_1, G_2 be two members of $\mathcal{G}$. By applying part 4 of Lemma 14.6.1 we conclude that

$$H(X(G_1 \setminus G_2)|X(G_1 \cap G_2), X(G_2 \setminus G_1)) \leq H(X(G_1 \setminus G_2)|X(G_1 \cap G_2)),$$

implying that

$$H(X(G_1 \cup G_2)) - H(X(G_2)) \leq H(X(G_1)) - H(X(G_1 \cap G_2)).$$

Therefore, $H((X(G_1 \cup G_2)) + H(X(G_1 \cap G_2)) \leq H(X(G_1)) + H(X(G_2))$. It follows that if we modify $\mathcal{G}$ by replacing G_1 and G_2 by their union and intersection, then the sum $\sum_{G \in \mathcal{G}} H(X(G))$ can only decrease. After a finite number of such modifications we can reach the case in which one of the sets in $\mathcal{G}$ is $\{1, \ldots, n\}$, and as this case has already been proved, this completes the proof. ■

Corollary 14.6.5 *Let $\mathcal{F}$ be a family of vectors in $S_1 \times S_2 \ldots \times S_n$. Let $\mathcal{G} = \{G_1, G_2, \ldots G_m\}$ be a collection of subsets of $N = \{1, 2, \ldots, n\}$, and suppose that each element $i \in N$ belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$ let $\mathcal{F}_i$ be the set of all projections of the members of $\mathcal{F}$ on G_i. Then*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

Proof. Let $X = (X_1, \ldots, X_n)$ be the random variable taking values in $\mathcal{F}$, where $P(X = F) = \frac{1}{|\mathcal{F}|}$ for all $F \in \mathcal{F}$. By Proposition 14.6.4

$$kH(X) \leq \sum_{i=1}^{m} H(X(G_i)).$$

But $H(X) = \log_2 |\mathcal{F}|$, whereas by Lemma 14.6.1, part 1, $H(X(G_i)) \leq \log_2 |\mathcal{F}_i|$, implying the desired result. ■

Since the volume of every d-dimensional measurable set in R^n can be approximated by the volume of an appropriate approximation of it by standard aligned boxes in a fine enough grid, the last result has the following geometric application, proved in Loomis and Whitney (1949) in a different manner.

Corollary 14.6.6 *Let B be a measurable body in the n dimensional Euclidean space, let $Vol(B)$ denote its (n-dimensional) volume, and let $Vol(B_i)$ denote the $(n - 1)$-dimensional volume of the projection of B on the hyperplane spanned by all coordinates besides the i-th one. Then*

$$(Vol(B))^{n-1} \leq \prod_{i=1}^{n} Vol(B_i).$$

■

If $S_i = \{0, 1\}$ for all i in Corollary 14.6.5, we get the following statement about set systems.

Corollary 14.6.7 [Chung et al. (1986)] *Let N be a finite set, and let $\mathcal{F}$ be a family of subsets of N. Let $\mathcal{G} = \{G_1, \dots G_m\}$ be a collection of subsets of N, and suppose that each element of S belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$ define $\mathcal{F}_i = \{F \cap G_i : F \in \mathcal{F}\}$. Then*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

■

We close the section with the following application of the last result, given in Chung et al. (1986) .

Corollary 14.6.8 *Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, \dots t\}$, and suppose that for any two members of $\mathcal{F}$ there is a triangle contained in both of them. Then*

$$|\mathcal{F}| < \frac{1}{4} 2^{\binom{t}{2}}.$$

Proof. Let N be the set of all $\binom{t}{2}$ unordered pairs of vertices in $T = \{1, 2 \dots, t\}$, and consider $\mathcal{F}$ as a family of subsets of N. Let $\mathcal{G}$ be the family of all subsets of N consisting of the edge-sets of unions of two vertex disjoint nearly equal complete graphs in T. Let

$$s = \binom{\lceil t/2 \rceil}{2} + \binom{\lfloor t/2 \rfloor}{2}$$

denote the number of edges of such a union, and let m denote the total number of members in $\mathcal{G}$. By symmetry, each edge in N lies in precisely $k = \frac{sm}{\binom{t}{2}}$ members of $\mathcal{G}$. The crucial point is that every two graphs in $\mathcal{F}$ must have at least one common edge in each $G \in \mathcal{G}$, since their intersection contains a triangle (and there are no triangles in the complement of G.) Therefore, in the notation of Corollary 14.6.7, the cardinality of each $\mathcal{F}_i$ is at most 2^{s-1}. We thus conclude that

$$|\mathcal{F}|^{\frac{sm}{\binom{t}{2}}} \leq (2^{s-1})^m,$$

implying that

$$|\mathcal{F}| \leq 2^{\binom{t}{2} - \binom{t}{2}/s},$$

and the desired result follows, as $s < \binom{t}{2}/2$. ■

Simonovits and Sós conjectured that if $\mathcal{F}$ satisfies the assumptions of the last corollary, then, in fact,

$$|\mathcal{F}| \leq \frac{1}{8} 2^{\binom{t}{2}},$$

which, if true, is tight. This remains open. It seems plausible to conjecture that there is some absolute constant $\epsilon > 0$, such that for any fixed graph H which is not a star-forest (that is, a forest each connected component of which is a star), the following holds. Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, \ldots t\}$, and suppose that for any two members of $\mathcal{F}$ there is a copy of H contained in both of them. Then

$$|\mathcal{F}| < (\frac{1}{2} - \epsilon) 2^{\binom{t}{2}}.$$

This is also open, though it is not difficult to show that it is true for every H of chromatic number at least 3, and that the conclusion fails for every star-forest H.

14.7 EXERCISES

1. Suppose that in the $(x_1, \ldots, x_k)$ tenure game of §14.3 the object of Paul is to maximize the number of faculty receiving tenure while the object of Carole is to minimize that number. Let v be that number with perfect play. Prove $v = \lfloor \sum_{i=1}^{k} x_i 2^{-k} \rfloor$.

2. Let $A_1, \ldots, A_n \subseteq \{1, \ldots, m\}$ with $\sum_{i=1}^{n} 2^{-|A_i|} < 1$. Paul and Carole alternately select distinct vertices from $\{1, \ldots, m\}$, Paul having the first move, until all vertices have been selected. Carole wins if she has selected all the vertices of some A_i. Paul wins if Carole does not win. Give a winning strategy for Paul.

3. Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, \ldots 2t\}$, and suppose that for any two members of $\mathcal{F}$ there is a perfect matching of t edges contained in both of them. Prove that

$$|\mathcal{F}| \leq 2^{\binom{2t}{2} - t}.$$

4. (Han's inequality). Let $X = (X_1, \ldots, X_m)$ be a random variable, and let $H(X)$ denote its entropy. For a subset I of $\{1, 2, \ldots, m\}$, let $X(I)$ denote the random variable $(X_i)_{i \in I}$. For $1 \leq q \leq m$, define

$$H_q(X) = \frac{1}{\binom{m-1}{q-1}} \sum_{Q \subset \{1, \ldots m\}, |Q| = q} H(X(Q)).$$

Prove that

$$H_1(X) \geq H_2(X) \geq \ldots \geq H_m(X) = H(X).$$

5. Let $X_i = \pm 1$, $1 \leq i \leq n$ be uniform and independent and let $S_n = \sum_{i=1}^{n} X_i$. Let $0 \leq p \leq \frac{1}{2}$. Prove

$$\Pr[S_n \geq (1 - 2p)n] \leq 2^{H(p)n} 2^{-n}$$

by computing precisely the Chernoff bound $\min_{\lambda \geq 0} E[e^{\lambda S_n}] e^{-\lambda(1-2p)n}$. (The case $p = 0$ shall require a slight adjustment in the method though the end result is the same.)

THE PROBABILISTIC LENS:

An Extremal Graph

Let T (top) and B (bottom) be disjoint sets of size m and let G be a bipartite graph, all edges between T and B. Suppose G contains no 4-cycle. How many edges can G have? This is a question from Extremal Graph Theory. Surprisingly, for some m we may give the precise answer.

Suppose $m = n^2 + n + 1$ and that a projective plane P of order n (and hence containing m points) exists. Identify T with the points of P and B with the lines of P and define $G = G_P$ by letting $t \in T$ be adjacent to $b \in B$ if and only if point t is on line b in P. As two points cannot lie on two lines, G_P contains no 4-cycle. We claim that such a G_P has the largest number of edges of any G containing no 4-cycle and further that any G containing no 4-cycle and having that many edges can be written in the form $G = G_P$.

Suppose G contains no 4-cycle. Let $b_1, b_2 \in B$ be a uniformly selected pair of distinct elements. For $t \in T$ let $D(t)$ be the set of $b \in B$ adjacent to t and $d(t) = |D(t)|$, the degree of t. Let I_t be the indicator random variable for t being adjacent to b_1, b_2. Then

$$E[I_t] = \Pr[b_1, b_2 \in D(t)] = \binom{d(t)}{2} \Big/ \binom{m}{2}$$

Now set

$$X = \sum_{t \in T} I_t,$$

the number of $t \in T$ adjacent to b_1, b_2. Then $X \leq 1$, i.e., all b_1, b_2 have at most one common neighbor. ($X \leq 1$ is actually equivalent to G containing no 4-cycle.)

Linearity of expectation gives

$$E[X] = \sum_{t \in T} E[I_t] = \sum_{t \in T} \binom{d(t)}{2} / \binom{m}{2}.$$

Let $\overline{d} = m^{-1} \sum_{t \in T} d(t)$ be the average degree. Convexity of the function $\binom{y}{2}$ gives

$$\sum_{t \in T} \binom{d(t)}{2} / \binom{m}{2} \geq m \binom{\overline{d}}{2} / \binom{m}{2}$$

with equality if and only if all $t \in T$ have the same degree. Now

$$1 \geq \max X \geq E[X] \geq m \binom{\overline{d}}{2} / \binom{m}{2}.$$

When $G = G_P$ all $d(x) = \overline{d}$ (every line has $m+1$ points) and $X = 1$ always (two points determine precisely one line) so that the above inequalities are all equalities and

$$1 = m \binom{\overline{d}}{2} / \binom{m}{2}.$$

Any graph with more edges would have a strictly larger $\overline{d}$ so that $1 \geq m\binom{\overline{d}}{2} / \binom{m}{2}$ would fail and the graph would contain a 4-cycle.

Suppose further G has the same number of edges as G_P and contains no 4-cycle. The inequalities then must be equalities and so $X = 1$ always. Define a geometry with points T and lines given by the neighbor sets of $b \in B$. As $X = 1$ any two points determine a unique line. Reversing the roles of T, B one also has that any two lines must determine a unique point. Thus G is generated from a projective plane.

15
Derandomization

> Math is natural. Nobody could have invented the mathematical universe. It was there, waiting to be discovered, and it's crazy; it's bizarre.
> – John Conway

As mentioned in Chapter 1, the probabilistic method supplies, in many cases, effective randomized algorithms for various algorithmic problems. In some cases, these algorithms can be derandomized and converted into deterministic ones. In this chapter we discuss some examples.

15.1 THE METHOD OF CONDITIONAL PROBABILITIES

An easy application of the basic probabilistic method implies the following statement, which is a special case of Theorem 2.3.1.

Proposition 15.1.1 *For every integer n there exists a coloring of the edges of the complete graph K_n by two colors so that the total number of monochromatic copies of K_4 is at most $\binom{n}{4} \cdot 2^{-5}$.*

Indeed, $\binom{n}{4} \cdot 2^{-5}$ is the expected number of monochromatic copies of K_4 in a random 2-edge-coloring of K_n, and hence a coloring as above exists.

Can we actually find *deterministically* such a coloring in time which is polynomial in n ? Let us describe a procedure that does it, and is a special case of a general technique called the *method of conditional probabilities.*

We first need to define a weight function for any partially colored K_n. Given a coloring of some of the edges of K_n by red and blue, we define, for each copy K of

K_4 in K_n, a weight $w(K)$ as follows. If at least one edge of K is colored red and at least one edge is colored blue then $w(K) = 0$. If no edge of K is colored, then $w(K) = 2^{-5}$, and if $r \geq 1$ edges of K are colored, all with the same color, then $w(K) = 2^{r-6}$. Also define the total weight W of the partially colored K_n as the sum $\sum w(K)$, as K ranges over all copies of K_4 in K_n. Observe that the weight of each copy K of K_4 is precisely the probability that it will be monochromatic, if all the presently uncolored edges of K_n will be assigned randomly and independently one of the two colors red and blue. Hence, by linearity of expectation, the total weight W is simply the expected number of monochromatic copies of K_4 in such a random extension of the partial coloring of K_n to a full coloring.

We can now describe the procedure for finding a coloring as in proposition 15.1.1. Order the $\binom{n}{2}$ edges of K_n arbitrarily, and construct the desired 2-coloring by coloring each edge either red or blue in its turn. Suppose $e_1, \ldots, e_{i-1}$ have already been colored, and we now have to color e_i. Let W be the weight of K_n, as defined above, with respect to the given partial coloring c of $e_1, \ldots, e_{i-1}$. Similarly, let W_{red} be the weight of K_n with respect to the partial coloring obtained from c by coloring e_i red, and let W_{blue} be the weight of K_n with respect to the partial coloring obtained from c by coloring e_i blue. By the definition of W (and as follows from its interpretation as an expected value)

$$W = \frac{W_{red} + W_{blue}}{2}.$$

The color of e_i is now chosen so as to minimize the resulting weight, i.e., if $W_{red} \leq W_{blue}$ then we color e_i red, otherwise, we color it blue. By the above inequality, the weight function never increases during the algorithm. Since at the beginning its value is exactly $\binom{n}{4}2^{-5}$, its value at the end is at most this quantity. However, at the end all edges are colored, and the weight is precisely the number of monochromatic copies of K_4. Thus the procedure above produces, deterministically and in polynomial time, a 2-edge-coloring of K_n satisfying the conclusion of Proposition 15.1.1.

