17. SEVENTEENTH LECTURE : 1/12

A slightly more natural notion than that of 'basis' with respect to the positive integers is the following modification :

Definition. Let $A \subseteq \mathbb{N}_0$ and $h \in \mathbb{N}$. We say that A is an *asymptotic basis* for $\mathbb{N}_0$ of order h if the difference $\mathbb{N}_0 \backslash hA$ is a finite set, whereas $\mathbb{N}_0 \backslash (h-1)A$ is not.

In words, an asymptotic basis A of order h has the property that every sufficiently large number can be expressed as a sum of h elements of A, and h is the smallest integer for which this is the case.

Remark 17.1. You might object that it is even more natural to say that an asymptotic basis of order h should have the property that every sufficiently large number is expressible as a sum of AT MOST h elements of A. I agree, but note that, if A is an asymptotic basis of order h according to this alternative criterion, then $A \cup \{0\}$ is an asymptotic basis of order h according to the definition given above.

Similarly, if A is an asymptotic basis of order h, then clearly there is a superset B of A which is a basis of order at most h and such that $B \backslash A$ is a finite set. The kinds of questions we will be dealing with below concern the density of bases. Since any basis of any order is an infinite set, adding a finite number of elements does not affect anything in this sense.

Examples. (i) The solution to Waring's problem states that the k:th powers are a basis of order $g(k)$ and an asymptotic basis of order $G(k)$. Note that $G(k) < g(k)$ for every $k > 2$.
(ii) Vinogradov's theorem states that the set of primes $\mathbb{P}$ forms an asymptotic basis of order at most 4. It is not known whether they are an asymptotic basis of order 3, something which would follow from Goldbach's conjecture. It is not even known whether $\{0,1\} \cup \mathbb{P}$ is a basis of order at most 4 (since the number coming out of Vinogradov's theorem is computationally unreachable), let alone of order 3 as Goldbach asserts.

Probably the most important question of a combinatorial nature about bases 'in general' is the issue of how efficiently they can be constructed. In linear algebra (where the notion of basis is somewhat different), a basis is an optimally efficient spanning set for a vector space in two respects :

(i) it is a spanning set of minimal size,

(ii) every vector has exactly one representation as a linear combination of the basis vectors, so there is no redundancy.

Similarly, there are two basic ways of measuring the efficiency of a basis in number theory :

METHOD 1 : For $A \subseteq \mathbb{N}$ and $n \in \mathbb{N}$, let

$$A(n) := \#(A \cap \{1, ..., n\}). \tag{17.1}$$

We would like to know, if A is an asymptotic basis of order h, then how slowly can the function $A(n)$ grow, in other words, how sparse a set can A be ?

METHOD 2 : For $A \subseteq \mathbb{N}$ and $h, n \in \mathbb{N}$ define $r_h(A, n)$ to be the number of ways of writing n as a sum of h elements of A. This is called the *(unordered) h-fold representation function* of A. Note that here I do not distinguish between two representations of a number n which consist of the same parts, only permuted. Thus, if A is an asymptotic basis of order h it implies that $r_h(A, n) > 0$ for all $n >> 0$. We would like to know, if A is an asymptotic basis of order h, then how small can the numbers $r_h(A, n)$ be ? Ideally, can we have $r_h(A, n) = 1$ for all $n >> 0$, i.e.: no redundancy whatsoever ?

In contrast to the situation in linear algebra, we will see that the answer to our question in the second case above is 'No'. This seems to be a highly non-trivial fact, though. Indeed, the deeper results in the literature at present seem to concern the second notion of efficiency for bases presented above. The precise connection between the two notions is not as simple to work out as in the linear algebra setting. To begin with, I will present some basic results concerning the first notion.

Proposition 17.2. *Let A be an asymptotic basis of order h. Then*

$$A(n) \gtrsim (h!)^{1/h} n^{1/h}. \tag{17.2}$$

The proof of this will require a basic combinatorial formula which you may have seen before :

Lemma 17.3. *Let k, l be positive integers. The number of ordered solutions in non-negative integers to the equation*

$$x_1 + \cdots + x_k = l \tag{17.3}$$

is $\begin{pmatrix} k+l-1 \\ l \end{pmatrix}$. This is also the number of ways of choosing l elements from a k-element set, where repitition is allowed and ordering of choice is unimportant.

