The Probabilistic Method

Third Edition

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

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

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

This Page Intentionally Left Blank

Preface

The Probabilistic Method is one of the most powerful and widely used tools applied in combinatorics. One of the major reasons for its rapid development is the important role of randomness in theoretical computer science and in statistical physics.

The interplay between discrete mathematics 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 book 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 usually be 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 was initiated by Paul Erdös, who contributed so much to its development over a fifty year period, that it seems appropriate to call it "The Erdös Method." His contribution can be measured not 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 $f = (1 + o(1))g$; that is, 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 (*). The exercises enable readers to check their understanding of the material and also provide the possibility of using the book as a textbook.

This is the third edition of the book; it contains several improved results and covers various additional topics that developed extensively during the last few years. The additions include a modern treatment of the Erdös–Rényi phase transition discussed in Chapter 11, focusing on the behavior of the random graph near the emergence of the giant component and briefly exploring its connection to classical percolation theory. Another addition is Chapter 17, Graph Property Testing—a recent topic that combines combinatorial, probabilistic and algorithmic techniques. This chapter also includes a proof of the Regularity Lemma of Szemerédi (described in a probabilistic language) and a presentation of some of its applications in the area. Further additions are two new *Probabilistic Lenses,* several additional exercises, and a new part in Appendix A focused on lower bounds.

It is a special pleasure to thank our wives, Nurit and Mary Ann. Their patience, understanding and encouragement have been key ingredients 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 third edition by joint research, helpful discussions and useful comments. These include Miklos Bona, Andrzej Dudek, Mathieu Dutour, Juliana Freire, Sariel Har-Peled, Johan Hastad, Rani Hod, Mihyun Kang, Michael Krivelevich, Eyal Lubetzky, Russell Lyons, Nabil Mustafa, Nathan Linial, Yuval Peres, Xue Rui, Alexander Sapozhenko, Asaf Shapira, Aravind Srinivasan, Benny Sudakov, Prasad Tetali and William Wu, who pointed out various inaccuracies and misprints, and suggested improvements in the presentation as well as in the results. Needless to say, the responsibility for the remaining mistakes, as well as the responsibility for the (hopefully not many) new ones, is solely ours.

It is a pleasure to thank Rani Hod and Eyal Lubetzky for their great technical help in the preparation of the final manuscript for this book.

This Page Intentionally Left Blank

Part I

METHODS

This Page Intentionally Left Blank

1

The Basic Method

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

1.1 THE PROBABILISTIC METHOD

The probabilistic method is a powerful tool for tackling many problems in discrete mathematics. Roughly speaking, the method works as follows: Trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this 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* $k \geq 3$.

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

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; that is, $R(k,k) > n$. Note that if $k \geq 3$ and we take $n = \lfloor 2^{k/2} \rfloor$ then

$$\binom{n}{k}2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{k^2/2}} < 1$$

and hence $R(k,k) > \lfloor 2^{k/2} \rfloor$ for all $k \geq 3$. ■

This simple example demonstrates the essence of the probabilistic method. To prove the existence of a good coloring we do not present one explicitly, but rather show, in a nonconstructive way, that it exists. This example appeared in a paper of P. Erdős from 1947. Although Szele had applied the probabilistic method to another combinatorial problem, mentioned in Chapter 2, already in 1943, Erdős was certainly the first one who understood the full power of this method and applied it successfully over the years to numerous problems. One can, of course, claim that the probability is not essential in the proof given above. An equally simple proof can be described by counting; we just check that the total number of two-colorings of K_n is larger 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, for example, 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 that is exponential in the size $[= \binom{n}{2}]$ of the problem. Algorithms whose running time is more than polynomial in the size of the problem are usually considered impractical. The class of problems that can be solved in polynomial time, usually denoted by **P** [see, e.g., Aho, Hopcroft and Ullman (1974)], is, in a sense, the class of all solvable problems. In this sense, the exhaustive search approach suggested above for finding a good coloring of K_n is not acceptable, and this is the reason for our remark that the proof of Proposition 1.1.1 is nonconstructive; it does not supply a constructive, efficient and deterministic way of