Let us describe, now, the method of conditional probabilities in a more general setting. An instance of this method is due, implicitly, to Erdős and Selfridge (1973) , and more explicit examples appear in Spencer (1987) and in Raghavan (1988) . Suppose we have a probability space, and assume, for simplicity, that it is symmetric and contains 2^l points, denoted by the binary vectors of length l. Let $A_1, \ldots, A_s$ be a collection of events and suppose that $\sum_{i=1}^{s} Pr(A_i) = k$. Thus, k is the expected value of the number of events A_i that hold, and hence there is a point $(\epsilon_1, \ldots, \epsilon_l)$ in the space in which at most k events hold. Our objective is to find such a point deterministically.

For each choice of $(\epsilon_1, \ldots, \epsilon_{j-1})$ and for each event A_i, the conditional probability

$$Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1})$$

of the event A_i given the values of $\epsilon_1, \ldots, \epsilon_{j-1}$ is clearly the average of the two conditional probabilities corresponding to the two possible choices for ϵ_j. I.e.,

$$Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}) = \frac{Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 0) + Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 1)}{2}.$$

Consequently,

$$\sum_{i=1}^{s} Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1})$$

$$= \frac{\sum_{i=1}^{s} Pr(A_i \mid \epsilon_1,\ldots,\epsilon_{j-1},0)+\sum_{i=1}^{s} Pr(A_i \mid \epsilon_1,\ldots,\epsilon_{j-1},1)}{2}$$

$$\geq Min\,\{\sum_{i=1}^{s} Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 0), \sum_{i=1}^{s} Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 1)\}.$$

Therefore, if the values of ϵ_j are chosen, each one in its turn, so as to minimize the value of $\sum_{i=1}^{s} Pr(A_i \mid \epsilon_1, \ldots, \epsilon_j)$, then the value of this sum cannot increase. Since this sum is k at the beginning, it follows that it is at most k at the end. But at the end each ϵ_j is fixed, and hence the value of this sum is precisely the number of events A_i that hold at the point $(\epsilon_1, \ldots, \epsilon_l)$, showing that our procedure works.

Note that the assumptions that the probability space is symmetric and that it has 2^l points can be relaxed. The procedure above is efficient provided l is not too large (as is usually the case in combinatorial examples), and, more importantly, provided the conditional probabilities $Pr(A_i \mid \epsilon_1, \ldots, \epsilon_j)$ can be computed efficiently for each of the events A_i and for each possible value of $\epsilon_1, \ldots, \epsilon_j$. This is, indeed, the case in the example considered in Propostion 15.1.1. However, there are many interesting examples where this is not the case. A trick that can be useful in such cases is the introduction of *pessimistic estimators*, introduced by Raghavan (1988) . Consider, again, the symmetric probability space with 2^l points described above, and the events $A_1, \ldots, A_s$ in it. Suppose that for each event A_i, and for each $0 \leq j \leq l$, we have a function $f_j^i(\epsilon_1, \ldots, \epsilon_j)$, which can be efficiently computed. Assume, also , that

$$f_{j-1}^i(\epsilon_1, \ldots, \epsilon_{j-1}) \geq Min\,\{f_j^i(\epsilon_1, \ldots, \epsilon_{j-1}, 0), f_j^i(\epsilon_1, \ldots, \epsilon_{j-1}, 1)\}, \qquad (15.1)$$

and that f_j^i is an upper bound on the conditional probabilities for the event A_i, i.e.,

$$f_j^i(\epsilon_1, \ldots, \epsilon_j) \geq Pr(A_i \mid \epsilon_1, \ldots, \epsilon_j). \qquad (15.2)$$

(In fact, it suffices to assume such inequalities for the sums over i, but the version here suffices for our purpose.) In this case, if in the beginning $\sum_{i=1}^{s} f_0^i \leq t$, and we choose the values of the ϵ_j so as to minimize the sum $\sum_{i=1}^{s} f_j^i(\epsilon_1 \ldots, \epsilon_j)$ in each step, we get in the end a point $(\epsilon_1, \ldots, \epsilon_l)$ for which the sum $\sum_{i=1}^{s} f_l^i(\epsilon_1, \ldots, \epsilon_l) \leq t$. The number of events A_i that hold in this point is at most t. The functions f_j^i in the argument above are called pessimistic estimators.

This enables us to obtain efficient algorithms in some cases where there is no known efficient way of computing the required conditional probabilities. The following theorem is an example; it is related to some of the results in Chapter 14 and Chapter 12.

Theorem 15.1.2 *Let $(a_{ij})_{i,j=1}^n$ be an n by n matrix of reals, where $-1 \leq a_{ij} \leq 1$ for all i, j. Then one can find, in polynomial time, $\epsilon_1, \ldots, \epsilon_n \in \{-1, 1\}$ such that for every i, $1 \leq i \leq n$, the inequality $|\sum_{j=1}^{n} \epsilon_j a_{ij}| \leq \sqrt{2n \ln(2n)}$ holds.*

Proof. Consider the symmetric probability space on the 2^n points corresponding to the 2^n possible vectors $(\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n$. Define $\beta = \sqrt{2n\ln(2n)}$ and let A_i be the event $|\sum_{j=1}^{n} \epsilon_j a_{ij}| > \beta$. We next show that the method of conditional probabilities with appropriate pessimistic estimators enables us to find efficiently a point of the space in which no event A_i holds.

Define $\alpha = \beta/n$ and let $G(x)$ be the function

$$G(x) = cosh(\alpha x) = \frac{e^{\alpha x} + e^{-\alpha x}}{2}.$$

By comparing the terms of the corresponding Taylor series it is easy to see that for every real x:

$$G(x) \le e^{\frac{\alpha^2 x^2}{2}},$$

with strict inequality if both x and α are not 0. It is also simple to check that for every real x and y:

$$G(x)G(y) = \frac{G(x+y) + G(x-y)}{2}.$$

We can now define the functions f_p^i which will form our pessimistic estimators. For each $1 \le i \le n$ and for each $\epsilon_1, \ldots, \epsilon_p \in \{-1, 1\}$ we define

$$f_p^i(\epsilon_1, \ldots, \epsilon_p) = 2e^{-\alpha\beta} G(\sum_{j=1}^{p} \epsilon_j a_{ij}) \prod_{j=p+1}^{n} G(a_{ij}).$$

Obviously, these functions can be efficiently computed. It remains to check that they satisfy the conditions described in equations (15.1) and (15.2), and that the sum $\sum_{i=1}^{n} f_0^i$ is less than 1. This is proved in the following claims.

Claim 6 *For every* $1 \le i \le n$ *and every* $\epsilon_1, \ldots, \epsilon_{p-1} \in \{-1, 1\}$*:*

$$f_{p-1}^i(\epsilon_1, \ldots, \epsilon_{p-1}) \ge Min\, \{f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, -1), f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, 1)\}.$$

Proof. Put $v = \sum_{j=1}^{p-1} \epsilon_j a_{ij}$. By the definition of f_p^i and by the properties of G:

$$\begin{aligned}
f_{p-1}^i(\epsilon_1, \ldots, \epsilon_{p-1}) &= 2e^{-\alpha\beta} G(v) G(a_{ip}) \prod_{j=p+1}^{n} G(a_{ij}) \\
&= 2e^{-\alpha\beta} \frac{G(v - a_{ip}) + G(v + a_{ip})}{2} \prod_{j=p+1}^{n} G(a_{ij}) \\
&= \frac{f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, -1) + f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, 1)}{2} \\
&\ge Min\, \{ f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, -1), f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, 1)\},
\end{aligned}$$

completing the proof of the claim. ■

Claim 7 *For every* $1 \le i \le n$ *and every* $\epsilon_1, \ldots, \epsilon_{p-1} \in \{-1, 1\}$*:*

$$f^i_{p-1}(\epsilon_1, \ldots, \epsilon_{p-1}) \ge Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{p-1}).$$

Proof. Define v as in the proof of Claim 6. Then:

$$\begin{aligned}
Pr(A_i \mid \epsilon_1, \ldots, \epsilon_{p-1}) &\le Pr(v + \sum_{j \ge p} \epsilon_j a_{ij} > \beta) + Pr(-v - \sum_{j \ge p} \epsilon_j a_{ij} > \beta) \\
&= Pr(e^{\alpha(v + \sum_{j \ge p} \epsilon_j a_{ij})} > e^{\alpha\beta}) + Pr(e^{-\alpha(v + \sum_{j \ge p} \epsilon_j a_{ij})} > e^{\alpha\beta}) \\
&\le e^{\alpha v} e^{-\alpha\beta} E(e^{\alpha(\sum_{j \ge p} \epsilon_j a_{ij})}) + e^{-\alpha v} e^{-\alpha\beta} E(e^{-\alpha(\sum_{j \ge p} \epsilon_j a_{ij})}) \\
&= 2e^{-\alpha\beta} G(v) \prod_{j \ge p} G(a_{ij}) = f^i_{p-1}(\epsilon_1, \ldots, \epsilon_{p-1}).
\end{aligned}$$

This completes the proof of Claim 7. ■

To establish the theorem it remains to show that $\sum_{i=1}^n f^i_0 < 1$. Indeed, by the properties of G and by the choice of α and β:

$$\begin{aligned}
\sum_{i=1}^n f^i_0 &= \sum_{i=1}^n 2e^{-\alpha\beta} \prod_{j=1}^n G(a_{ij}) \\
&\le \sum_{i=1}^n 2e^{-\alpha\beta} \prod_{j=1}^n e^{\frac{\alpha^2 a_{ij}^2}{2}} \\
&\le \sum_{i=1}^n 2e^{-\alpha\beta} e^{\frac{\alpha^2 n}{2}} \\
&= 2ne^{\frac{\alpha^2 n}{2} - \alpha\beta} = 2ne^{-\frac{\alpha^2 n}{2}} = 1.
\end{aligned}$$

Moreover, the first inequality is strict unless $a_{ij} = 0$ for all i, j, whereas the second is strict unless $a_{ij}^2 = 1$ for all i, j. This completes the proof of the theorem. ■

15.2 D-WISE INDEPENDENT RANDOM VARIABLES IN SMALL SAMPLE SPACES

The complexity class NC is, roughly speaking, the class of all problems that can be solved in time which is poly-logarithmic (in the size of the input) using a polynomial number of parallel processors. Several models of computation, which are a theoretical abstraction of the parallel computer, have been used in considering this class. The most common one is the EREW (=Exclusive Read, Exclusive Write) PRAM, in which different processors are not allowed to read from or write into the same memory cell simultanously. See Karp and Ramachandran (1990) for more details.

Let n denote the size of the input. There are several simple tasks that can be easily performed in NC. For example, it is possible to copy the content of a cell c into $m = n^{O(1)}$ cells in time $O(\log n)$, using, say, m processors. To do so, consider a complete binary tree with m leaves and associate each of its internal vertices with a processor. At first, the processor corresponding to the root of the tree reads from c and writes its content in two cells, corresponding to its two children. Next, each of these two, in parallel, reads from its cell and writes its content in two cells corresponding to its two children. In general, at the i-th step all the processors whose distance from the root of the tree is $i - 1$,in parallel, read the content of c previously stored in their cells and write it twice. The procedure clearly ends in time $O(\log m)$, as claimed. (In fact, it can be shown that $O(m/\log m)$ processors suffice for this task but we do not try to optimize this number here).

A similar technique can be used for computing the sum of m numbers with m processors in time $O(\log m)$; We consider the numbers as if they lie on the leaves of a complete binary tree with m leaves, and in the $i-th$ step each one of the processors whose distance from the leaves is i computes, in parallel, the sum of the two numbers previously computed by its children. The root will clearly have, in such a way, the desired sum in time $O(\log m)$.

Let us now return to the edge-coloring problem of the complete graph K_n discussed in Proposition 15.1.1. By the remarks above, the problem of *checking* if in a given edge-coloring there are at most $\binom{n}{4}2^{-5}$ monochromatic copies of K_4 is in NC, i.e., this checking can be done in time $(\log n)^{O(1)}$ -(in fact, in time $O(\log n)$)- using $n^{O(1)}$ processors. Indeed, we can first copy the given coloring $\binom{n}{4}$ times. Then we assign a processor for each copy of K_4 in K_n, and this processor checks if its copy is monochromatic or not (all these checkings can be done in parallel, since we have enough copies of the coloring). Finally, we sum the number of processors whose copies are monochromatic. Clearly we can complete the work in time $O(\log n)$ using $n^{O(1)}$ parallel processors.