Proof. of Lemma. Imagine you have l identical dots and $k-1$ identical vertical dashes. Then you have $k+l-1$ symbols in all. There are $\begin{pmatrix} k+l-1 \\ l \end{pmatrix}$ ways of arranging these symbols in a line, since the only choice that makes a difference is where you place the dots. But there is a 1-1 correspondence between the arrangements of dots and dashes and the non-negative integer solutions to (17.3), namely : given an arrangement of dots and dashes, interpret x_i as the number of dots between the $(i-1)$:st and i:th dashes.

Finally, to see that the solutions to (17.3) are in 1-1 correspondence with the unordered choices of l elements from a k-set, say the set $\{1, ..., k\}$, just observe that, given such a choice, we can interpret x_i as the number of times you choose the number i. □

Proof. of Proposition 17.2. Since A is an aysmptotic basis there are $O(1)$ elements of $\mathbb{N}$ which cannot be expressed as a sum of h elements of A. Now let n be a large integer. All but $O(1)$ of the numbers among $\{1, ..., n\}$ have an expression as

$$a_1 + \cdots + a_h, \quad a_i \in A. \tag{17.4}$$

Obviously, in any such expression, each $a_i \in \{1, ..., n\}$. Let $T := A(n)$. By Lemma 17.3, the total number of expressions (17.4), satisfying that each $a_i \leq n$, is $\begin{pmatrix} T+h-1 \\ h \end{pmatrix}$. Thus

$$\begin{pmatrix} T+h-1 \\ h \end{pmatrix} \geq n - O(1). \tag{17.5}$$

Let $n \to \infty$. Obviously, $n - O(1) \sim n$. Also,

$$\begin{pmatrix} T+h-1 \\ h \end{pmatrix} = \frac{(T+h-1)(T+h-2)\cdots(T)}{h!} \sim \frac{T^h}{h!}. \tag{17.6}$$

Hence $T^h/h! \gtrsim n$ and hence $T = A(n) \gtrsim (h!)^{1/h} n^{1/h}$, v.s.v. □

Remark 17.4. By Stirling's formula, as $h \to \infty$,

$$(h!)^{1/h} \sim (h^h e^{-h} \sqrt{2\pi h})^{1/h} \sim \frac{h}{e}. \tag{17.7}$$

Let $A = (a_n)$ denote a generic infinite subset of $\mathbb{N}_0$, where the elements of A are listed in increasing order. If A is an asymptotic basis of order h, then the lower bound (17.2) for $A(n)$ can be easily converted to an upper bound for a_n. For, by definition, $A(a_n) = n$ and hence $n \gtrsim (h!)^{1/h} a_n^{1/h}$. In other words,

$$a_n \lesssim \frac{n^h}{h!}. \tag{17.8}$$

This is perhaps an easier way to think about things, as it gives an upper bound on how 'sparse' a basis of order h can be. Bases whose sparsity is of the right order of magnitude have been constructed :

Theorem 17.5. *For every $h \in \mathbb{N}$ there exists a positive number γ_h and an asymptotic basis $A = (a_n)$ of order h such that $a_n \sim \gamma_h n^h$.*

There are several explicit constructions of bases in the literature which imply this theorem, the most important one probably being due to Cassels[1]. His construction is complicated, however. Below, we give an example of a fairly simple construction which at least has the right order of magnitude. Note, though, that it is an unsolved problem to determine the largest possible constants γ_h such that an asymptotic basis of order h exists satisfying $a_n \sim \gamma_h n^h$. We know from (17.8) that $\gamma_h \leq 1/h!$. However, the true maximal value of γ_h is not known for a single $h > 1$ (as far as I know !), and the problem seems to be pretty intractable.

Example. Let $h \geq 2$. For each $n \in \mathbb{N}$, write n in base 2, say

$$n = x_k x_{k-1} \cdots x_1 x_0, \quad x_i \in \{0, 1\},\ x_k = 1. \tag{17.9}$$

Let $A = \cup_{i=0}^{h-1} A_i$ where, for each i, the set A_i is defined by

$$A_i = \{n \in \mathbb{N} : x_r = 0 \text{ for all } r \not\equiv i \pmod h\}. \tag{17.10}$$

[1] J.W.S. Cassels, Über Basen der natürliche Zahlenreihe, *Abh. Math. Sem. Univ. Hamburg* **21** (1957), 247-257.