producing a coloring with the desired properties. However, a closer look at the proof shows that, in fact, it can be used to produce, effectively, a coloring that is very likely to be good. This is because for large k, if $n = \lfloor 2^{k/2} \rfloor$ then

$$\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \left(\frac{n}{2^{k/2}}\right)^k \leq \frac{2^{1+\frac{k}{2}}}{k!} \ll 1 .$$

Hence, a random coloring of K_n is very likely not to contain a monochromatic $K_{2\log n}$. This means that if, for some reason, we *must* present a two-coloring of the edges of K_{1024} without a monochromatic K_{20} we can simply produce a random two-coloring by flipping a fair coin $\binom{1024}{2}$ times. We can then deliver the resulting coloring safely; the probability that it contains a monochromatic K_{20} is less than $2^{11}/20!$, probably much smaller than our chances of making a mistake in any rigorous proof that a certain coloring is good! Therefore, in some cases the probabilistic, nonconstructive method does supply effective probabilistic algorithms. Moreover, these algorithms can sometimes be converted into deterministic ones. This topic is discussed in some detail in Chapter 16.

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; that is, the probability space considered is symmetric. It is worth noting that we often use in applications symmetric probability spaces. In these cases, we shall sometimes refer to an element of the space as a *random element*, without describing explicitly the probability distribution. Thus, for example, in the proof of Proposition 1.1.1 random two-colorings of K_n were considered; that is, 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 that 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; that is, 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 $\binom{n}{k} < (en/k)^k$ and $(1-2^{-k})^{n-k} < e^{-(n-k)/2^k}$, Theorem 1.2.1 implies that $f(k) \leq k^2 \cdot 2^k \cdot (\ln 2)(1+o(1))$. 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* $n[1+\ln(\delta+1)]/(\delta+1)$ *vertices.*

Proof. Let $p \in [0,1]$ be, for the moment, arbitrary. Let us pick, randomly and independently, each vertex of V with probability p. Let X be the (random) set of all vertices picked and let $Y = Y_X$ be the random set of all vertices in $V - X$ that do not have any neighbor in X. The expected value of $|X|$ is clearly np. For each fixed vertex $v \in V$, $\Pr[v \in Y] = \Pr[v \text{ and its neighbors are not in } X] \leq (1-p)^{\delta+1}$. Since the expected value of a sum of random variables is the sum of their expectations (even

if they are not independent) and since the random variable $|Y|$ can be written as a sum of n indicator random variables χ_v ($v \in V$), where $\chi_v = 1$ if $v \in Y$ and $\chi_v = 0$ otherwise, we conclude that the expected value of $|X| + |Y|$ is at most $np + n(1-p)^{\delta+1}$. Consequently, there is at least one choice of $X \subseteq V$ such that $|X| + |Y_X| \leq np + n(1-p)^{\delta+1}$. The set $U = X \cup Y_X$ is clearly a dominating set of G whose cardinality is at most this size.

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

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

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

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

Formally, we set p equal to this value in the first line of the proof. We now have $|U| \leq n[1 + \ln(\delta+1)]/(\delta+1)$ 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 perhaps more subtle and is an example of the "alteration" principle that 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 that 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 $n \ln \delta/\delta$.

It can easily be 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 neighbors. 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 $r(1-(\delta+1)/n)$. 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 $n\ln(\delta+1)/(\delta+1)$ steps there will be at most $n/(\delta+1)$ yet uncovered vertices that can now be added to the set of chosen vertices to form a dominating set of size at most equal to the one in the conclusion of Theorem 1.2.2.