Thus we can *check*, in NC, if a given coloring of K_n satisfies the assertion of Proposition 15.1.1. Can we *find* such a coloring deterministically in NC? The method described in the previous section does not suffice, as the edges have been colored one by one, so the procedure is sequential and requires time $\Omega(n^2)$. However, it turns out that in fact we can find , in NC, a coloring with the desired properties by applying a method which relies on a technique first suggested by Joffe (1974) , and later developed by many researchers. This method is a general technique for converting randomized algorithms whose analysis only depends on d-wise rather than fully independent random choices (for some constant d) into deterministic (and in many cases also parallel) ones. Our approach here follows the one of Alon, Babai and Itai (1986) , but for simplicity we only consider here the case of random variables that take the two values $0, 1$ with equal probability.

The basic idea is to replace an exponentially large sample space by one of polynomial size. If a random variable on such a space takes a certain value with positive probability, then we can find a point in the sample space in which this happens simply by deterministically checking all the points. This can be done with no loss of time by using a polynomial number of parallel processors. Note that for the edge-coloring problem considered in Proposition 15.1.1, 6-wise independence of the random variables corresponding to the colors of the edges suffice- since this already gives a probability of 2^{-5} for each copy of K_4 to be monochromatic, and hence gives the required expected value of monochromatic copies. Therefore, for this specific example it suffices to construct a sample space of size $n^{O(1)}$ and $\binom{n}{2}$ random variables in it, each taking the values 0 and 1 with probability $1/2$, such that each 6 of the random variables are independent.

Small sample spaces with many d-wise independent $0,1$-random variables in them can be constructed from any linear error correcting code with appropriate parameters. The construction we describe here is based on the binary BCH codes (see, e.g., MacWilliams and Sloane (1977)).

Theorem 15.2.1 *Suppose $n = 2^k - 1$ and $d = 2t + 1$. Then there exists a symmetric probability space Ω of size $2(n+1)^t$ and d-wise independent random variables $y_1, \ldots, y_n$ over Ω each of which takes the values 0 and 1 with probability $1/2$.*

The space and the variables are explicitly constructed, given a representation of the field $F = GF(2^k)$ as a k-dimensional algebra over $GF(2)$.

Proof. Let $x_1, \ldots, x_n$ be the n nonzero elements of F, represented as column-vectors of length k over $GF(2)$. Let H be the following $1 + kt$ by n matrix over $GF(2)$:

$$\begin{array}{cccc} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1^3 & x_2^3 & \ldots & x_n^3 \\ .. & & . & \\ .. & & . & \\ .. & & . & \\ x_1^{2t-1} & x_2^{2t-1} & \ldots & x_n^{2t-1} \end{array}$$

This is the parity check matrix of the extended binary BCH code of length n and designed distance $2t + 2$. It is well known that any $d = 2t + 1$ columns of H are linearly independent over $GF(2)$. For completeness, we present the proof in the next lemma.

Lemma 15.2.2 *Any set of $d = 2t + 1$ columns of H is linearly independent over $GF(2)$.*

Proof. Let $J \subset \{1, 2, \ldots, n\}$ be a subset of cardinality $|J| = 2t + 1$ of the set of indices of the columns of H. Suppose that $\sum_{j \in J} z_j H_j = 0$, where H_j denotes the j-th column of H and $z_j \in GF(2)$. To complete the proof we must show that $z_j = 0$ for all $j \in J$. By the assumption,

$$\sum_{j \in J} z_j x_j^i = 0 \tag{15.3}$$

for $i = 0$ and for every odd i satisfying $1 \leq i \leq 2t - 1$. Suppose, now, that $a = 2^b \cdot l$, where $l \leq 2t - 1$ is an odd number. By squaring the equation (15.3) for $i = l$ b

times, using the fact that in characteristic 2, $(u+v)^2 = u^2 + v^2$, and the fact that since each z_j is either 0 or 1, the equality $z_j = z_j^2$ holds for all j, we conclude that equation (15.3) holds for $i = a$. Consequently, (15.3) holds for all i, $0 \le i \le 2t$. This is a homogeneous system of $2t+1$ linear equations in $2t+1$ variables. The matrix of the coefficients is a Vandermonde matrix, which is nonsingular. Thus, the only solution is the trivial one $z_j = 0$ for all $j \in J$, completing the proof of the lemma. ■

Returning to the proof of the theorem, we define $\Omega = \{1, 2, \ldots, 2(n+1)^t\}$, and let $A = (a_{ij}), i \in \Omega, 1 \le j \le n$ be the $(0,1)$-matrix whose $2(n+1)^t = 2^{kt+1}$ rows are all the linear combinations (over $GF(2)$) of the rows of H. The sample space Ω is now endowed with the uniform probability measure, and the random variable y_j is defined by the formula $y_j(i) = a_{ij}$ for all $i \in \Omega, 1 \le j \le n$.

It remains to show that the variables y_j are d-wise independent, and that each of them takes the values 0 and 1 with equal probability. For this we have to show that for every set J of up to d columns of A, the rows of the $|\Omega|$ by $|J|$ submatrix $A_J = (a_{ij}), i \in \Omega, j \in J$ take on each of the $2^{|J|}$ $(0,1)$-vectors of length $|J|$ equally often. However, by Lemma 15.2.2 the columns of the corresponding submatrix H_J of H are linearly independent. The number of rows of A_J that are equal to any given vector is precisely the number of linear combinations of the rows of H_J that are equal to this vector. This number is the number of solutions of a system of $|J|$ linearly independent linear equations in $kt+1$ variables, which is, of course, $2^{kt+1-|J|}$, independent of the vector of free coefficients. This completes the proof of the theorem. ■

Theorem 15.2.1 supplies an efficient way of constructing, for every fixed d and every n, a sample space of size $O(n^{\lfloor d/2 \rfloor})$ and n d-wise independent random variables in it, each taking the values 0 and 1 with equal probability. In particular, we can use such a space of size $O(\binom{n}{2}^3) = O(n^6)$ for finding a coloring as in Proposition 15.1.1 in NC. Several other applications of Theorem 15.2.1 appear in the paper of Alon et al. (1986) .

It is natural to ask if the size $O(n^{\lfloor d/2 \rfloor})$ can be improved. We next show that this size is optimal, up to a constant factor (depending on d).

Let us call a random variable *almost constant* if it attains a single value with probability 1. Let $m(n,d)$ denote the function defined by:

$$m(n,d) = \sum_{j=0}^{d/2} \binom{n}{j} \quad if\ \ d\ \ is\ \ even,$$

and

$$m(n,d) = \sum_{j=0}^{(d-1)/2} \binom{n}{j} + \binom{n-1}{(d-1)/2} \quad if\ \ d\ \ is\ \ odd.$$

Observe that for every fixed d, $m(n,d) = \Omega(n^{\lfloor d/2 \rfloor})$.

Proposition 15.2.3 *If the random variables $y_1, \ldots, y_n$ over the sample space Ω are d-wise independent and none of them is almost constant then $|\Omega| \ge m(n,d)$.*

Note that we assume here neither that Ω is a symmetric space nor that the variables y_j are $(0,1)$-variables.

Proof. Clearly we may assume that the expected value of each y_j is 0 (since otherwise we can replace y_j by $y_j - E(y_j)$). For each subset S of $\{1,\ldots,n\}$, define $\alpha_S = \prod_{j\in S} y_j$. Observe that since no y_j is almost constant and since the variables are d-wise independent

$$E(\alpha_S\alpha_S) = \prod_{j\in S} Var(y_j) > 0 \tag{15.4}$$

for all S satisfying $|S| \le d$. Similarly, for all S and T satisfying $|S\cup T| \le d$ and $S \ne T$ we have

$$E(\alpha_s\alpha_T) = \prod_{j\in S\cap T} Var(y_j) \prod_{j\in S\cup T\setminus(S\cap T)} E(y_j) = 0. \tag{15.5}$$

Let $S_1,\ldots S_m$, where $m = m(n,d)$, be subsets of $\{1,\ldots,n\}$ such that the union of each two is of size at most d. (Take all subsets of size at most $d/2$, and if d is odd add all the subsets of size $(d+1)/2$ containing 1.)

To complete the proof, we show that the m functions α_{S_j} (considered as real vectors of length $|\Omega|$) are linearly independent. This implies that $|\Omega| \ge m = m(n,d)$, as stated in the proposition.

To prove linear independence, suppose $\sum_{j=1}^m c_j\alpha_{S_j} = 0$. Multiplying by α_{S_i} and computing expected values we obtain, by (15.5):

$$0 = \sum_{j=1}^m c_j E(\alpha_{S_j}\alpha_{S_i}) = c_i E(\alpha_{S_i}\alpha_{S_i}).$$

This implies, by (15.4), that $c_i = 0$ for all i. The required linear independence follows, completing the proof. ■

The last proposition shows that the size of a sample space with n d-wise independent nontrivial random variables can be polynomial in n only when d is fixed. However, as shown by Naor and Naor (1990) , if we only require the random variables to be *almost* d-wise independent, the size can be polynomial even when $d = \Omega(\log n)$. Such sample spaces and random variables, that can be constructed explicitly in several ways, have various interesting applications in which almost d-wise independence suffices. More details appear in Naor and Naor (1990) and in Alon, Goldreich, stad and Peralta (1990) .

15.3 EXERCISES

1. Let $A_1,\ldots,A_n \subseteq \{1,\ldots,m\}$ with $\sum_{i=1}^n 2^{1-|A_i|} < 1$. Prove there exists a two coloring $\chi : \{1,\ldots,m\} \to \{0,1\}$ with no A_i monochromatic. With $m = n$ give a deterministic algorithm to find such a χ in polynomial time.

2. Describe a deterministic algorithm which, given n, constructs, in time polynomial in n, a family $\mathcal{F}$ of n^{10} subsets of the set $N = \{1, 2, \ldots, n\}$, where each $F \in \mathcal{F}$ is of size at most $10 \log n$ and for every family $\mathcal{G}$ of n subsets each of cardinality $n/2$ of N, there is an $F \in \mathcal{F}$ that intersects all members of $\mathcal{G}$.

THE PROBABILISTIC LENS:
Crossing Numbers, Incidences, Sums and Products

In this lens we start with a simple result in graph theory, whose proof is probabilistic, and then describe some of its fascinating consequences in Combinatorial Geometry and Combinatorial Number Theory. Some versions of most of these seemingly unrelated consequences have been proved before, in a far more complicated manner. Before the discovery of the new proofs shown here, the only clue that there might be a connection between all of them has been the fact that Endre Szemerédi is one of the co-authors of each of the papers providing the first proofs.

An *embedding* of a graph $G = (V, E)$ in the plane is a planar representation of it, where each vertex is represented by a point in the plane, and each edge uv is represented by a curve connecting the points corresponding to the vertices u and v. The *crossing number* of such an embedding is the number of pairs of intersecting curves that correspond to pairs of edges with no common endpoints. The *crossing number* $cr(G)$ of G is the minimum possible crossing number in an embedding of it in the plane. The following theorem was proved by Ajtai, Chvátal, Newborn and Szemerédi (1982) and, independently, by Leighton. Here we describe a very short probabilistic proof.

Theorem 1 *The crossing number of any simple graph $G = (V, E)$ with $|E| \geq 4|V|$ is at least $\frac{|E|^3}{64|V|^2}$.*

Proof. By Euler's formula any simple planar graph with n vertices has at most $3n - 6$ edges, implying that the crossing number of any simple graph with n vertices and m

edges is at least $m-(3n-6) > m-3n$. Let $G=(V,E)$ be a graph with $|E| \geq 4|V|$ embedded in the plane with $t = cr(G)$ crossings. Let H be the random induced subgraph of G obtained by picking each vertex of G, randomly and independently, to be a vertex of H with probability p (where p will be chosen later). The expected number of vertices of H is $p|V|$, the expected number of its edges is $p^2|E|$, and the expected number of crossings in its given embedding is p^4t, implying that the expected value of its crossing number is at most p^4t. Therefore, $p^4t \geq p^2|E| - 3p|V|$, implying that

$$cr(G) = t \geq \frac{|E|}{p^2} - 3\frac{|V|}{p^3}.$$

Without trying to optimize the constant factor, substitute $p = 4|V|/|E|$ (≤ 1), to get the desired result. ■

Székely (1997) noticed that this result can be applied to obtain a surprisingly simple proof of a result of Szemerédi and Trotter in Combinatorial Geometry. The original proof is far more complicated.

Theorem 2 *Let P be a set of n distinct points in the plane, and let L be a set of m distinct lines. Then, the number of incidences between the members of P and those of L (that is, the number of pairs (p,l) with $p \in P$, $l \in L$ and $p \in l$) is at most $c(m^{2/3}n^{2/3}+m+n)$, for some absolute constant c.*

Proof. Denote the number of incidences by I. Let $G=(V,E)$ be the graph whose vertices are all members of P, where two are adjacent if and only if they are consecutive points of P on some line in L. Clearly, $|V| = n$ and $|E| = I - m$. Note that G is already given embedded in the plane, where the edges are represented by segments of the corresponding lines in L. In this embedding, every crossing is an intersection point of two members of L, implying that $cr(G) \leq \binom{m}{2} \leq m^2/2$. By Theorem 1, either $I - m = |E| < 4|V| = 4n$, that is, $I \leq m+4n$, or

$$\frac{m^2}{2} \geq cr(G) \geq \frac{(I-m)^3}{64n^2},$$

implying that $I \leq (32)^{1/3}m^{2/3}n^{2/3}+m$. In both cases $I \leq 4(m^{2/3}n^{2/3}+m+n)$, completing the proof. ■

An analogous argument shows that the maximum possible number of incidences between a set of n points and a set of m unit cycles in the plane does not exceed $O(m^{2/3}n^{2/3}+m+n)$, and this implies that the number of unit distances determined by a set of n points in the plane is at most $O(n^{4/3})$. While the above upper bound for the number of incidences of points and lines is sharp, up to a constant factor, an old conjecture of Erdős asserts that the maximum possible number of unit distances determined by a set of n points in the plane is at most $c_\epsilon n^{1+\epsilon}$ for any $\epsilon > 0$. The $O(n^{4/3})$ estimate is, however, the best known upper bound, and has first been proved by Spencer, Szemerédi and Trotter in a far more complicated way.