Then it is easy to see that $A \cup \{0\}$ is a basis (not just an asymptotic basis) of order h and, for any $t \in \mathbb{N}$,

$$A(2^{ht} - 1) = h \cdot 2^t - (h - 1). \tag{17.11}$$

From this one can check that

$$a_n \sim \frac{n^h}{h^h}. \tag{17.12}$$

By (17.7) and (17.11), this basis is in a sense, for large h, at most a factor $1/e$ from 'optimal'.

We now turn to our second notion of what it means to efficiently construct a basis, a notion which emphasises lack of redundancy rather than size. There is the following fundamental result :

Theorem 17.6. *Let A be an asymptotic basis of order h. Then the representation function $r_h(A, n)$ cannot be ultimately constant. In particular, we cannot have $r_h(A, n) = 1$ for all sufficiently large n.*

This theorem was proven for arbitrary h by Vaughan[2] using Fourier analysis methods. No really 'elementary' proof is known, even for $h = 2$. In that special case, however, a clever proof was provided earlier by Erdős and Turán using so-called *generating functions*.

I started this proof, but didn't finish it, so I will present it in full next day ...

[2]R.C. Vaughan, On the addition of sequences of integers, *J. Number Theory* **4** (1972), 1-16.

18. EIGHTEENTH LECTURE : 3/12

Proof. of Theorem 17.6. Suppose the contrary and let A be an asymptotic basis of order 2 such that $r_2(A,n) = l$ for all sufficiently large n and some constant $l > 0$. We consider the generating function of the set A, which is the power series

$$G(z) = G_A(z) := \sum_{a \in A} z^a \quad (z \in \mathbb{C}). \tag{18.1}$$

The power series certainly converges when $|z| < 1$, so we will work in this region so that all our algebraic manipulations will be valid. The connection between the generating function and the representation function is that

$$[G(z)]^2 + G(z^2) = 2 \cdot \sum_{n=1}^{\infty} r_2(A,n) z^n. \tag{18.2}$$

Suppose that $r_2(A,n) = l$ for all $n \geq n_0$. Then (18.2) can be written as

$$[G(z)]^2 + G(z^2) = \sum_{n=1}^{n_0-1} r_2(A,n) z^n + 2l \cdot \sum_{n=n_0}^{\infty} z^n. \tag{18.3}$$

The first sum on the right of (18.3) is some polynomial in z. The second sum is a geometric series, so has a simple formula. We thus find that

$$[G(z)]^2 + G(z^2) = P(z) + 2l \cdot \frac{z^{n_0}}{1-z}. \tag{18.4}$$

We obtain a contradiction by seeing what happens as $z \to -1^+$, along the real axis. Because of all the squares present, the left-hand side of (18.4) heads inexorably towards positive infinity. But the right-hand side heads towards some finite value, namely $P(1) \pm l$. This contradiction completes the proof. □

Notation. Previously, we have introduced the notation $f = O(g)$ to denote that $|f(x)/g(x)|$ is bounded as $x \to \infty$. Two similar pieces of notation whcih are useful to have are :
1. $f = \Omega(g)$ means the same thing as $g = O(f)$, in other words, the quotient$|f(x)/g(x)|$ has a positive lower bound as $x \to \infty$. In words, it means that f grows at least as fast as g.
2. $f = \Theta(g)$ means that both $f = O(g)$ and $g = O(f)$. In other words, it means that there exist positive constants $0 < c_1 \leq c_2$ such that

$$c_1 \leq \liminf_{x \to \infty} \left| \frac{f(x)}{g(x)} \right| \leq \limsup_{x \to \infty} \left| \frac{f(x)}{g(x)} \right| \leq c_2. \tag{18.5}$$

In words, f and g have the same rate of growth.

One of the major open problems in combinatorial additive number theory is

Conjecture 18.1. (Erdős-Turán) *Let $h \geq 2$ and $A \subseteq \mathbb{N}$ be an asymptotic basis of order h. Then the representation function $r_h(A,n)$ is unbounded.*

Bases whose representation functions are 'slowly growing' are called *thin*. The state-of-the-art with regard to the construction of thin bases is

Theorem 18.2. *For each $h \geq 2$ there exists an asymptotic basis A of order h satisfying*

$$r_h(A, n) = \Theta(\log n). \tag{18.6}$$

This theorem is one of the classical applications of the so-called *probabilistic method* in number theory. The case $h = 2$ was proven by Erdős in 1956 using *Chernoff's inequality*. The general case wasn't completed until 1990, by Erdős and Tetali. It is most succinctly presented using the so-called *Janson inequalities*, though Erdős and Tetali did not actually use these in their original proof.