Combining this with some ideas of Podderyugin and Matula, we can obtain a very efficient algorithm to decide if a given undirected graph on n vertices is, say, $n/2$ 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; that is, 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 time $O(n^2)$. We now find, for each i, $2\leq i\leq k$, the minimum size s_i of a cut that separates u_1 from u_i. Each of these problems can be solved by solving a standard network flow problem in time $O(n^{8/3})$ [see, e.g., Tarjan (1983)]. By Lemma 1.2.3 the edge connectivity of G is simply the minimum between δ and $\min_{2\leq i\leq k} s_i$. The total time of the algorithm is $O(n^{8/3}\log n)$, as claimed.

1.3 COMBINATORICS

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

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

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

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

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

In Section 3.5 we present a more delicate argument, due to Radhakrishnan and Srinivasan, and based on an idea of Beck, that shows that

$$m(n) \geq \Omega\left(\left(\frac{n}{\ln n}\right)^{1/2} 2^n\right).$$

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

$$\Pr[S \text{ is monochromatic under } \chi] \geq p,$$

where we set

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

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

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

There are 2^v colorings so

$$\Pr\left[\bigvee_\chi A_\chi\right] \leq 2^v (1-p)^m .$$

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

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

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

then $2^v(1-p)^m < 2^v e^{-pm} \leq 1$ so $m(n) \leq m$. Now we need to find v to minimize v/p. We may interpret p as twice the probability of picking n white balls from an urn with $v/2$ white and $v/2$ black balls, sampling without replacement. It is tempting to estimate p by 2^{-n+1}, the probability for sampling with replacement. This approximation would yield $m \sim v2^{n-1}(\ln 2)$. As v gets smaller, however, the approximation becomes less accurate and, as we wish to minimize m, the trade-off 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

$$\frac{v-2i}{v-i} = 1 - \frac{i}{v} + O\left(\frac{i^2}{v^2}\right) = e^{-i/v + O(i^2/v^2)} .$$

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

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 $1 \leq i \leq h$, $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ for all distinct i, j with $1 \leq i, j \leq h$. 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* $h \leq \binom{k+\ell}{k}$.

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

$$1 \geq \Pr\left[\bigvee_{i=1}^{h} X_i\right] = \sum_{i=1}^{h} \Pr[X_i] = h/\binom{k+\ell}{k},$$

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 $(A + A) \cap A = \emptyset$; that is, 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 $p > 2\max\{|b_i|\}_{i=1}^{n}$ and put $C = \{k+1, k+2, \ldots, 2k+1\}$. Observe that C is a sum-free subset of the cyclic group $\mathbb{Z}_p$ and that

$$\frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.$$

Let us choose at random an integer x, $1 \leq x < p$, according to a uniform distribution on $\{1, 2, \ldots, p-1\}$, and define $d_1, \ldots, d_n$ by $d_i \equiv xb_i \pmod{p}$, $0 \leq d_i < p$. Trivially, for every fixed i, $1 \leq i \leq n$, as x ranges over all numbers $1, 2, \ldots, p-1$, d_i ranges over all nonzero elements of $\mathbb{Z}_p$ and hence $\Pr[d_i \in C] = |C|/(p-1) > \frac{1}{3}$. Therefore the expected number of elements b_i such that $d_i \in C$ is more than $n/3$. Consequently, there is an x, $1 \leq x < p$ and a subsequence A of B of cardinality $|A| > n/3$, such that $xa \pmod{p} \in C$ 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 $xa_1 + xa_2 \equiv xa_3 \pmod{p}$, contradicting the fact that C is a sum-free subset of $\mathbb{Z}_p$. This completes the proof. ■

Remark. The above proof works whenever p is a prime that does not divide any of the numbers b_i. This can be used to design an efficient deterministic algorithm for finding a sum-free subset A of size bigger than $|B|/3$ in a given set B as above. 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; that is,