Elekes (1997) found several applications of Theorem 2 in Additive Number Theory. Here, too, the proofs are amazingly simple. Here is a representative result.

Theorem 3 *For any three sets A,B and C of s real numbers each,*

$$|A \cdot B + C| = |\{ab + c : \ a \in A, b \in B, c \in C\}| \geq \Omega(s^{3/2}).$$

Proof. Put $R = A \cdot B + C$, $|R| = r$ and define

$$P = \{(a,t) : a \in A, t \in R\}, \quad L = \{y = bx + c : b \in B, c \in C\}.$$

Thus P is a set of $n = sr$ points in the plane, L is a set of $m = s^2$ lines in the plane, and each line $y = bx + c$ in L is incident with s points of P, that is, with all the points $\{a, ab + c) : a \in A\}$. Therefore, by Theorem 2, $s^3 \leq 4(s^{4/3}(sr)^{2/3} + sr + s^2)$, implying that $r \geq \Omega(s^{3/2})$, as needed. ■

The same method implies that for every set A of n reals, either $|A + A| \geq \Omega(n^{5/4})$ or $|A \cdot A| \geq n^{5/4}$, greatly improving and simplifying a result of Erdős and Szemerédi.

Appendix A
Bounding of Large Deviations

A.1 BOUNDING OF LARGE DEVIATIONS

We give here some basic bounds on large deviations that are useful when employing the probabilistic method. Our treatment is self-contained. Most of the results may be found in, or immediately derived from, the seminal paper of Chernoff (1952) . While we are guided by asymptotic considerations the inequalities are proven for all values of the parameters in the specified region. The first result, while specialized, contains basic ideas found throughout the Appendix.

Theorem A.1.1 *Let $X_i, 1 \leq i \leq n$, be mutually independent random variables with*

$$\Pr[X_i = +1] = \Pr[X_i = -1] = \frac{1}{2}$$

and set, following the usual convention,

$$S_n = X_1 + \ldots X_n.$$

Let $a > 0$. Then

$$\Pr[S_n > a] < e^{-a^2/2n}.$$

Remark. For large n the Central Limit Theorem implies that S_n is approximately normal with zero mean and standard deviation $\sqrt{n}$. In particular, for any fixed u

$$\lim_{n\to\infty} \Pr[S_n > u\sqrt{n}] = \int_{t=u}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

which one can show directly is less than $e^{-u^2/2}$. Our proof, we emphasize once again, is valid for all n and all $a > 0$.

We require Markov's inequality which states: Suppose that Y is an arbitrary nonnegative random variable, $\alpha > 0$. Then

$$\Pr[Y > \alpha E[Y]] < 1/\alpha.$$

Proof.[A.1.1] Fix n, a and let, for the moment, $\lambda > 0$ be arbitrary. For $1 \le i \le n$

$$E[e^{\lambda X_i}] = (e^{\lambda} + e^{-\lambda})/2 = \cosh(\lambda).$$

We require the inequality

$$\cosh(\lambda) \le e^{\lambda^2/2},$$

valid for all $\lambda > 0$, the special case $\alpha = 0$ of Lemma A.5 below. (The inequality may be more easily shown by comparing the Taylor series of the two functions termwise.)

$$e^{\lambda S_n} = \prod_{i=1}^{n} e^{\lambda X_i}.$$

Since the X_i are mutually independent so are the $e^{\lambda X_i}$, expectations multiply and

$$E[e^{\lambda S_n}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] = [\cosh(\lambda)]^n < e^{\lambda^2 n/2}.$$

We note that $S_n > a$ if and only if $e^{\lambda S_n} > e^{\lambda a}$ and apply Markov's inequality so that

$$\Pr[S_n > a] = \Pr[e^{\lambda S_n} > e^{\lambda a}] < E[e^{\lambda S_n}]/e^{\lambda a} \le e^{\lambda^2 n/2 - \lambda a}.$$

We set $\lambda = a/n$ to optimize the inequality, $\Pr[S_n > a] < e^{-a^2/2n}$ as claimed. ■

By symmetry we immediately have

Corollary A.1.2 *Under the assumptions of Theorem A.1.1*

$$\Pr[|S_n| > a] < 2e^{-a^2/2n}.$$

Our remaining results will deal with distributions X of the following prescribed type.

Assumptions A.1.3

$$p_1, \ldots, p_n \in [0, 1]$$

$$p = (p_1 + \ldots + p_n)/n$$

$$X_1, \ldots, X_n \text{ mutually independent with}$$

$$\Pr[X_i = 1 - p_i] = p_i$$

$$\Pr[X_i = -p_i] = 1 - p_i$$

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

Remark. Clearly $E[X] = E[X_i] = 0$. When all $p_i = 1/2$, X has distribution $S_n/2$. When all $p_i = p$, X has distribution $B(n,p) - np$ where $B(n,p)$ is the usual Binomial Distribution.

Theorem A.1.4 *Under assumptions A.1.3 and with* $a > 0$

$$\Pr[X > a] < e^{-2a^2/n}.$$

Lemma A.1.5 *For all reals* α, β *with* $|\alpha| \le 1$

$$\cosh(\beta) + \alpha \sinh(\beta) \le e^{\beta^2/2 + \alpha\beta}.$$

Proof. This is immediate if $\alpha = +1$ or $\alpha = -1$ or $|\beta| \ge 100$. If the Lemma were false the function

$$f(\alpha, \beta) = \cosh(\beta) + \alpha \sinh(\beta) - e^{\beta^2/2 + \alpha\beta}$$

would assume a positive global maximum in the interior of the rectangle

$$R = \{(\alpha, \beta) : |\alpha| \le 1, |\beta| \le 100\}.$$

Setting partial derivatives equal to zero we find

$$\sinh(\beta) + \alpha \cosh(\beta) = (\alpha + \beta) e^{\beta^2/2 + \alpha\beta},$$

$$\sinh(\beta) = \beta e^{\beta^2/2 + \alpha\beta},$$

and thus $\tanh(\beta) = \beta$ which implies $\beta = 0$. But $f(\alpha, 0) = 0$ for all α, a contradiction. ■

Lemma A.1.6 *For all* $\theta \in [0, 1]$ *and all* λ

$$\theta e^{\lambda(1-\theta)} + (1-\theta) e^{-\lambda\theta} \le e^{\lambda^2/8}.$$

Proof. Setting $\theta = (1+\alpha)/2$ and $\lambda = 2\beta$, Lemma A.1.6 reduces to Lemma A.5. ■

Proof.[Theorem A.1.4] Let, for the moment, $\lambda > 0$ be arbitrary.

$$E[e^{\lambda X_i}] = p_i e^{\lambda(1-p_i)} + (1-p_i) e^{-\lambda p_i} \le e^{\lambda^2/8}$$

by Lemma A.1.6. Then

$$E[e^{\lambda X}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] \le e^{\lambda^2 n/8}.$$

Applying Markov's inequality

$$\Pr[X > a] = \Pr[e^{\lambda X} > e^{\lambda a}] < E[e^{\lambda X}]/e^{\lambda a} \le e^{\lambda^2 n/8 - \lambda a}.$$

We set $\lambda = 4a/n$ to optimize the inequality: $\Pr[X > a] < e^{-2a^2/n}$ as claimed. ■

Again by symmetry we immediately have

Corollary A.1.7 *Under assumptions A.1.3 and with* $a > 0$

$$\Pr[|X| > a] < 2e^{-2a^2/n}.$$

Under assumptions A.1.3 with λ arbitrary

$$E[e^{\lambda X}] = \prod_{i=1}^n E[e^{\lambda X_i}] = \prod_{i=1}^n [p_i e^{\lambda(1-p_i)} + (1-p_i)e^{-\lambda p_i}] = e^{-\lambda pn} \prod_{i=1}^n [p_i e^{\lambda} + (1-p_i)].$$

With λ fixed the function

$$f(x) = \ln[xe^{\lambda} + 1 - x] = \ln[Bx + 1] \text{ with } B = e^{\lambda} - 1$$

is concave and hence (Jensen's Inequality)

$$\sum_{i=1}^{n} f(p_i) \le nf(p).$$

Exponentiating both sides

$$\prod_{i=1}^{n} [p_i e^{\lambda} + (1-p_i)] \le [pe^{\lambda} + (1-p)]^n,$$

so that

Lemma A.1.8 *Under the assumptions A.1.3*

$$E[e^{\lambda X}] < e^{-\lambda pn}[pe^{\lambda} + (1-p)]^n.$$

Theorem A.1.9 *Under the assumptions A.1.3 and with* $a > 0$

$$\Pr[X \ge a] < e^{-\lambda pn}[pe^{\lambda} + (1-p)]^n e^{-\lambda a}$$

for all $\lambda > 0$.

Proof. $\Pr[X > a] = \Pr[e^{\lambda X} > e^{\lambda a}] < E[e^{\lambda X}]/e^{\lambda a}$. Now apply Lemma A.1.8. ■

Remark. For given p, n, a, an optimal assignment of λ in Theorem A.1.9 is found by elementary calculus to be

$$\lambda = \ln[(\frac{1-p}{p})(\frac{a+np}{n-(a+np)})].$$

This value is oftentimes too cumbersome to be useful. We employ suboptimal λ to achieve more convenient results.

Setting $\lambda = \ln[1 + a/pn]$ and using the fact that $(1 + a/n)^n \leq e^a$, Theorem A.1.9 implies

Corollary A.1.10

$$\Pr[X \geq a] < e^{a - pn\ln(1+a/pn) - a\ln(1+a/pn)}.$$

Theorem A.1.11

$$\Pr[X \geq a] < e^{-a^2/2pn + a^3/2(pn)^2}.$$

Proof. With $u = a/pn$ apply the inequality

$$\ln(1+u) \geq u - u^2/2,$$

valid for all $u \geq 0$, to Corollary A.1.10. ■

When all $p_i = p$, X has variance $np(1-p)$. With $p = o(1)$ and $a = o(pn)$ this bound reflects the approximation of X by a Normal Distribution with variance $\sim np$. The bound of Theorem A.1.11 hits a minimum at $a = 2pn/3$. For $a > 2pn/3$ we have the simple bound

$$\Pr[X > a] \leq \Pr[X > 2pn/3] < e^{-2pn/27}.$$

This is improved by the following.

Theorem A.1.12 *For* $\beta > 1$

$$\Pr[X \geq (\beta - 1)pn] < [e^{\beta-1}\beta^{-\beta}]^{pn}.$$

Proof. Direct “plug in” to Corollary A.1.10. ■

$X + pn$ may be interpreted as the number of successes in n independent trials when the probability of success in the i-th trial is p_i.

Theorem A.1.13 *Under assumptions A.1.3 and with* $a > 0$*,*

$$\Pr[X < -a] < e^{-a^2/2pn}.$$

Note that one cannot simply employ “symmetry” as then the roles of p and $1-p$ are interchanged.

Proof. Let $\lambda > 0$ be, for the moment, arbitrary. Then by the argument preceding A.1.8

$$E[e^{-\lambda X}] \leq e^{\lambda pn}[pe^{-\lambda} + (1-p)]^n.$$

Thus

$$\Pr[X < -a] = \Pr[e^{-\lambda X} > e^{\lambda a}] < e^{\lambda pn}[pe^{-\lambda} + (1-p)]^n e^{-\lambda a},$$

analogous to Theorem A.1.9. We employ the inequality

$$1 + u \leq e^u,$$

valid for all u, so that

$$pe^{-\lambda} + (1-p) = 1 + (e^{-\lambda} - 1)p < e^{p(e^{-\lambda}-1)}$$

and

$$\Pr[X < -a] \leq e^{\lambda pn + np(e^{-\lambda}-1) - \lambda a} = e^{np(e^{-\lambda}-1+\lambda)-\lambda a}.$$

We employ the inequality

$$e^{-\lambda} \leq 1 - \lambda + \lambda^2/2,$$

valid for all $\lambda > 0$. (Note: The analogous inequality $e^{\lambda} \leq 1 + \lambda + \lambda^2/2$ is *not* valid for $\lambda > 0$ and so this method, when applied to $\Pr[X > a]$, requires an "error" term as the one found in Theorem A.1.11.) Now

$$\Pr[X < -a] \leq e^{np\lambda^2/2 - \lambda a}.$$

We set $\lambda = a/np$ to optimize the inequality: $Pr[X < -a] < e^{-a^2/2pn}$ as claimed. ■

For clarity the following result is often useful.