The more important point, though, is that these proofs use probabilistic arguments, and in each case use tools from probability theory which were state-of-the-art at the time. My intention in coming lectures is to give a proof of the case $h = 2$. On the way I will state, but not prove, Chernoff's inequality. As a lead-up to the main result, I will give an introduction to the probabilistic method in general (a method which is widely used in all areas of combinatorics) by means of the historically significant example of *Ramsey numbers.*

This material will occupy us in the next two lectures. Before embarking upon it, I wish to make some extra comments and state some open problems arising out of Theorem 17.6.

Definition. A set $A \subseteq \mathbb{Z}$ is called a *Sidon set* if all the sums $a_1 + a_2$ are distinct, for $a_1, a_2 \in A$.

Example. $A = \{0, 1, 3, 7\}$ is a Sidon set. There are 10 possible sums, and all are distinct, namely

$$0+0=0,\ \ 0+1=1,\ \ 1+1=2,\ \ 0+3=3,\ \ 1+3=4,$$
$$3+3=6,\ \ 0+7=7,\ \ 1+7=8,\ \ 3+7=10,\ \ 7+7=14.$$

Theorem 17.6 implies that an asymptotic basis of order 2 for $\mathbb{N}$ cannot be a Sidon set. There are many open problems regarding Sidon sets. Here are some :

Problem 18.3. *Can an asymptotic basis of order 3 be a Sidon set ?*

Problem 18.4. *Can an asymptotic basis of order 4 be a Sidon set ?*

It can be deduced from the Erdős-Tetali result on thin bases that there exist asymptotic bases of order 5 which are Sidon sets. However, because of the probabilistic nature of the argument, it yields no explicit examples.

Problem 18.5. *Give an explicit example of a Sidon set which is an asymptotic basis of order 5.*

Some interesting problems revolve around the connection between finite and infinite (dense) Sidon sets. For $n > 0$, let $S(n)$ denote the maximum size of a Sidon set in $\{1, ..., n\}$. Since the number of possible sums of pairs of elements in a k-element set is $k + \binom{k}{2} = \frac{k(k+1)}{2}$. Since a sum of any two elements amongst $1, ..., n$ must lie

amongst $2, ..., 2n$, we get that

$$\frac{S(n)[S(n)+1]}{2} \leq 2n-1 \Rightarrow S(n) \lesssim 2\sqrt{n}. \tag{18.7}$$

The following theorem was proven in the 1940s by combining work of Erdős, Turán and Chowla :

Theorem 18.6. *There exist positive constants c_1, c_2 such that, for all $n \in \mathbb{N}$,*

$$n^{1/2} - c_1 n^{5/16} < S(n) < n^{1/2} + c_2 n^{1/4}. \tag{18.8}$$

There remains to this day a gap between the known lower and upper bounds for $S(n)$, but since this gap is already fairly small, by (18.8), this problem is perhaps not so interesting. More intriguingly, the following result of Erdős (1955) implies that dense finite Sidon sets cannot be 'glued' together to give equally dense infinite Sidon sets :

Theorem 18.7. *Let $A \subseteq \mathbb{N}$ be a Sidon set. Then there exists a universal constant $c > 0$ such that*

$$\liminf_{n\to\infty} \frac{A(n)}{\sqrt{n/\log n}} \leq c. \tag{18.9}$$

Indeed the densest known infinite Sidon sets are given by :

Theorem 18.8. (Ruzsa 1998) *There is a Sidon set $A \subseteq \mathbb{N}$ such that*

$$A(n) = n^{\sqrt{2}-1+o_n(1)}. \tag{18.10}$$

Ruzsa's proof uses a probabilistic argument and thus does not yield explicit examples. Indeed, basically nothing is known in terms of explicitly constructing dense infinite Sidon sets, beyond what one gets from a greedy choice procedure. This involves constructing an infinite Sidon set $A \subseteq \mathbb{N}_0$ according to the following rules :

First take $0 \in A$. At each successive step, add to A the smallest number you can which maintains the Sidon property.

The set one gets this way starts off as $A = \{0, 1, 3, 7, 12, ...\}$ and one can quite easily verify that $A(n) = \Theta(n^{1/3})$. Any improvement on this exponent would be a significant achievement.

