Lecture 1 (Nov. 1, 2011)

Various notations are used when comparing the rates of growth of different functions, and it is a good idea for us to get these out of the way before we start.

NOTATION : Let $f, g : \mathbb{N} \to \mathbb{R}$ be any functions. We write

(i) $f = O(g)$ if the quotient $|f(n)/g(n)|$ is bounded as $n \to \infty$.
(ii) $f = \Omega(g)$ if $g = O(f)$.
(iii) $f = o(g)$ if $f(n)/g(n) \to 0$ as $n \to \infty$.
(iv) $f \sim g$ if $f(n)/g(n) \to 1$ as $n \to \infty$.
(v) $f \preceq g$ if $\limsup |f(n)/g(n)| \leq 1$.
(vi) $f \succeq g$ if $g \preceq f$.

First, some general words of wisdom (or waffle) :

The basic application of probabilisitc techniques to combinatorics is to prove existence of a structure from amongst a certain class $\mathcal{X}$ of structures, which possesses some desired property $\mathcal{P}$.

One does so by introducing some appropriate probability measure μ on the collection $\mathcal{X}$ and showing (somehow) that, if one chooses at random, according to the distribution μ, an element of $\mathcal{X}$, then with positive probability one's choice possesses the property $\mathcal{P}$.

An important remark :

Usually, though not always, μ is just a simple uniform distribution. Also, since we're interested in combinatorial applications, $\mathcal{X}$ is usually (though not always) a finite collection[1] This means that

(i) there is usually no great mystery about how the probability theory is introduced to the problem under consideration. It is intuitively clear what is meant by 'choosing at random' and one doesn't need to be an expert in probability theory to understand what's going on.

(ii) also, since the sets under consideration are usually finite, one can in principle present most of the same arguments without ever mentioning probability theory at all, i.e.: by 'purely combinatorial' reasoning. Though this is the case, for more sophisticated applications, the advantages of using notions of probability in terms of the clarity of exposition outweigh the disadvantages of having to learn these notions.

[1]We will see some applications, for example in number theory, where $\mathcal{X}$ is infinite. But even here, the underlying set of interest, namely the natural numbers, is discrete.

Note that the probabilistic method is usually employed to show that some desired structure exists. It doesn't usually tell you how to actually find such a structure. This is an *algorithmic* problem, but obviously for real-world applications, one can conceive that it might be essential to actually be able to find what one is looking for. Sometimes the probabilistic method gives a good *randomized algorithm*, basically an algorithm that is fast but has a certain probability of failure[2].

Intuitively, it is clear how this would work. One shows that a structure with property $\mathcal{P}$ exists by showing that if one chooses at random, then one finds something with property $\mathcal{P}$ with probability $\epsilon > 0$. Often it turns out that the proof yields a value of ϵ which is close to 1. This means that a random choice is very likely to be a good one.

The course is roughly divided into three parts :

I. Introduction to the basics of the probabilistic method by means of a variety of examples.
II. Some more sophisticated proabilistic techniques, in particular so-called *concentration inequalities*.
III. To be decided (depends on time considerations etc.).

We will discuss applications of the method to a variety of combinatorial problems, for example in graph theory and number theory, as well as applications in computer science.

Example 1 : Ramsey Numbers

DEFINITION 1 : The *complete graph* on n vertices, denoted K_n, is the graph in which each pair of vertices is joined by a single edge. Thus K_n contains $\left(\begin{array}{c} n \\ 2 \end{array} \right)$ edges.

I will now state and prove an abridged form of what has become known as *Ramsey's theorem*. It is abridged in the sense that, in its' full generality, the number of colors in the statement below can be any finite number, not just two.

Theorem 1 *Let $k, l \geq 2$ be fixed positive integers. Then for all sufficiently large positive integers n (how large depends on k, l), the following holds :*

If each edge of K_n is colored either red or blue, then there must exist either a red K_k or a blue K_l.

[2]There is also a whole theory of *derandomization*, which deals with how to turn fast randomizsed algorithms into decent deterministic ones. We will not discuss this topic in our course. There is, however, a chapter devoted to it in the book of Alon and Spencer

Before proving this, we introduce some notation :

NOTATION : We denote by $R(k,l)$ the smallest integer n for which the above statement holds. It is called the (k,l)-*th Ramsey number*. Theorem 1 states that these numbers exist, for every $k,l \geq 2$.

PROOF OF THEOREM : We present the standard argument, which is basically an induction on $k+l$.

Step 0 : Note that $R(k,l) = R(l,k)$ by symmetry.