Corollary A.1.14 *Let Y be the sum of mutually independent indicator random variables, $\mu = E[Y]$. For all $\epsilon > 0$*

$$\Pr[|Y - \mu| > \epsilon\mu] < 2e^{-c_\epsilon \mu}$$

where $c_\epsilon > 0$ depends only on ϵ.

Proof. Apply theorems A.1.12, A.1.13 with $Y = X + pn$ and

$$c_\epsilon = \min[-\ln(e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}), \epsilon^2/2].$$

■

The asymmetry between $\Pr[X < a]$ and $\Pr[X > a]$ given by Theorems A.1.12, A.1.13 is real. The estimation of X by a normal distribution with zero mean and variance np is roughly valid for estimating $\Pr[X < a]$ for any a and for estimating $\Pr[X > a]$ while $a = o(np)$. But when a and np are comparable or when $a >> np$

the Poisson behaviour "takes over" and $\Pr[X > a]$ *cannot* be accurately estimated by using the normal distribution.

We conclude with two large deviation results involving distributions other than sums of indicator random variables.

Theorem A.1.15 *Let P have Poisson distribution with mean μ. For $\epsilon > 0$*

$$\Pr[P \leq \mu(1-\epsilon)] \leq e^{-\epsilon^2\mu/2},$$

$$\Pr[P \geq \mu(1+\epsilon)] \leq \left[e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}\right]^{\mu}.$$

Proof. For any s

$$\Pr[P = s] = \lim_{n\to\infty} \Pr[B(n, \frac{\mu}{n}) = s].$$

Apply Theorems A.1.12,A.1.13. ■

Theorem A.1.16 *Let X_i, $1 \leq i \leq n$, be mutually indpendent with all $E[X_i] = 0$ and all $|X_i| \leq 1$. Set $S = X_1 + \ldots + X_n$. Then*

$$\Pr[S > a] < e^{-a^2/2n}.$$

Proof. Set, as in the proof of theorem A.1.1, $\lambda = a/n$. Set

$$h(x) = \frac{e^{\lambda} + e^{-\lambda}}{2} + \frac{e^{\lambda} - e^{-\lambda}}{2}x$$

For $x \in [-1, 1]$, $e^{\lambda x} \leq h(x)$. ($y = h(x)$ is the chord through the points $x = \pm 1$ of the convex curve $y = e^{\lambda x}$.) Thus

$$E[e^{\lambda X_i}] \leq E[h(X_i)] = h(E[X_i]) = h(0) = \cosh \lambda.$$

The remainder of the proof follows as in A.1.1. ■

Theorem A.1.17 *Suppose $E[X] = 0$ and no two values of X are ever more than one apart. Then for all $\lambda \geq 0$*

$$E[e^{\lambda X}] \leq e^{\lambda^2/8}.$$

Proof. Fix $b \in [-\frac{1}{2}, \frac{1}{2}]$ with $X \in [\frac{-1+b}{2}, \frac{+1+b}{2}]$. Let $y = h(x)$ be the straight line intersecting the curve $y = e^{\lambda x}$ at the points $(\pm 1 + b)/2)$. As $e^{\lambda x}$ is a convex function, $e^{\lambda x} \leq h(x)$ for all $x \in [\frac{-1+b}{2}, \frac{+1+b}{2}]$. Thus

$$E[e^{\lambda X}] \leq E[h(X)] = h[E(X)] = h(0).$$

We calculate $h(0) = e^{\lambda b/2}[\cosh(\lambda/2) - b\sinh(\lambda/2)]$ which is at most $e^{\lambda^2/8}$ by Lemma A.1.5 ■

Theorem A.1.18 *Let X_i, $1 \le i \le n$ be independent random variables with each $E[X_i] = 0$ and no two values of any X_i ever more than one apart. [We allow, however, values of different X_i, X_j to be further apart.] Set $S = X_1 + \ldots + X_n$. Then*

$$\Pr[S > a] < e^{-2a^2}.$$

Proof. $E[e^{\lambda S}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] \le e^{n\lambda^2/8}$ by Theorem A.1.17 Then for $\lambda \ge 0$

$$\Pr[S > a] = \Pr[e^{\lambda S} \ge e^{\lambda a}] \le \exp\left[\frac{n\lambda^2}{8} - \lambda a\right]$$

and we set $\lambda = 4a/n$. ■

We have been roughly guided by the notion that if X has mean zero and variance σ^2 then $\Pr[X \ge a\sigma]$ should go like $e^{-a^2/2}$. There are times this idea is badly wrong. Consider Assumptions A.3 with all $p_i = 1/n$ so that $X = P_n - 1$ where P_n has the Binomial Distribution $B(n, 1/n)$ which is asymptotically P, the Poisson distribution with mean one. Then $E[X] = 0$ and $Var[X] \sim 1$. For a fixed $\Pr[X = a] \to \frac{1}{e(a+1)!}$ which is far bigger than $e^{-a^2/2}$. With this cautionary preamble, we give a general situation for which the notion is asymptotically correct when a is not too large.

Theorem A.1.19 *For every $C > 0$ and $\varepsilon > 0$ there exists $\delta > 0$ so that the following holds: Let X_i, $1 \le i \le n$, n arbitrary, be independent random variables with $E[X_i] = 0$, $|X_i| \le C$ and $Var[X_i] = \sigma_i^2$. Set $X = \sum_{i=1}^{n} X_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$ so that $Var[X] = \sigma^2$. Then for $0 < a \le \delta\sigma$*

$$\Pr[X > a\sigma] < e^{-\frac{a^2}{2}(1-\varepsilon)}.$$

Proof. We set $\lambda = a/\sigma$ so that $0 \le \lambda \le \delta$. Then

$$E[e^{\lambda X_i}] = \sum_{k=0}^{\infty} E[\frac{\lambda^k}{k!} X_i^k] = 1 + \frac{\lambda^2}{2}\sigma_i^2 + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!} E[X_i^k].$$

As $|X_i^k| \le C^{k-2} X_i^2$ we bound

$$E[X_i^k] \le E[|X_i^k|] \le C^{k-2} E[X_i^2] = C^{k-2}\sigma_i^2.$$

For $k \ge 3$ we bound $\frac{2}{k!} \le \frac{1}{(k-2)!}$ so that

$$E[e^{\lambda X_i}] \le 1 + \frac{\lambda^2}{2}\sigma_i^2 \left[1 + \sum_{k=3}^{\infty} \frac{(C\lambda)^{k-2}}{(k-2)!}\right] = 1 + \frac{\lambda^2}{2}\sigma_i^2 e^{\lambda C}.$$

We choose δ to satisfy $e^{C\delta} \le 1 + \varepsilon$. As $\lambda \le \delta$

$$E[e^{\lambda X_i}] \le 1 + \frac{\lambda^2}{2}\sigma_i^2(1+\varepsilon) < \exp\left[\frac{\lambda^2}{2}\sigma_i^2(1+\varepsilon)\right].$$

This inequality has held for all X_i so

$$E[e^{\lambda X}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] < \exp\left[\frac{\lambda^2}{2}\sigma^2(1+\varepsilon)\right]$$

and

$$\Pr[X > a\sigma] \leq E[e^{\lambda X}]e^{-\lambda a\sigma} < e^{-\frac{a^2}{2}(1-\varepsilon)}$$

■

A.2 EXERCISES

1. The *Hajós number* of a graph G is the maximum number k such that there are k vertices in G with a path between each pair so that all the $\binom{k}{2}$ paths are internally pairwise vertex disjoint (and no vertex is an internal vertex of a path and an endpoint of another). Is there a graph whose chromatic number exceeds twice its Hajós number ?

2. For two subsets A and B of the set Z_m of integers modulo m, and for a $g \in Z_m$, denote
$$s(A, B, g) = |\{(a, b) : a \in A, b \in B, a + b = g\}|.$$
For a partition of Z_m into two disjoint sets $Z_m = A \cup B$, $A \cap B = \emptyset$ denote
$$c(A, B) = max_{x \in Z_m}|s(A, A, x) + s(B, B, x) - 2s(A, B, x)|.$$
Prove that for every odd m there is a partition of Z_m into two disjoint sets A and B such that $c(A, B) = O(\sqrt{m \log m})$.

3. Let N be the standard normal random variable, with distribution function $f(t) = (2\pi)^{-1/2}e^{-t^2/2}$. Find $E[e^{\lambda N}]$ exactly. For $a > 0$ find
$\min_{\lambda>0} E[e^{\lambda N}]e^{-\lambda a}$ precisely. This gives the Chernoff Bound on $\Pr[N \geq a]$. How does this compare for a large with the actual $\Pr[N \geq a]$?

THE PROBABILISTIC LENS:

Triangle-free graphs have large independence numbers

Let $\alpha(G)$ denote the independence number of a graph G. It is easy and well known that for every graph G on n vertices with maximum degree d, $\alpha(G) \geq n/(d+1)$. Ajtai, Komlós and Szemerédi (1980) showed that in case G is triangle-free, this can be improved by a logarithmic factor and in fact $\alpha(G) \geq cn \log d/d$, where c is an absolute positive constant. Shearer (1983) simplified the proof and improved the constant factor to its best possible value $c = 1 + o(1)$. Here is a very short proof, without any attempts to optimize c, which is based on a different technique of Shearer (1995) and its modification in Alon (1996) .

Proposition 1 *Let $G = (V, E)$ be a triangle-free graph on n vertices with maximum degree at most $d \geq 1$. Then*

$$\alpha(G) \geq \frac{n \log d}{8d},$$

where the logarithm here and in what follows is in base 2.

Proof. If, say, $d < 16$ the result follows from the trivial bound $\alpha(G) \geq n/(d+1)$ and hence we may and will assume that $d \geq 16$. Let W be a random independent set of vertices in G, chosen uniformly among all independent sets in G. For each vertex $v \in V$ define a random variable $X_v = d|\{v\} \cap W| + |N(v) \cap W|$, where $N(v)$ denotes the set of all neighbors of v. We claim that the expectation of X_v satisfies $E(X_v) \geq \frac{\log d}{4}$.
To prove this claim, let H denote the induced subgraph of G on $V - (N(v) \cup \{v\})$, fix an independent set S in H and let X denote the set of all non-neighbors of S in

the set $N(v)$, $|X| = x$. It suffices to show that the conditional expectation

$$E(X_v \mid W \cap V(H) = S) \geq \frac{\log d}{4} \tag{1.1}$$

for each possible S. Conditioning on the intersection $W \cap V(H) = S$ there are precisely $2^x + 1$ possibilities for W: one in which $W = S \cup \{v\}$ and 2^x in which $v \notin W$ and W is the union of S with a subset of X. It follows that the conditional expectation considered in (1.1) is precisely $\frac{d}{2^x+1} + x2^{x-1}/(2^x+1)$. To check that the last quantity is at least $\log d/4$ observe that the assumption that this is false implies that $x \geq 1$ and $2^x(\log d - 2x) > 4d - \log d$, showing that $\log d > 2x \geq 2$ and hence $4d - \log d < \sqrt{d}(\log d - 2)$, which is false for all $d \geq 16$. Therefore,

$$E(X_v \mid W \cap V(H) = S) \geq \frac{\log d}{4},$$

establishing the claim.

By linearity of expectation we conclude that the expected value of the sum $\sum_{v \in V} X_v$ is at least $\frac{n \log d}{4}$. On the other hand, this sum is clearly at most $2d|W|$, since each vertex $u \in W$ contributes d to the term X_u in this sum, and its degree in G, which is at most d, to the sum of all other terms X_v. It follows that the expected size of W is at least $\frac{n \log d}{8d}$, and hence there is an independent set of size at least this expectation, completing the proof. ■

The *Ramsey Number* $r(3, k)$ is the minimum number r such that any graph with at least r vertices contains either a triangle or an independent set of size k. The asymptotic behaviour of this function has been studied for over fifty years. It turns out that $r(3, k) = \Theta(k^2/\log k)$. The lower bound is a recent result of Kim (1995) , based on a delicate probabilistic construction together with some thirty pages of computation. There is no known explicit construction of such a graph, and the largest known explicit triangle-free graph with no independent set of size k, described in Alon (1994) , has only $\Theta(k^{3/2})$ vertices. The tight upper bound for $r(3, k)$, proved in Ajtai et al. (1980) , is a very easy consequence of the above proposition.

Theorem 2 [Ajtai et al. (1980)] *There exists an absolute constant b such that $r(3, k) \leq bk^2/\log k$ for every $k > 1$.*

Proof. Let $G = (V, E)$ be a triangle-free graph on $8k^2/\log k$ vertices. If G has a vertex of degree at least k then its neighborhood contains an independent set of size k. Otherwise, by proposition 1 above, G contains an independent set of size at least $\frac{8k^2}{\log k} \frac{\log k}{8k} = k$. Therefore, in any case $\alpha(G) \geq k$, completing the proof. ■

Appendix B
Paul Erdős

Working with Paul Erdős was like taking a walk in the hills. Every time when I thought that we had achieved our goal and deserved a rest, Paul pointed to the top of another hill and off we would go.
– Fan Chung

B.1 PAPERS

Paul Erdős was the most prolific mathematician of the twentieth century, with over 1500 written papers and more than 490 collaborators. This highly subjective list gives only some of the papers that created and shaped the subject matter of this volume.