One final problem. As we've stated several times, there is no subset $A \subseteq \mathbb{N}$ such that every sufficiently large number can be uniquely expressed as $a_1 + a_2$. However, consider the set A consisting of all $n \in \mathbb{N}_0$ in whose binary representation, there are zeroes in all the even positions, reading from right to left (recall the example after Theorem 17.5). Then it easy to see that every $n \in \mathbb{N}_0$ has a unique expression as $a_1 + 2a_2$, for some $a_1, a_2 \in A$. The following problem is open :

Problem 18.9. *For which pairs $\{u, v\}$ of relatively prime positive integers does there exist a set $A \subseteq \mathbb{N}$ such that every sufficiently large number has a unique representation as $ua_1 + va_2$, for some $a_1, a_2 \in A$?*

The above comments imply that $\{1, 2\}$ is such a pair, whereas $\{1, 1\}$ isn't.

19. NINETEENTH LECTURE : 5/12

To introduce the proabilistic method, we will take a detour away from number theory per se and discuss the problem of computing so-called Ramsey numbers. The standard reference for an introduction to the probabilistic method in discrete mathematics is

N. Alon and J. Spencer, *The Probabilistic Method,* 3rd edition, Wiley (2008).

Definition. Let $k, l \in \mathbb{N}_{\geq 2}$. The *Ramsey number* $R(k,l)$ is the smallest $n \in \mathbb{N}$ such that, if the edges of the complete graph K_n are each colored either red or blue then, no matter how the coloring is done, there must exist either a totally red K_k subgraph or a totally blue K_l subgraph.

The fundamental result proven by Frank Ramsey in the 1920s was :

Theorem 19.1. *The numbers $R(k,l)$ all exist, i.e.: are finite. In fact, we have*

$$R(k,l) = R(l,k), \tag{19.1}$$

$$R(k,2) = R(2,k) = k, \tag{19.2}$$

and, in general,

$$R(k,l) \leq R(k-1,l) + R(k,l-1). \tag{19.3}$$

Proof. Note that the three statements together imply that all the numbers $R(k,l)$ are finite, say by an induction on $k+l$. The first statement is totally obvious and, for the second, note that since a K_2 is just a single edge, if K_n is colored in such a way that it has no red K_2 subgraphs, then it means that the entire K_n itself is blue.

So we turn to the third statement. Let $n = R(k-1,l) + R(k,l-1)$. We must show that any coloring of the edges of K_n must yield either a red K_k or a blue K_l subgraph. Consider an arbitrary coloring of K_n. Pick out any vertex, call it v_0. There are $n-1$ other vertices and, since

$$n - 1 > [R(k-1,l) - 1] + [R(k,l-1) - 1], \tag{19.4}$$

at least one of the following events must occur :

CASE 1 : v_0 is connected to at least $R(k-1,l)$ other vertices by a red edge,
CASE 2 : v_0 is connected to at least $R(k,l-1)$ other vertices by a blue edge.

We present the remainder of the argument in Case 1 only, for Case 2 is dealt with similarly. By definition of the number $R(k-1,l)$, amongst the vertices to which v_0 is connected by a red edge, there must exist at least one of the following patterns :

(a) an entirely blue K_l,
(b) an entirely red K_{k-1}.

In the former case, we're already done - we have a blue K_l in the whole graph if we have one in any part of it (wow !). But in the latter case, just tag on the vertex v_0 and we obtain instead a red K_k in the whole graph. □

From this result we can deduce explicit upper bounds for Ramsey numbers :

Corollary 19.2. *For every $k, l \geq 2$ we have*

$$R(k,l) \leq \left(\begin{array}{c} k+l-2 \\ k-1 \end{array} \right). \tag{19.5}$$

Proof. For $l = 2$ this reduces to

$$R(k,2) \leq \left(\begin{array}{c} k \\ k-1 \end{array} \right), \tag{19.6}$$

which is true, since both sides equal k. To establish the inequality in general, we proceed by induction on $k + l$. By (19.3), it is easily seen to suffice that

$$\left(\begin{array}{c} k+l-2 \\ k-1 \end{array} \right) \geq \left(\begin{array}{c} k+l-3 \\ k-2 \end{array} \right) + \left(\begin{array}{c} k+l-3 \\ k-1 \end{array} \right). \tag{19.7}$$

And this is true, in fact we have equality again, by Pascal's identity for binomial coefficients :

$$\left(\begin{array}{c} n \\ r \end{array} \right) = \left(\begin{array}{c} n-1 \\ r-1 \end{array} \right) + \left(\begin{array}{c} n-1 \\ r \end{array} \right). \tag{19.8}$$