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

Daykin and Erdős conjectured that if $m = 2^{(1/2+\delta)n}$, 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 $m = 2^{(1/2+\delta)n}$ subsets of $X = \{1, 2, \ldots, n\}$, where $\delta > 0$. Then*

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

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

In fact,

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

Define

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

Clearly,

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

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

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

Since $Y \leq m$ we conclude that

$$\Pr\left[Y \geq m^{1-t\delta^2/2} \right] \geq m^{-t\delta^2/2} . \tag{1.3}$$

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 (1.3) is greater than the right-hand side of (1.2). Thus, with positive probability, $|A_1 \cup A_2 \cup \cdots \cup A_t| > n/2$ and still this union is disjoint to more than $2^{n/2}$ members of F. This contradiction implies inequality (1.1). ■

1.6 EXERCISES

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

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

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

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

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

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

$$\Pr[|X - Y| \leq 2] \leq 3\Pr[|X - Y| \leq 1] .$$

4. (*) Let $G = (V, E)$ be a graph with n vertices and minimum degree $\delta > 10$. Prove that there is a partition of V into two disjoint subsets A and B so that

$|A| \leq O(n \ln \delta/\delta)$, 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 $h \leq (k + l)^{k+l}/(k^k l^l)$.

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

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

This Page Intentionally Left Blank

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, $X = c_1X_1 + \cdots + c_nX_n$. Linearity of expectation states that

$$\mathrm{E}[X] = c_1\mathrm{E}[X_1] + \cdots + c_n\mathrm{E}[X_n] .$$

The power of this principle comes from there being no restrictions on the dependence or independence of the X_i. In many instances $\mathrm{E}[X]$ can easily be 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 $\mathrm{E}[X]$ we decompose $X = X_1 + \cdots + X_n$ where X_i is the indicator random variable of the event $\sigma(i) = i$. Then

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

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

so that

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

In applications we often use that there is a point in the probability space for which $X \geq \mathrm{E}[X]$ and a point for which $X \leq \mathrm{E}[X]$. We have selected results with a purpose of describing this basic methodology. The following result of Szele (1943) is oftentimes 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; that is, satisfying $(\sigma(i), \sigma(i+1)) \in T$ for $1 \leq i < n$. Then $X = \sum X_\sigma$ and

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

Thus some tournament has at least $\mathrm{E}[X]$ Hamiltonian paths. ■

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

2.2 SPLITTING GRAPHS

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

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

$$\mathrm{E}[X_{xy}] = \frac{1}{2}$$

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

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

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 (which is tight for complete graphs).

Theorem 2.2.2 *If G has $2n$ vertices and e edges then it contains a bipartite subgraph with at least $en/(2n-1)$ edges. If G has $2n+1$ vertices and e edges then it contains a bipartite subgraph with at least $e(n+1)/2n+1$ 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 $n/(2n-1)$ 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 $V = V_1 \cup \cdots \cup V_k$, where the V_i are disjoint sets of size n. Let $h : V^k \to \{\pm 1\}$ 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

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

these choices being mutually independent, with 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 $|E \cap V_i| = a_i$, $1 \le i \le k$. For these E,

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

Combining terms by type

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

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

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

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

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

Thus

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

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

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

$$\mathrm{E}[|X|] \ge |\mathrm{E}[X]| \ge c_k n^k .$$

Some particular value of $|X|$ must exceed or equal its expectation. Hence there is a particular set $S \subseteq V$ with $|X| = |h(S)| \ge c_k n^k$. ■

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

2.3 TWO QUICKIES

Linearity of expectation sometimes gives very quick results.

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

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

monochromatic K_a.

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

In Chapter 16 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 $\mathrm{E}[X]$. 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* $v_1, \ldots, v_n \in \mathbb{R}^n$, *all* $|v_i| = 1$. *Then there exist* $\epsilon_1, \ldots, \epsilon_n = \pm 1$ *so that*

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

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

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

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

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

Then

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

Thus

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

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

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

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 translation of the $p_1 = \cdots = p_n = 1/2$ case.