- A Combinatorial problem in geometry, *Compositio Math* **2** (1935), 463-470 (with George Szekeres) **Zbl.** 12,270.
  Written when Erdős was still a teenager this gem contains a rediscovery of Ramsey's Theorem and the Monotone Subsequence Theorem. Many authors have written that this paper played a key role in moving Erdős towards a more combinatorial view of mathematics.

- Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292-294, **MR** 8,479d; **Zbl** 32,192.

The three page paper that "started" the probabilistic method, giving an exponential lower bound on Ramsey $R(k,k)$ §1.1.

- The Gaussian law of errors in the theory of additive number theoretic functions, *Amer. J. Math.* **62** (1940), 738-742 (with Mark Kac) **MR** 2,42c; **Zbl.** 24,102.
  Showing that the number of prime factors of x chosen uniformly from 1 to n has an asymptotically normal distribution. A connection between probability and number theory that was extraordinary for its time. §4.2.

- Problems and results in additive number theory, *Colloque sur la Théorie des Nombres, Bruxelles, 1955*, pp. 127-137, George Thone, Liège; Masson and Cie, Paris, 1956; **MR** 18,18a; **Zbl.** 73,31.
  Using random subsets to prove the existence of a set of integers such that every n is represented $n = x + y$ at least once but at most $c \ln n$ times. Resolving a problem Sidon posed to Erdős in the 1930s. This problem continued to fascinate Erdős , see, e.g., Erdős and Tetali (1990) , §8.6.

- On a combinatorial problem, *Nordisk. Mat. Tidskr.* **11** (1963), 220-223 **MR** 28# 4068; **Zbl.** 122,248.
  On a combinatorial problem II., *Acta. Math. Acad. Sci. Hungar.* **15** (1964), 445-447; **MR** 29# 4700; **Zbl.** 201,337.
  Property B. Probabilistic proofs that any $m < 2^{n-1}$ n-sets can be two colored with no set monochromatic yet there exist $cn^2 2^n$ n-sets that cannot be so colored. §1.3.

- On the evolution of random graphs, *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17-61 (with Alfred Rényi); **MR** 23# A2338; **Zbl.** 103,163.
  Rarely in mathematics can an entire subject be traced to one paper. For Random Graphs this is the paper. Chapter 10.

- Graph theory and probability, *Canad. J. Math.* **11** (1959), 34-38; **MR** 21# 876; **Zbl.** 84,396.
  Proving by probabilistic methods the existence of graphs with arbitrarily high girth and chromatic number. This paper convinced many of the power of the methodology as the problem had received much attention but no construction had been found. Lens, following Chapter 3.

- Graph theory and probability II., *Canad. J. Math.* **13** (1961), 346-352 **MR** 22# 10925; **Zbl.** 97,391.
  Showing the existence of a triangle free graph on n vertices with no independent set of size $cn^{1/2} \ln n$ vertices, and hence that the Ramsey $R(3,k) = \Omega(k^2 \ln^{-2} k)$. A technical *tour de force* that uses probabilistic methods in a very subtle way, particularly considering the early date of publication.

- On circuits and subgraphs of chromatic graphs, *Mathematika* **9** (1962), 170-175; **MR** 25 # 3035; **Zbl.** 109,165.
  Destroying the notion that chromatic number is necessarily a local property,

Erdős proves the existence of a graph on n vertices that cannot be k-colored but for which every εn vertices can be three colored. Lens, following Chapter 8.

- On a combinatorial game, *J. Combinatorial Theory Ser. A* **14** (1973), 298-301 (with John Selfridge) **MR** 48# 5655; **Zbl.** 293,05004.
  Players alternate turns selecting vertices and the second player tries to stop the first from getting a winning set. The weight function method used was basically probabilistic and was an early use of derandomization. §15.1.

B.2 CONJECTURES

Conjectures were always an essential part of the mathematical life of Paul Erdős. Here are some of our favorites.

- Do sets of integers of positive density necessarily contain arithmetic progressions of arbitrary length? In finite form, is there for all k and all $\varepsilon > 0$, an n_0 so that if $n \geq n_0$ and S is a subset of the first n integers of size at least εn then S necessarily contains an arithmetic progression of length k? This conjecture was first made by Paul Erdős and Paul Turán in the 1930s. It was solved (positively) by Szemerédi in the 1970s. Let $F(k, \varepsilon)$ denote the minimal n_0 that suffices above. The growth rate of F remains an intriguing question with very recent results due to Gowers.

- Call distinct S, T, U a Δ-system if $S \cap T = S \cap U = T \cap U$. Let $F(n)$ be the minimal m such that given any m n-sets some three form a Δ-system. Erdős and Rado showed that $F(n)$ exists and gave the upper bound $F(n) \leq 2^n n!$. Erdős conjectured that $F(n) < C^n$ for some constant C.

- What are the asymptotics of the Ramsey function $R(k, k)$? In particular, what is the value c (if it exists) of $\lim_k R(k, k)^{1/k}$? The classic 1947 paper of Erdős gives $c \geq \sqrt{2}$ and $c \leq 4$ follows from the proof of Ramsey's Theorem but a half century has seen no further improvements in c, though there have been some results on lower order terms.

- Write $r_S(n)$ for the number of solutions to the equation $n = x + y$ with $x, y \in S$. Does there exist a set S of positive integers such that $r_S(n) > 0$ for all but finitely many n yet $r_S(n)$ is bounded by some constant K? The 1955 paper of Erdős referenced above gives S with $r_S(n) = \Theta(\ln n)$.

- Let $m(n)$, as defined in §1.3, denote the minimal size of a family of n-sets that cannot be two colored without forming a monochromatic set. What are the asymptotics of $m(n)$? In 1963 and 1964 Erdős found the bounds $\Omega(2^n) \leq m(n) = O(2^n n^2)$ and the lower bound of Radhakrishnan and Srinivasan shown in §3.5, is now $\Omega(2^n (n/\ln n)^{1/2})$.

- Given $2^{n-2}+1$ points in the plane, no three on a line, must some n of them form a convex set? This conjecture dates back to the 1935 paper of Erdős and Szekeres referenced above.
- Let $m(n,k,l)$ denote the size of the largest family of k-element subsets of an n-set such that no l-set is contained in more than one of them. Simple counting gives $m(n,k,l) \leq \binom{n}{l}/\binom{k}{l}$. Erdős and Haim Hanani conjectured in 1963 that for fixed $l < k$ this bound is asymptotically correct - that is, that the ratio of $m(n,k,l)$ to $\binom{n}{l}/\binom{k}{l}$ goes to one as $n \to \infty$. Erdős had a remarkable ability to select problems that were very difficult but not impossible. This conjecture was settled affirmatively by Vojtech Rödl in 1985, as discussed in §4.7. The asymptotics of the difference $\binom{n}{l}/\binom{k}{l} - m(n,k,l)$ remains open.

B.3 ON ERDŐS

There have been numerous books and papers written about the life and mathematics of Paul Erdős. Three deserving particular mention are

- The Mathematics of Paul Erdős (Ron Graham and Jarik Nešetřil, eds.), Berlin: Springer-Verlag, 1996. (Vols I. and II.)
- Combinatorics, Paul Erdős is Eighty (D. Miklós, V.T. Sós, T. Szönyi, eds.), Bolyai Soc. Math. Studies, Vol I (1990) and Vol II (1993).
- Erdős on Graphs - His Legacy of Unsolved Problems, Fan Chung and Ron Graham, A.K. Peters, 1998.

Of the many papers by mathematicians we note

- László Babai, In and out of Hungary: Paul Erdős, his friends, and times. In Combinatorics, Paul Erdős is Eighty (listed above), Vol II, 7-93.
- Béla Bollobás, Paul Erdős- Life and work, in The Mathematics of Paul Erdős (listed above), Vol. II, 1-42.
- A. Hajnal, Paul Erdős' Set theory, in: The mathematics of Paul Erdős(listed above), Vol. II, 352–393.
- János Pach, Two places at once: a remembrance of Paul Erdős, Math Intelligencer, Vol. 19 (1997), no. 2, 38-48.

Two popular biographies of Erdős have appeared

- The man who loved only numbers, Paul Hoffman, Hyperion (New York), 1998.
- My brain is open - The mathematical journies of Paul Erdős, Bruce Schechter, Simon & Schuster (New York), 1998.

Finally, George Csicsery has made a documentary film *N is a Number, A Portrait of Paul Erdős*, available from the publishers A. K. Peters, which allows one to see and hear Erdős in lecture and amongst friends, proving and conjecturing.

B.4 UNCLE PAUL

Paul Erdős died in September 1996 at the age of 83. His theorems and conjectures permeate this volume. This tribute[1]*, given by Joel Spencer at the National Meeting of the American Mathematical Society in January 1997, attempts to convey some of the special spirit that we and countless others took from this extraordinary man.*

Paul Erdős was a searcher, a searcher for mathematical truth.

Paul's place in the mathematical pantheon will be a matter of strong debate for in that rarefied atmosphere he had a unique style. The late Ernst Straus said it best, in a commemoration of Erdős's 70-th birthday.

> In our century, in which mathematics is so strongly dominated by "theory constructors" he has remained the prince of problem solvers and the absolute monarch of problem posers. One of my friends - a great mathematician in his own right - complained to me that "Erdős only gives us corollaries of the great metatheorems which remain unformulated in the back of his mind." I think there is much truth to that observation but I don't agree that it would have been either feasible or desirable for Erdős to stop producing corollaries and concentrate on the formulation of his metatheorems. In many ways Paul Erdős is the Euler of our times. Just as the "special" problems that Euler solved pointed the way to analytic and algebraic number theory, topology, combinatorics, function spaces, etc.; so the methods and results of Erdős's work already let us see the outline of great new disciplines, such as combinatorial and probabilistic number theory, combinatorial geometry, probabilistic and transfinite combinatorics and graph theory, as well as many more yet to arise from his ideas.

Straus, who worked as an assistant to Albert Einstein, noted that Einstein chose physics over mathematics because he feared that one would waste one's powers in persuing the many beautiful and attractive questions of mathematics without finding the central questions. Straus goes on,

> Erdős has consistently and successfully violated every one of Einstein's prescriptions. He has succumbed to the seduction of every beautiful problem he has encountered - and a great many have succumbed to him. This just proves to me that in the search for truth there is room for Don Juans like Erdős and Sir Galahads like Einstein.

I believe, and I'm certainly most prejudiced on this score, that Paul's legacy will be strongest in Discrete Math. Paul's interest in this area dates back to a marvelous paper with George Szekeres in 1935 but it was after World War II that it really flourished. The rise of the Discrete over the past half century has, I feel, two main causes. The first was The Computer, how wonderful that this physical object has led

[1]Reprinted with permission from the Bulletin of the American Mathematical Society.

to such intriguing mathematical questions. The second, with due respect to the many others, was the constant attention of Paul Erdős with his famous admonition “Prove and Conjecture!" Ramsey Theory, Extremal Graph Theory, Random Graphs, how many turrets in our mathematical castle were built one brick at a time with Paul’s theorems and, equally important, his frequent and always penetrating conjectures.

My own research specialty, The Probabilistic Method, could surely be called The Erdős Method. It was begun in 1947 with a 3 page paper in the Bulletin of the American Math Society. Paul proved the existence of a graph having certain Ramsey property without actually constructing it. In modern language he showed that an appropriately defined random graph would have the property with positive probability and hence there must exist a graph with the property. For the next twenty years Paul was a “voice in the wilderness", his colleagues admired his amazing results but adaption of the methodology was slow. But Paul persevered - he was always driven by his personal sense of mathematical aesthetics in which he had supreme confidence - and today the method is widely used in both Discrete Math and in Theoretical Computer Science.

There is no dispute over Paul’s contribution to the spirit of mathematics. Paul Erdős was the most inspirational man I have every met. I began working with Paul in the late 1960-s, a tumultuous time when “do your own thing" was the admonition that resonated so powerfully. But while others spoke of it, this was Paul’s modus operandi. He had no job; he worked constantly. He had no home; the world was his home. Possessions were a nuisance, money a bore. He lived on a web of trust, travelling ceaselessly from Center to Center, spreading his mathematical pollen.

What drew so many of us into his circle. What explains the joy we have in speaking of this gentle man. Why do we love to tell Erdős stories. I’ve thought a great deal about this and I think it comes down to a matter of belief, or faith. We mathematicians know the beauties of our subject and we hold a belief in its transcendent quality. God created the integers, the rest is the work of Man. Mathematical truth is immutable, it lies outside physical reality. When we show, for example, that two n-th powers never add to an n-th power for $n \geq 3$ we have discovered a Truth. This is our belief, this is our core motivating force. Yet our attempts to describe this belief to our nonmathematical friends are akin to describing the Almighty to an atheist. Paul embodied this belief in mathematical truth. His enormous talents and energies were given entirely to the Temple of Mathematics. He harbored no doubts about the importance, the absoluteness, of his quest. To see his faith was to be given faith. The religious world might better have understood Paul’s special personal qualities. We knew him as Uncle Paul.

I do hope that one cornerstone of Paul’s, if you will, theology will long survive. I refer to The Book. The Book consists of all the theorems of mathematics. For each theorem there is in The Book just one proof. It is the most aesthetic proof, the most insightful proof, what Paul called The Book Proof. And when one of Paul’s myriad conjectures was resolved in an “ugly" way Paul would be very happy in congratulating the prover but would add, “Now, let’s look for The Book Proof." This platonic ideal spoke strongly to those of us in his circle. The mathematics was there, we had only to discover it.