□

It is natural to consider the special (symmetric) case $k = l$. Then (19.5) becomes

$$R(k,k) \leq \left(\begin{array}{c} 2k-2 \\ k-1 \end{array} \right). \tag{19.9}$$

Using simply the fact that

$$\sum_{r=0}^{n} \left(\begin{array}{c} n \\ r \end{array} \right) = 2^n \tag{19.10}$$

(both sides count the number of subsets of an n-element set), it follows that

$$R(k,k) \leq 4^{k-1}. \tag{19.11}$$

A slightly better estimate can be got using Stirling's formula, but the important point is that the upper bound we obtain for $R(k,k)$ is exponential in k. A problem which Ramsey didn't solve, and which remained open for a few years after his untimely death in 1930 at the age of 26, was whether these symmetric Ramsey numbers really do grow exponentially. It turns out that they do, and the proof of this fact is perhaps the oldest application of what has become known as the probabilistic method in combinatorics[3]. The formal result is

Theorem 19.3. *Let $k \geq 3$. If the positive integer n satisfies*

$$\left(\begin{array}{c} n \\ k \end{array} \right) 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)} < 1, \tag{19.12}$$

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

Corollary 19.4.

$$R(k,k) \geq 2^{k/2}. \tag{19.13}$$

[3]Only a finite (and very small !) collection of Ramsey numbers $R(k,l)$, for $k, l \geq 3$, have been computed exactly. For example, I showed in class that $R(3,3) = 6$.

Proof. of Corollary. Note that we have equality in (19.13) for $k = 2$. Now for $k \geq 3$ it suffices to show that if $n = 2^{k/2}$ then (19.12) is satisfied. Since $\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n(n-1)\cdots(n-k+1)}{k!} < \frac{n^k}{k!}$ and $\left(\begin{array}{c} k \\ 2 \end{array} \right) = \frac{k(k-1)}{2}$, the left-hand side of (19.12) will be less than

$$\frac{2^{\frac{k^2}{2}+1}}{k!\, 2^{\frac{k(k-1)}{2}}} = \frac{2^{1+\frac{k}{2}}}{k!}. \tag{19.14}$$

So we just need to check that $2^{1+\frac{k}{2}} \leq k!$, for every $k \geq 3$. This is easily done by induction on k (the case $k = 3$ reducing to $4\sqrt{2} < 6$). □

Proof. of Theorem. Let n be an integer satisfying (19.12). We must show that there is a way to color the edges of K_n red and blue so that there will be no monochromatic K_k subgraph. Consider a 'totally random' coloring of the edges, i.e.: for each of the $\left(\begin{array}{c} n \\ 2 \end{array} \right)$ edges independently, toss a fair coin and color the edge red or blue depending on whether you get heads or tails respectively. Thus, each edge is equally likely to be colored red or blue, and in fact all $2^{\left(\begin{array}{c} n \\ 2 \end{array} \right)}$ possible colorings of the entire graph are equally likely. Let A be any K_k subgraph. It has $\left(\begin{array}{c} k \\ 2 \end{array} \right)$ edges, and hence the probability that all of these will be colored red is $2^{-\left(\begin{array}{c} k \\ 2 \end{array} \right)}$. There is the same probability that the entire subgraph is blue, hence

$$\mathbb{P}(A \text{ is monochromatic}) = 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}. \tag{19.15}$$

The total number of K_k subgraphs in K_n is just $\left(\begin{array}{c} n \\ k \end{array} \right)$. The probability of any particular one being colored monochromatically is given by (19.15). Now, the probability of at least one of a collection of events occurring cannot exceed the sum of the probabilitites of the individual events[4]. Thus the probability that some K_k is monochromatic is at most $\left(\begin{array}{c} n \\ k \end{array} \right) 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}$. By assumption, this quantity is strictly less than one. In other words, a random coloring of K_n yields, with strictly positive probability, a coloring for which there is no monochromatic K_k subgraph. Since there are only finitely many possible colorings a priori ($2^{\left(\begin{array}{c} n \\ 2 \end{array} \right)}$ of them), this means that there must be SOME way to color K_n such that no monochromatic K_k arises. □

[4]This is referred to as the *union bound* by probabilists. Formally, if $E_1, ..., E_r$ are any events, then $\mathbb{P}(\cup_{i=1}^r E_i) \leq \sum_{i=1}^r \mathbb{P}(E_i)$.