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

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

Proof. Pick ϵ_i independently with

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

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

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

For $i \neq j$,

$$\mathrm{E}[(p_i - \epsilon_i)(p_j - \epsilon_j)] = \mathrm{E}[p_i - \epsilon_i]\,\mathrm{E}[p_j - \epsilon_j] = 0.$$

For $i = j$,

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

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

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

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

2.5 UNBALANCING LIGHTS

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

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

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 $(\sqrt{2/\pi} + o(1))n^{3/2}$.

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

$$R_i = \sum_{j=1}^{n} a_{ij}y_j\,, \quad R = \sum_{i=1}^{n} |R_i|\,.$$

Fix i. Regardless of a_{ij}, $a_{ij}y_j$ is ± 1 with probability $1/2$ and their values (over j) are independent; that is, whatever the ith 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

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

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

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

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

Now apply linearity of expectation to R:

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

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

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

Another result on unbalancing lights appears in The Probabilistic Lens: Unbalancing Lights (following Chapter 13). The existence of Hadamard matrices and the discussion in Section 9.1 show that the estimate in the last theorem cannot be improved to anything bigger than $n^{3/2}$.

2.6 WITHOUT COIN FLIPS

A non probabilistic 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 $w = \epsilon_1 v_1 + \cdots + \epsilon_{i-1} v_{i-1}$ 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 obtuse (or right) angle with w. If the sum need be big make the angle acute 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 $v_1, \ldots, v_n \in \mathbb{R}^n$, $p_1, \ldots, p_n \in [0,1]$ 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

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

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

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

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

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

While the proofs appear similar, a direct implementation of the proof of Theorem 2.4.2 to find $\epsilon_1, \ldots, \epsilon_n$ might take an exhaustive search with exponential time. In applying the greedy algorithm at the sth 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 16 we discuss several similar examples in a more general setting.

2.7 EXERCISES

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

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

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

4. Suppose $p > n > 10m^2$, with p prime, and let $0 < a_1 < a_2, < \cdots < a_m < p$ be integers. Prove that there is an integer x, $0 < x < p$ for which the m numbers
$$(xa_i \bmod p) \bmod n, \qquad 1 \leq i \leq m$$
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 shown in The Probabilistic Lens: 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 Hamiltonian cycles in K (i.e., 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 $\mathbb{R}^n$ 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 $\epsilon^2 = k/2^r < 1$.

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 ith 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 $\mathrm{per}(A)$ is simply $|S|$. The following result was conjectured by Minc and proved by Brégman (1973). The proof presented here is similar to that of Schrijver (1978).

Theorem 1 [Brégman's Theorem] $\mathrm{per}(A) \le \prod_{1 \le i \le n} (r_i!)^{1/r_i}$.

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 $\sigma\tau 1, \ldots, \sigma\tau(i-1)$ 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 different 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, $G[Y] = e^{\mathrm{E}[\ln Y]}$. Linearity of expectation translates into the geometric mean of a product being the product of the geometric means.

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

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

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

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

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

and so

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

Lemma 2 $\left(\prod_{j=1}^{r} t_j^{t_j} \right)^{1/r} \geq t^t$.

Proof. Taking logarithms, this is equivalent to

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

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

Applying the lemma,

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

■

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

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

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

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

3

Alterations

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

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

3.1 RAMSEY NUMBERS

Recall from Section 1.1 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, $R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$.

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

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. By linearity of expectation, $\mathrm{E}[X] = \sum \mathrm{E}[X_R] = m$ 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

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

For Theorem 3.1.2, $\mathrm{E}[X] < 1$ 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), $s \leq \mathrm{E}[X]$. 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 and 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,

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

so by linearity of expectation,

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

Clearly $\mathrm{E}[X] = np$, so, again by linearity of expectation,

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

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

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

Thus there exists a specific S for which 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 occurring is less than $0.6n^{-2}$ and so

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

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

3.4 PACKING

Let C be a bounded measurable subset of $\mathbb{R}^d$ 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 proved 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,

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

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 $n - (n^2/2)2^d x^{-d}\mu(C)$ 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 14).