Step 1 : Observe that $R(2,l) = l$ since a K_2 is just a graph with a single edge, so if we're to avoid a red K_2 then we must color every edge of our graph blue. And then we'll have a blue K_l as soon as we have l or more vertices.

Step 2 : The general induction step involves verifying the following inequality :

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

So we assume the two Ramsey numbers on the right hand side of (1) exist and consider a 2-coloring of the graph K_n, where $n = R(k-1,l)+R(k,l-1)$. We must prove the existence of either a red K_k or a blue K_l. Pick any one of the n vertices and give it a name, say v. Now v is joined by an edge to $n-1$ other vertices. Since

$$n-1 > [R(k-1,l)-1] + [R(k,l-1)-1],$$

one of the following must occur :

(i) v is joined to at least $R(k-1,l)$ vertices by a red edge, or
(ii) v is joined to at least $R(k,l-1)$ vertices by a blue edge.

Suppose (i) occurs. By definition of the Ramsey numbers, amongst the vertices joined to v by a red edge, there must exist either a red K_{k-1} or a blue K_l. In the latter case we're done already. In the former case, adding on the vertex v gives a red K_k, and again we're done.

If instead (ii) holds, then the argument is similar. It is left to the reader to write out the details.

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

$$R(k,l) \leq \begin{pmatrix} k+l-2 \\ k-1 \end{pmatrix}. \tag{2}$$

PROOF : This follows from (1) and the well-known Pascal identity for binomial coefficients

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

The details are left as an exercise.

It is natural to consider the special case $k = l$. Then (2) becomes

$$R(k,k) \leq \left(\begin{array}{c} 2k-2 \\ k-1 \end{array} \right). \tag{3}$$

Using simply the fact that

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

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

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

Using Stirling's formula[3] (details left as an exercise), we can obtain a slightly better estimate, namely

$$R(k,k) \preceq \frac{4^{k-1}}{\sqrt{\pi(k-1)}}. \tag{5}$$

But the important point is that (4) and (5) both say that the Ramsey numbers $R(k,k)$ grow at worst exponentially.

Now, finally, we intorduce probabilistic ideas to the discussion, in order to show that the numbers $R(k,k)$ do, in fact, exhibit exponential growth. We do this by proving

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

$$\left(\begin{array}{c} n \\ k \end{array} \right) 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)} < 1, \tag{6}$$

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

For the moment, let us assume the theorem and prove what we're really after, namely

Corollary 4

$$R(k,k) > 2^{k/2}. \tag{7}$$

---

[3] $$n! \sim n^n e^{-n} \sqrt{2\pi n}.$$

PROOF OF COROLLARY : We must show that if $k \geq 3$ and $n = 2^{k/2}$, then (6) is satisfied. Since $\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n(n-1)\cdots(n-k+1)}{k!}$ and, in particular, $\left(\begin{array}{c} k \\ 2 \end{array} \right) = \frac{k(k-1)}{2}$, the left-hand side of (6) is thus at most

$$n^k \cdot \frac{2^{1+\frac{k}{2}}}{k! 2^{\frac{k^2}{2}}}.$$

Taking $n = 2^{k/2}$ this becomes simply $2^{1+k/2}/k!$. It is then a simple exericse to verify that $2^{1+k/2}/k! < 1$ for all $k \geq 3$.

We remark that a more careful analysis, again based on Stirling's formula and left as an exercise for the reader, shows that

$$R(k,k) \succeq \frac{1}{e\sqrt{2}} k 2^{k/2}. \tag{8}$$

But again the main point is that both (7) and (8) say that the numbers $R(k,k)$ exhibit exponential growth. Combining all our results, the essence of what we have found is expressed in the following :

$$\sqrt{2} \leq \liminf_{k\to\infty} R(k,k)^{1/k} \leq \limsup_{k\to\infty} R(k,k)^{1/k} \leq 4. \tag{9}$$

BIG Open Problem *Does*

$$\lim_{k\to\infty} R(k,k)^{1/k}$$

exist and, if so, what is it ?

This problem has been open for 70 years without any progress whatsoever having been made beyond (9). An even more daunting task, however, is to compute Ramsey numbers $R(k,l)$ exactly. In fact, for $k, l > 2$ only a small (finite) collection of Ramsey numbers have been computed exactly.