Remark 19.5. It's important to note here that this proof suffers from the typical weakness of an application of a probabilistic method, namely that it only proves existence of the desired object (a certain good coloring of K_n), not any explicit example. From an algorithmic/practical viewpoint, this is not such a big deal. If n is significantly smaller than $2^{k/2}$, then the proof implies that a random coloring of K_n will yield no monochromatic K_k with very high probability (because the quantity $\left(\begin{array}{c} n \\ k \end{array} \right) 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}$ will be very close to zero). Thus if we just perform the coloring at random, we will get a good coloring unless we're very unlucky. Of course, we won't actually KNOW that our coloring is good.

As a prelude to developing some machinery which we will need in order to prove Theorem 18.2, it is instructive, I think, at this point to reformulate the above proof using somewhat more formal probabilistic terminology. Formally, probability theory revolves around the study of triples (X, Ω, μ), where X is a so-called *random variable*, Ω is a *probability space* and μ is a *probability measure*. I don't want to get into any technical details here, but I will just say that, in general, Ω is just some set, X is a function from Ω to $\mathbb{R}$, and μ is a function from 2^{Ω} to $[0, 1]$, where 2^{Ω} denotes the collection of all subsets of Ω.

In Theorem 19.3, Ω is the set of all possible colorings of the edges of K_n, where each edge is colored red or blue. Hence Ω is a finite set and $|\Omega| = 2^{\left(\begin{array}{c} n \\ 2 \end{array} \right)}$. μ is so-called *uniform measure* on Ω which means, concretely, that our rules for randomly coloring the graph lead to every possible coloring being equally likely. X is the number of monochromatic K_k subgraphs. Clearly, this is indeed a function from Ω to $\mathbb{R}$. In fact, X only takes on non-negative integer values. Since we choose a coloring from Ω randomly (according to μ), the value of X is also 'random' (hence the name 'random variable'). The strategy of the proof above was to show that

$$\mathbb{E}(X) < 1. \tag{19.16}$$

Since X is non-negative integer valued, this inequality allows one to conclude that

$$\mathbb{P}(X = 0) > 0, \tag{19.17}$$

which proves the theorem. The idea behind proving (19.16) was to write X as a sum of so-called *indicator variables*. Let A be a K_k subgraph in K_n and let $\mathscr{E}_A$ denote the event that a randomly chosen coloring of K_n yields a monochromatic A. We showed that

$$\mathbb{P}(\mathscr{E}_A) = 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}. \tag{19.18}$$

Let X_A denote the following random variable :

$$X_A = \begin{cases} 1, & \text{if } \mathscr{E}_A \text{ occurs,} \\ 0, & \text{otherwise.} \end{cases} \tag{19.19}$$

X_A is called the *indicator variable* of the event $\mathscr{E}_A$. Clearly,

$$\mathbb{E}(X_A) = \mathbb{P}(\mathscr{E}_A). \tag{19.20}$$

The important point is that

$$X = \sum_A X_A, \tag{19.21}$$

the sum being taken over all K_k subgraphs of K_n. Now since the average of a sum equals the sum of the averages, it follows that

$$\mathbb{E}(X) = \sum_A \mathbb{E}(X_A). \tag{19.22}$$

This property of expectation values (that they commute with sums) is referred to by probabilists as *linearity of expectation*. In our example, there are $\left(\begin{array}{c} n \\ k \end{array} \right)$ possibilities for A, hence from (19.18) we could conclude that

$$\mathbb{E}(X) = \left(\begin{array}{c} n \\ k \end{array} \right) 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)} < 1. \tag{19.23}$$

From a probabilistic viewpoint, the proof of Theorem 19.3 is 'easy' in the sense that it sufficed to have knowledge of the expectation value of the random variable X of interest (eq. (19.16)). This we could easily obtain by writing X as a sum of indicator variables and using linearity of expectation. What we didn't need was any information on how *strongly concentrated* X was about its mean. This would actually have been difficult to obtain, since the indicator variables X_A are not all independent. In situations where we have sums of INDEPENDENT, or mostly independent, random variables, it is often possible to prove that the sum X is very strongly concentrated about its mean. Some of the classical results of statistics and probability theory deal with this phenomenon, and it is such a result which will be needed in due course to prove Theorem 18.2. We will delve deeper into these matters next day ...