The intensity and the selflessness of the search for truth were described by the writer Jorge Luis Borges in his story The Library of Babel. The narrator is a worker in this library which contains on its infinite shelves all wisdom. He wanders its infinite corridors in search of what Paul Erdős might have called The Book. He cries out,

> To me, it does not seem unlikely that on some shelf of the universe there lies a total book. I pray the unknown gods that some man - even if only one man, and though it have been thousands of years ago! - may have examined and read it. If honor and wisdom and happiness are not for me, let them be for others. May heaven exist though my place be in hell. Let me be outraged and annihilated but may Thy enormous Library be justified, for one instant, in one being.

In the summer of 1985 I drove Paul to what many of us fondly remember as Yellow Pig Camp - a mathematics camp for talented high school students at Hampshire College. It was a beautiful day - the students loved Uncle Paul and Paul enjoyed nothing more than the company of eager young minds. In my introduction to his lecture I discussed The Book but I made the mistake of discribing it as being “held by God". Paul began his lecture with a gentle correction that I shall never forget. “You don’t have to believe in God," he said, “but you should believe in The Book."

Index

Author Index

References

Ahlswede, R. and Daykin, D. E. (1978). An inequality for the weights of two families of sets, their unions and intersections, *Z. Wahrscheinl. V. Geb* **43**: 183–185.

Aho, A. V., Hopcroft, J. E. and Ullman, J. D. (1974). *The Design and Analysis of Computer Algorithms*, Addison Wesley, Reading, MA.

Ajtai, M. (1983). Σ_1^1-formulae on finite structures, *Annals of Pure and Applied Logic* **24**: 1–48.

Ajtai, M., Chvátal, V., Newborn, M. M. and Szemerédi, E. (1982). Crossing-free subgraphs, *Theory and practice of combinatorics, North Holland Math. Stud.* **60**: 9–12.

Ajtai, M., Komlós, J. and Szemerédi, E. (1980). A note on Ramsey numbers, *J. Combinatorial Theory, Ser. A* **29**: 354–360.

Ajtai, M., Komlós, J. and Szemerédi, E. (1983). Sorting in $c \log n$ parallel steps, *Combinatorica* **3**: 1–19.

Ajtai, M., Komlós, J. and Szemerédi, E. (1987). Deterministic simulation in LOGSPACE, *Proc.* 19^{th} *annual ACM STOC*, New York, pp. 132–140.

Akiyama, J., Exoo, G. and Harary, F. (1981). Covering and packing in graphs IV: Linear arboricity, *Networks* **11**: 69–72.

Alon, N. (1986a). Eigenvalues and Expanders, *Combinatorica* **6**: 83–96.

Alon, N. (1986b). Eigenvalues, geometric expanders, sorting in rounds and Ramsey Theory, *Combinatorica* **6**: 207–219.

Alon, N. (1988). The linear arboricity of graphs, *Israel J. Math* **62**: 311–325.

Alon, N. (1990a). The maximum number of Hamiltonian paths in tournaments, *Combinatorica* **10**: 319–324.

Alon, N. (1990b). Transversal numbers of uniform hypergraphs, *Graphs and Combinatorics* **6**: 1–4.

Alon, N. (1994). Explicit Ramsey graphs and orthonormal labelings, *The Electronic J. Combinatorics* **1**: 8 pp. R12.

Alon, N. (1996). Independence numbers of locally sparse graphs and a Ramsey type problem, *Random Structures and Algorithms* **9**: 271–278.

Alon, N. and Boppana, R. B. (1987). The monotone circuit complexity of Boolean functions, *Combinatorica* **7**: 1–22.

Alon, N. and Chung, F. R. K. (1988). Explicit construction of linear sized tolerant networks, *Discrete Math.* **72**: 15–19.

Alon, N. and Frankl, P. (1985). The maximum number of disjoint pairs in a family of subsets, *Graphs and Combinatorics* **1**: 13–21.

Alon, N. and Kleitman, D. J. (1990). Sum-free subsets, *in: A tribute to Paul Erdős (A. Baker, B. Bollobás and A. Hajnál, eds.)*, Cambridge Univ. Press, Cambridge, England, pp. 13–26.

Alon, N. and Krivelevich, M. (1997). The concentration of the chromatic number of random graphs, *Combinatorica* **17**: 303–313.

Alon, N. and Linial, N. (1989). Cycles of length 0 modulo k in directed graphs, *J. Combinatorial Theory, Ser. B* **47**: 114–119.

Alon, N. and Milman, V. D. (1984). Eigenvalues, expanders and superconcentrators, *Proc.* 25^{th} *Annual FOCS*, IEEE, New York, pp. 320–322. See also: N. Alon and V. D. Milman, λ_1, isoperimetric inequalities for graphs and superconcentrators, J. Combinatorial Theory, Ser. B, 38, 1985, 73–88.

Alon, N., Babai, L. and Itai, A. (1986). A fast and simple randomized parallel algorithm for the maximal independent set problem, *J. of Algorithms* **7**: 567–583.

Alon, N., Frankl, P. and Rödl, V. (1985). Geometrical realization of set systems and probabilistic communication complexity, *Proc.* 26^{th} *FOCS*, IEEE, New York, pp. 277–280.

Alon, N., Goldreich, O., stad, J. H. and Peralta, R. (1990). Simple constructions of almost k-wise independent random variables, *Proc.* 31^{st} *FOCS, St. Louis*, IEEE, New York, pp. 544–553.

Alon, N., Kim, J. H. and Spencer, J. H. (1997). Nearly perfect matchings in regular simple hypergraphs, *Israel J. Math* **100**: 171–187.

Alon, N., Rónyai, L. and Szabó, T. (1999). Norm-graphs: variations and applications, *J. Combinatorial Theory, Ser. B* **76**: 280–290.

Andreev, A. E. (1985). On a method for obtaining lower bounds for the complexity of individual monotone functions, *Doklady Akademii Nauk SSSR* **282**(5): 1033–1037. (In Russian). English translation in Soviet Mathematics Doklady, 31:3, 530–534.

Andreev, A. E. (1987). On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes, *Vestnik Moskov. Univ. Ser. I Mat. Mekh* (1): 70–73. (In Russian).

B. Bollobás (1985). *Random Graphs*, Academic Press.

B. Bollobás (1988). The chromatic number of random graphs, *Combinatorica* **8**: 49–55.

Baik, J., Deift, P. and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations, *J. AMS* **12**: 1119–1178.

Bárány, I. and Fűredi, Z. (1987). Empty simplices in Euclidean Spaces, *Canad. Math. Bull.* **30**: 436–445.

Beck, J. (1978). On 3-Chromatic Hypergraphs, *Disc. Math.* **24**: 127–137.

Beck, J. (1981). Roth's estimate of the discrepancy of integer sequences is nearly optimal, *Combinatorica* **1**: 319–325.

Beck, J. (1991). An Algorithmic Approach to the Lovász Local Lemma. I., *Random Structures and Algorithms* **2**: 343–365.

Beck, J. and Fiala, T. (1981). Integer-making Theorems, *Disc. Appl. Math.* **3**: 1–8.

Bernstein, S. N. (1912). Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, *Comm. Soc. Math. Kharkov* **13**: 1–2.

Blum, N. (1984). A Boolean function requiring $3n$ network size, *Theoretical Computer Science* **28**: 337–345.

Bollobás, B. (1965). On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16**: 447–452.

Bollobás, B. and Erdős, P. (1976). Cliques in Random Graphs, *Math. Proc. Camb. Phil. Soc.* **80**: 419–427.

Boppana, R. B. and Spencer, J. H. (1989). A useful elementary correlation inequality, *J. Combinatorial Theory, Ser. A* **50**: 305–307.

Brégman, L. M. (1973). Some properties of nonnegative matrices and their permanents, *Soviet. Math. Dokl.* **14**: 945–949.

Chazelle, B. and Welzl, E. (1989). Quasi-optimal range searching in spaces of finite VC-dimension, *Discrete and Computational Geometry* **4**: 467–489.

Chernoff, H. (1952). A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23**: 493–509.

Chung, F. R. K., Frankl, P., Graham, R. L. and Shearer, J. B. (1986). Some intersection theorems for ordered sets and graphs, *J. Combinatorial Theory, Ser. A* **43**: 23–37.

Chung, F. R. K., Graham, R. L. and Wilson, R. M. (1989). Quasi-random graphs, *Combinatorica* **9**: 345–362.

Cohen, A. and Wigderson, A. (1989). Dispersers, deterministic amplification, and weak random sources, *Proc.* 30^{th} *IEEE FOCS*, IEEE, New York, pp. 14–19.

Danzer, L. and Grűnbaum, B. (1962). Ű̈ber zwei Probleme bezűglich konvexer Kőrper von P. Erdős und von V. L. Klee, *Math. Z.* **79**: 95–99.

de la Vega, W. F. (1983). On the maximal cardinality of a consistent set of arcs in a random tournament, *J. Combinatorial Theory, Ser. B* **35**: 328–332.

Dudley, R. M. (1978). Central limit theorems for empirical measures, *Ann. Probab.* **6**: 899–929.

Elekes, G. (1997). On the number of sums and products, *Acta Arith.* **81**: 365–367.

Erdős, P. (1947). Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53**: 292–294.

Erdős, P. (1956). Problems and results in additive number theory, *Colloque sur le Théorie des Nombres (CBRM, Bruselles)* pp. 127–137.

Erdős, P. (1959). Graph theory and probability, *Canad. J. Math.* **11**: 34–38.

Erdős, P. (1962). On Circuits and Subgraphs of Chromatic Graphs, *Mathematika* **9**: 170–175.

Erdős, P. (1963a). On a combinatorial problem, I, *Nordisk Mat. Tidskr.* **11**: 5–10.

Erdős, P. (1963b). On a problem of graph theory, *Math. Gaz.* **47**: 220–223.

Erdős, P. (1964). On a combinatorial problem II, *Acta Math. Acad. Sci. Hungar.* **15**: 445–447.

Erdős, P. (1965a). Extremal problems in number theory, *Proc. Symp. Pure Math. (AMS)* **VIII**: 181–189.

Erdős, P. (1965b). On Extremal Problems of Graphs and Generalized Graphs, *Israel J. Math.* **2**: 189–190.

Erdős P. and Hanani, H. (1963). On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10**: 10–13.

Erdős P. and Kac, M. (1940). The Gaussian law of errors in the theory of additive number theoretic functions, *Amer. J. Math.* **62**: 738–742.

Erdős P. and Lovász, L. (1975). Problems and results on **3**-chromatic hypergraphs and some related questions, *in: Infinite and Finite Sets (A. Hajnal et al., eds.)*, North-Holland, Amsterdam, pp. 609–628.

Erdős P. and Moon, J. W. (1965). On sets of consistent arcs in a tournament, *Canad. Math. Bull.* **8**: 269–271.

Erdős P. and Rényi, A. (1960). On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**: 17–61.

Erdős P. and Selfridge, J. L. (1973). On a combinatorial game, *J. Combinatorial Theory, Ser. A* **14**: 298–301.

Erdős P. and Spencer, J. H. (1991). Lopsided Lovász Local Lemma and Latin transversals, *Discrete Appl. Math.* **30**: 151–154.

Erdős P. and Tetali, P. (1990). Representations of integers as the sum of k terms, *Random Structures and Algorithms* **1**: 245–261.

Erdős P. and Z. Füredi (1983). The greatest angle among n points in the d-dimensional Euclidean space, *Annals of Discrete Math.* **17**: 275–283.

Fagin, R. (1976). Probabilities in finite models, *J. Symbolic Logic* **41**: 50–58.

Fishburn, P. (1992). Correlation in partially ordered sets, *Discrete Applied Math.* **39**: 173–191.

Fortuin, C. M., Kasteleyn, P. W. and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets, *Comm. Math. Phys.* **22**: 89–103.

Frankl, P. and Wilson, R. M. (1981). Intersection theorems with geometric consequences, *Combinatorica* **1**: 357–368.

Frankl, P., Rödl, V. and Wilson, R. M. (1988). The number of submatrices of given type in a Hadamard matrix and related results, *J. Combinatorial Theory, Ser. B* **44**: 317–328.

Furst, M., Saxe, J. and Sipser, M. (1984). Parity, circuits and the polynomial hierarchy, *Mathematical Systems Theory* **17**: 13–27.

Glebskii, Y. V., Kogan, D. I., Liagonkii, M. I. and Talanov, V. A. (1969). Range and degree of realizability of formulas the restricted predicate calculus, *Cybernetics* **5**: 142–154. (Russian original: Kibernetica **5**, 17–27).

Graham, R. L. and Spencer, J. H. (1971). A constructive solution to a tournament problem, *Canad. Math. Bull.* **14**: 45–48.

Graham, R. L., Rothschild, B. L. and Spencer, J. H. (1990). *Ramsey Theory, second edition*, John Wiley, New York.

Halberstam, H. and Roth, K. F. (1983). *Sequences, second edition*, Springer Verlag, Berlin.

Hall, M. (1986). *Combinatorial Theory, second edition*, Wiley, New York.

Harper, L. (1966). Optimal numberings and isoperimetric problems on graphs, *J. Combinatorial Theory* **1**: 385–394.