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 two possibilities are decreasing and increasing functions of p. Calculus then points us to the optimal p.

We use the notation of Section 1.3 on property B: $m(n) > m$ means that given any n-uniform hypergraph $H = (V, E)$ with m edges there exists a two-coloring of V so that no edge is monochromatic. Beck (1978) improved Erdős' 1963 bound to $m(n) = \Omega(2^n n^{1/3})$. Building on his methods, Radhakrishnan and Srinivasan (2000) proved $m(n) = \Omega\left(2^n (n/\ln n)^{1/2}\right)$ 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 $m(n) = \Omega\left(2^n (n/\ln n)^{1/2}\right)$.

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

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

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

The condition of Theorem 3.5.1 is somewhat typical; one wants the total failure probability to be less than 1 and there are two types of failure. Oftentimes one finds reasonable bounds by requiring the stronger condition that each failure type has probability less than one-half. Here $k^2 p \leq \frac{1}{2}$ 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 $k(1-p)^n + k^2 p$. The assumption of Theorem 3.5.1 then assures us that with positive probability the algorithm succeeds. This, by our usual magic, means that there is some running of the algorithm which yields a final coloring with no monochromatic e; that is, there exists a two-coloring of V with no monochromatic edge. For convenience, we bound the probability that some $e \in H$ is red in the final coloring; the failure probability for the algorithm is at most twice that.

An $e \in E$ can be red in the final coloring in two ways. Either e was red in the first coloring and remained red through to the final coloring or e was not red in the first coloring but was red in the final coloring (the structure of the algorithm assures us

that vertices cannot change color more than once). Let A_e be the first event and C_e the second. Then

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

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

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

giving the first addend of our failure probability.

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

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

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

Suppose C_e holds. Some points of e changed color from blue to red so there is a *last* point v that did so. But why did v flip its coin? It must have been still dangerous. That is, v must be in some (perhaps many) set f that was blue in the first coloring and was still blue when v was considered. Can e, f overlap in another vertex v' ? No! For such a v' would necessarily have been blue in the first coloring (as $v' \in f$) and red in the final coloring (as $v' \in e$), but then v' changed color before v. Hence f was no longer entirely blue when v was considered, contradicting the assumption on f. Therefore, when C_e holds, e blames some f. Let B_{ef} be the event that e blames f. Then $\sum_e \Pr[C_e] \leq \sum_{e \neq f} \Pr[B_{ef}]$. As there are less than $(2^{n-1}k)^2$ pairs $e \neq f$ it now suffices to bound $\Pr[B_{ef}] \leq 2^{1-2n}p$.

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

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

Let's take the factors one at a time. First, v itself must start blue and turn red. Second, all other $v' \in f$ must start blue. Third, all $v' \in f$ coming before v must have second coin tails. Fourth, all $v' \in e$ coming after v must start red (since v is the last point of e to change color). Finally, all $v' \in e$ coming before v must either start red or start blue and turn red. [The final factor may well be a substantial overestimate. Those $v' \in e$ coming before v which start blue must not only have second coin heads but must themselves lie in an $e' \in H$ monochromatic under the first coloring. Attempts to further improve bounds on $m(n)$ have often centered on this overestimate but (thus far!) to no avail.]

We can then write

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

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

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

Proof. Fix a matching between $e - \{v\}$ and $f - \{v\}$; think of Mr. & Mrs. Jones, Mr. & Mrs. Smith, and so on. Condition on how many of each pair (two Joneses, one Smith, no Taylors, etc.) come before v. This splits the space into 3^{n-1} parts, and it suffices to show that the conditional expectation in each of them is at most 1. Indeed, the factor contributed to $(1+p)^i (1-p)^j$ from each pair is at most 1, as follows: 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 $1+p$ (Brad) or $1-p$ (Angelina), giving a factor of one. Moreover, these factors are independent for different pairs (given the above conditioning). All factors are at most one, and hence so is their product. ■

The desired result follows. ■

3.6 CONTINUOUS TIME

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

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

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

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

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

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

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

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

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

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

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

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

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 infinitesimal time interval $[x, x + dx)$, Eve has probability dx of giving birth and these events are independent over disjoint intervals.] Our fertile Eve always gives birth to k-tuplets. Each child is born fertile under the same rules, so if Alice in born at time x she (in our unisexual model) has a Poisson distribution with mean x of births, uniformly distributed in $[0, x)$.

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

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

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

Example. $c = 10, k = 2$. Eve gives birth to Alice, Barbara at time 8.3 and then to Rachel, Siena at time 4.3. Alice gives birth to Nancy, Olive at time 5.7 and Rachel gives birth to Linda, Mayavati at time 0.4. There are no other births. Leaves Nancy, Olive, Linda, Mayavati, Barbara and Siena 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 birth tree $T = T_c$.

We outline the equivalence by defining a tree $T = T_c(v)$ for $v \in H$. For each edge e containing v with birth time $t = t_e < c$ 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 birth time $t' = t_{e'} < t$ 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 birth times 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 birth time $t_e < c$ has binomial distribution $\mathrm{BIN}[D, c/D]$ which approaches

(critically) the Poisson distribution with mean c. Given that there are l such e their birth times 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 birth time 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 about D edges e containing w and no other such w'. We now examine their birth times, the number with $t_e < x$ has binomial distribution $\mathrm{BIN}[D - o(D), x/D]$ 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 occurring in $[0, c)$. If Eve survives in T_d she must survive in T_c. Hence $f(d) \le f(c)$. We now claim

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

If not, the nondecreasing f would have a limit $L > 0$ and all $f(x) \ge L$. 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,

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

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

By linearity of expectation $\mathrm{E}[|S_c|] \to f(c)n$. As $(k+1)|P_c| + |S_c| = n$, $\mathrm{E}[|P_c|] \to (1 - f(c))n/(k+1)$. But $\mathrm{E}\left[|P^{\mathrm{FINAL}}|\right] \ge \mathrm{E}[|P_c|]$. We make $f(c)$ arbitrarily small by taking c appropriately big, so that $\mathrm{E}\left[|P^{\mathrm{FINAL}}|\right] \ge (1 - o(1))n/(k+1)$. As $|P^{\mathrm{FINAL}}| \le n/(k+1)$ 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 $\mathrm{E}[|S_D|] = O(ND^{-1/k})$, 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 that gives a packing with expected size $\sim N/(k+1)$ and our usual magic implies such a P must exist. ■

In particular, this gives an alternate proof to the Erdős–Hanani conjecture, first proved by Rödl as given in Section 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 Section 4.7) that k, l are fixed and $n \to \infty$ so this number of common edges is $o(D)$. The assumptions of Section 4.7 give $K+1$ fixed, $N, D \to \infty$ so that there exists P with

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

3.7 EXERCISES

1. As shown in Section 3.1, the Ramsey number $R(k,k)$ satisfies

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

for every integer n. Conclude that

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

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

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

3. Prove that every three-uniform hypergraph with n vertices and $m \geq n/3$ edges contains an independent set (i.e., a set of vertices containing no edges) 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 $p = n^{\theta-1}$; that is, 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

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

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

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

Set $x = \lceil (3/p) \ln n \rceil$ so that

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

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

$\alpha(G) < 3n^{1-\theta} \ln n$. Remove from G a vertex from each cycle of length at most l. This gives a graph G^* with at least $n/2$ vertices. G^* has girth greater than l and $\alpha(G^*) \leq \alpha(G)$. Thus

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

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