Harris, T. E. (1960). Lower bound for the critical probability in a certain percolation process, *Math. Proc. Cambridge Phil. Soc.* **56**: 13–20.

Haussler, D. (1995). Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension, *J. Combinatorial Theory, Ser. A* **69**: 217–232.

Haussler, D. and Welzl, E. (1987). ϵ-nets and simplex range queries, *Discrete and Computational Geometry* **2**: 127–151.

J. Håstad (1988). Almost optimal lower bounds for small depth circuits, *in* S. Micali (ed.), *Advances in Computer Research*, JAI Press, chapter 5: Randomness and Computation, pp. 143–170.

J. Håstad (1998). The shrinkage exponent of De Morgan formulas is 2, *SIAM J. Comput.* **27**: 48–64.

Janson, S. (1990). Poisson Approximation for Large Deviations, *Random Structures and Algorithms* **1**: 221–230.

Janson, S. (1998). New versions of Suen's correlation inequality, *Random Structures and Algorithms* **13**: 467–483.

Janson, S., Knuth, D., Łuczak, T. and Pittel, B. (1993). The birth of the giant component, *Random Structures and Algorithms* **4**: 233–358.

Joffe, A. (1974). On a set of almost deterministic k-independent random variables, *Ann. Probability* **2**: 161–162.

Kahn, J. (1996). Asymptotically good list colorings, *J. Combinatorial Theory, Ser. A* **73**: 1–59.

Karchmer, M. and Wigderson, A. (1990). Monotone circuits for connectivity require super-logarithmic depth, *SIAM J. Disc. Math.* **3**: 255–265.

Karp, R. M. (1990). The transitive closure of a Random Digraph, *Random Structures and Algorithms* **1**: 73–94.

Karp, R. M. and Ramachandran, V. (1990). Parallel algorithms for shared memory machines, *in: Handbook of Theoretical Computer Science (J. Van Leeuwen ed.), Vol. A, Chapter 17*, Elsevier, New York, pp. 871–941.

Katchalski, M. and Meir, A. (1988). On empty triangles determined by points in the plane, *Acta Math. Hungar.* **51**: 323–328.

Katona, G. O. H. (1972). A simple proof of the Erdős Ko-Rado Theorem, *J. Combinatorial Theory, Ser. B* **13**: 183–184.

Khrapchenko, V. M. (1971). A method of determining lower bounds for the complexity of Π-schemes, *Matematischi Zametki* **10**(1): 83–92. (In Russian.) English translation in Mathematical Notes of the Academy of Sciences of the USSR, 11, 1972, 474–479.

Kim, J. and Vu, V. (to appear). Concentration of Multivariate Polynomials and its Applications.

Kim, J. H. (1995). The Ramsey number $R(3, t)$ has order of magnitude $t^2/\log t$, *Random Structures and Algorithms* **7**: 173–207.

Kleitman, D. J. (1966a). On a combinatorial problem of Erdős, *J. Combinatorial Theory* **1**: 209–214.

Kleitman, D. J. (1966b). Families of non-disjoint subsets, *J. Combinatorial Theory* **1**: 153–155.

Kleitman, D. J., Shearer, J. B. and Sturtevant, D. (1981). Intersection of k-element sets, *Combinatorica* **1**: 381–384.

Kolountzakis, M. N. (1999). An effective additive basis for the integers, *Discrete Mathematics* **145**: 307–313.

Komlós, J., Pach, J. and Woeginger, G. (1992). Almost tight bounds on epsilon-nets, *Discrete Comput. Geom.* **7**: 163–173.

Komlós, J., Pintz, J. and Szemerédi, E. (1982). A lower bound for Heilbronn's problem, *J. London Math. Soc.* **25**(2): 13–24.

Loomis, L. H. and Whitney, H. (1949). An inequality related to the isoperimetric inequality, *Bull. Amer. Math. Soc.* **55**: 961–962.

Lovász, L., Spencer, J. H. and Vesztergombi, K. (1986). Discrepancy of set systems and matrices, *Europ. J. Comb.* **7**: 151–160.

Lubotzky, A., Phillips, R. and Sarnak, P. (1986). Explicit expanders and the Ramanujan conjectures, *Proc.* 18^{th} *ACM STOC*, pp. 240–246. See also: A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, Combinatorica 8, 1988, 261–277.

Łuczak, T. (1990). Component behavior near the critical point of the random graph process, *Random Structures and Algorithms* **1**: 287–310.

Łuczak, T. (1991). A note on the sharp concentration of the chromatic number of random graphs, *Combinatorica* **11**: 295–297.

MacWilliams, F. J. and Sloane, N. J. A. (1977). *The Theory of Error Correcting Codes*, North Holland, Amsterdam.

Mani-Levitska, P. and Pach, J. (1988). Decomposition problems for multiple coverings with unit balls, manuscript.

Margulis, G. A. (1973). Explicit constructions of concentrators, *Problemy Peredachi Informatsii* **9**: 71–80. (In Russian). English translation in Problems of Information Transmission 9, 325–332.

Margulis, G. A. (1988). Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and superconcentrators, *Problemy Peredachi Informatsii* **24**: 51–60. (In Russian.) English translation in Problems of Information Transmission 24, 1988, 39–46.

Marica, J. and Schonheim, J. (1969). Differences of sets and a problem of Graham, *Canad. Math. Bull.* **12**: 635–637.

Matoušek, J. (1997). On discrepancy bounds via dual shatter function, *Mathematika* **44**(1): 42–49.

Matoušek, J., Welzl, E. and Wernisch, L. (1993). Discrepancy and approximation for bounded VC dimension, *Combinatorica* **13**: 455–466.

Matula, D. W. (1976). The largest clique size in a random graph, *Technical report*, Southern Methodist University, Dallas.

Maurey, B. (1979). Construction de suites symétriques, *Compt. Rend. Acad. Sci. Paris* **288**: 679–681.

Milman, V. D. and Schechtman, G. (1986). *Asymptotic Theory of Finite Dimensional Normed Spaces, Lecture Notes in Mathematics*, Vol. 1200, Springer Verlag, Berlin and New York.

Moon, J. W. (1968). *Topics on Tournaments*, Holt, Reinhart and Winston, New York.

Nakayama, A. and Peroche, B. (1987). Linear arboricity of digraphs, *Networks* **17**: 39–53.

Naor, J. and Naor, M. (1990). Small-bias probability spaces: efficient constructions and applications, *Proc.* 22^{nd} *annual ACM STOC*, ACM Press, pp. 213–223.

Nilli, A. (1991). On the second eigenvalue of a graph, *Discrete Mathematics* **91**: 207–210.

Pach, J. and Agarwal, P. K. (1995). *Combinatorial Geometry*, J. Wiley and Sons, New York.

Pach, J. and Woeginger, G. (1990). Some new bounds for epsilon-nets, *Proc.* 6^{th} *Annual Symposium on Computational Geometry*, ACM Press, New York, pp. 10–15.

Paturi, R. and Simon, J. (1984). Probabilistic communication complexity, *Proc.* 25^{th} *FOCS*, IEEE, New York, pp. 118–126.

Paul, W. J. (1977). A $2.5n$ lower bound on the combinational complexity of Boolean functions, *SIAM Journal on Computing* **6**: 427–443.

Pinsker, M. (1973). On the complexity of a concentrator, 7^{th} *Internat. Teletraffic Conf.*, Stockholm, pp. 318/1–318/4.

Pippenger, N. and Spencer, J. H. (1989). Asymptotic behaviour of the chromatic index for hypergraphs, *J. Combinatorial Theory, Ser. A* **51**: 24–42.

Rabin, M. O. (1980). Probabilistic algorithms for testing primality, *J. Number Theory* **12**: 128–138.

Radhakrishnan, J. and Srinivasan, A. (2000). Improved bounds and algorithms for hypergraph two-coloring, *Random Structures and Algorithms* **16**: 4–32.

Raghavan, P. (1988). Probabilistic construction of deterministic algorithms: approximating packing integer programs, *J. of Computer and Systems Sciences* **37**: 130–143.

Ramsey, F. P. (1929). On a problem of formal logic, *Proc. London Math. Soc.* **30**(2): 264–286.

Razborov, A. A. (1985). Lower bounds on the monotone complexity of some Boolean functions, *Doklady Akademii Nauk SSSR* **281**(4): 798–801. (In Russian.) English translation in Soviet Mathematics Doklady 31, 354–357.

Razborov, A. A. (1987). Lower bounds on the size of bounded depth networks over a complete basis with logical addition, *Matematischi Zametki* **41**(4): 598–607. (In Russian.) English translation in Mathematical Notes of the Academy of Sciences of the USSR, 41, 4, 333–338.

Rödl, V. (1985). On a packing and covering problem, *European Journal of Combinatorics* **6**: 69–78.

Sauer, N. (1972). On the density of families of sets, *J. Combinatorial Theory, Ser. A* **13**: 145–147.

Schrijver, A. (1978). A short proof of Minc's conjecture, *J. Combinatorial Theory, Ser. A* **25**: 80–83.

Shamir, E. and Spencer, J. H. (1987). Sharp concentration of the chromatic number in random graphs $G_{n,p}$, *Combinatorica* **7**: 121–130.

Shearer, J. B. (1983). A note on the independence number of triangle-free graphs, *Discrete Math* **46**: 83–87.

Shearer, J. B. (1985). On a problem of Spencer, *Combinatorica* **5**: 241–245.

Shearer, J. B. (1995). On the independence number of sparse graphs, *Random Structures and Algorithms* **7**: 269–271.

Shelah, S. and Spencer, J. H. (1988). Zero-One Laws for Sparse Random Graphs, *J. Amer. Math. Soc.* **1**: 97–115.

Shepp, L. A. (1982). The XYZ-conjecture and the FKG-inequality, *Ann. of Probab.* **10**: 824–827.

Smolensky, R. (1987). Algebraic methods in the theory of lower bounds for Boolean circuit complexity, *Proceedings of the 19th ACM STOC*, ACM Press, New York, pp. 77–82.

Spencer, J. H. (1977). Asymptotic lower bounds for Ramsey functions, *Disc. Math.* **20**: 69–76.

Spencer, J. H. (1985a). Six Standard Deviations Suffice, Trans. Amer. Math. Soc., **289**: 679–706.

Spencer, J. H. (1985b). Probabilistic methods, *Graphs and Combinatorics* **1**: 357–382.

Spencer, J. H. (1987). *Ten Lectures on the Probabilistic Method*, SIAM, Philadelphia.

Spencer, J. H. (1990a). Threshold functions for extension statements, *J. Combinatorial Theory, Ser. A* **53**: 286–305.

Spencer, J. H. (1990b). Counting Extensions, *J. Combinatorial Theory, Ser. A* **55**: 247–255.

Spencer, J. H. (1995). Asymptotic packing via a branching process, *Random Structures and Algorithms* **7**: 167–172.

Subbotovskaya, B. A. (1961). Realizations of linear functions by formulas using $+, \cdot, -$, *Doklady Akademii Nauk SSSR* **136**(3): 553–555. (In Russian.) English translation in Soviet Mathematics Doklady, 2, 110–112.

Suen, W. C. (1990). A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph, *Random Structures and Algorithms* **1**: 231–242.

Székely, L. (1997). Crossing numbers and hard Erdős problems in discrete geometry, *Combin. Probab. Comput.* **6**: 353–358.

Szele, T. (1943). Kombinatorikai vizsgálatok az irányitott teljes gráffal kapcsolatban, *Mat. Fiz. Lapok 50* pp. 223–256. For a German translation see: T. Szele, Publ. Math. Debrecen, 13, 1966, 145–168.

Talagrand, M. (1996). Concentration of measures and isopermetric inequalites in product spaces, *Publications Mathematiques de l'I.H.E.S.* **81**: 73–205.

Tanner, R. M. (1984). Explicit construction of concentrators from generalized N-gons, *SIAM J. Alg. Disc. Meth.* **5**: 287–293.

Tarjan, R. E. (1983). *Data Structures and Network Algorithms*, SIAM, Philadelphia.

Thomason, A. (1987). Pseudo-random graphs, *Annals of Discrete Math.* **33**: 307–331.

Turán, P. (1934). On a theorem of Hardy and Ramanujan, *J. London Math Soc.* **9**: 274–276.

Turán, P. (1941). On an extremal problem in Graph Theory, *Mat. Fiz. Lapok* **48**: 436–452.

Valtr, P. (1995). On the minimum number of empty polygons in planar point sets, *Studia Sci. Math. Hungar.* **30**: 155–163.

Vapnik, V. N. and Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* **16**: 264–280.

Wegener, I. (1987). *The Complexity of Boolean Functions*, Wiley-Teubner, New York.

Weil, A. (1948). *Sur les courbes algébriques et les variéstés qui sèn déduisent*, Actualités Sci. Ind, no. 1041. iv+85pp.

Wendel, J. G. (1962). A problem in geometric probability, *Math. Scand.* **11**: 109–111.

Wright, E. M. (1977). The number of connected sparsely edged graphs, *Journal of Graph Theory* **1**: 317–330.

Yao, A. C. (1985). Separating the polynomial-time hierarchy by oracles, *Proceedings of the 26th Annual IEEE FOCS*, IEEE, New York, pp. 1–10.

Z. Fűredi (1988). Matchings and covers in hypergraphs, *Graphs and Combinatorics* **4**: 115–206.