We conclude this discussion by proving Theorem 3 :

PROOF OF THEOREM 3 : The proof will use the following simple facts about probabilities :

(I) For any two events A and B,

$$P(A \cup B) \leq P(A) + P(B), \tag{10}$$

with equality if the events are *mutually exclusive*, i.e.: if $P(A \cap B) = 0$.
(II) If A and B are *independent* events, then

$$P(A \cap B) = P(A) \cdot P(B). \tag{11}$$

Consider now a fixed n and k, and a random 2-coloring of the graph K_n. This means that each of the $\left(\begin{array}{c} n \\ 2 \end{array} \right)$ edges is colored independently red or blue, each with probability $1/2$. We want to estimate the probability of obtaining a monochromatic K_k. We divide this procedure up into three steps :

(i) the probability of a given K_k being entirely red is $\left[\frac{1}{2}\right]^{\left(\begin{array}{c} k \\ 2 \end{array} \right)}$. This follows from (11) and the fact that the probability of any particular edge being red is $1/2$. Obviously, we have the same expression for the probability of a given K_k being entirely blue.

(ii) hence, the probability of a given K_k being monochromatic is $2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}$. This follows from (i) and (10), since there are two mutually exclusive ways to have monochromaticity, namely redness or blueness, and these have equal probability.

(iii) hence, the probability of there being some monochromatic K_k is at most $\left(\begin{array}{c} n \\ k \end{array} \right) \cdot 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}$. This follows from the previous steps and (10), since there are $\left(\begin{array}{c} n \\ k \end{array} \right)$ complete k-subgraphs in K_n.

From these estimates it follows that, if (6) holds, then there is a positive probability that a randomly chosen coloring of K_n will include no monocrhomatic K_k. In other words, at least one such good coloring exists, and thus $R(k,k) > n$. This completes the proof of the theorem.

Remark 5 The proofs of Theorem 3 and Corollary 4 show that, if $k \geq 3$ and $n = 2^{k/2}$, then the probability that a random 2-coloring of K_n yields some monochromatic K_k is at most $2^{1+k/2}/k!$. This goes to zero very quickly as $k \to \infty$. Hence, for large k, a random coloring is highly likely to be successful. No reasonable deterministic algorithm for producing a successful coloring is known, however.

Remark 6 Small improvements to both the upper and lower bounds for $R(k,k)$, as given in (5) and (8) respectively, are known. To improve on (8) one can use a more refined probabilistic technique known as the Lovasz Local Lemma - we may do this later in the course. The currently best upper bounds can be found in [1] (note that the author is Irish and was a Ph.D. student at Cambridge at the time !!). However, all these improvements are at the end of the day very minor, as nothing tighter than (9) is yet known.

Example 2 : Van der Waerden numbers

DEFINITION 2 : Let $k \geq 1$. An increasing sequence $a_1 < a_2 < \cdots < a_k$ of k integers is said to be an *arithmetic progression of length k and common difference d* if $a_{i+1} - a_i = d$ for $i = 1, ..., k-1$. We will use the abbreviation 'k-AP' to denote an arithmetic progression of length k.

The following theorem was proven by the Dutch mathematician Bartel van der Waerden in the 1920s and has been given his name :

Theorem 7 (van der Waerden's Theorem) *Let $k, l \geq 1$ be given integers. Then for all sufficiently large positive integers n (depending on k and l), the following holds :*

If the integers $1, 2, ..., n$ are colored with at most l colors, then there must exist a monochromatic k-AP.

NOTATION : The *van der Waerden number* $W(k, l)$ is the least integer n for which any l-coloring of $\{1, ..., n\}$ must yield a monochromatic k-AP. The theorem states that these numbers exist.

It is beyond the scope of this course to give a fully rigorous proof of Theorem 7. A proof may be found, for example, in [2]. Basically, it involves two nice ideas and a lot of horrible notation. The two ideas are

(i) observe that $W(2, l) = l + 1$ (why ?). This allows us to get an induction started. The induction proceeds by proving that the numbers $W(k+1, l)$ exist for all l and a fixed k, assuming that the numbers $W(k, l)$ all exist.
(ii) for a fixed l and k, the proof of the existence of $W(k, l)$ in this inductive manner involves an idea which has become called *color focusing*. It is basically the same idea for all l but because the numbers involved grow so drastically with l, it becomes something of a technical nightmare to write down the details. The idea itself is quite beautiful, though.

We will be content to illustrate the method by proving that

$$W(3, 2) \leq 325. \tag{12}$$

Note that, according to the program outlined above, our proof of this should at some point use the knowledge that $W(2, l) = l+1$. I'll leave it as an amusing exercise for you to spot where this is used, since it would be easy to miss it !

So let us suppose the numbers from 1 through 325 have been colored red or blue in some manner. We must prove the existence of a monochromatic 3-AP. The first step is to divide the 325 numbers into 65 blocks $B_1, ..., B_{65}$ of 5 consecutive numbers. So $B_1 = \{1, 2, 3, 4, 5\}$, $B_2 = \{6, 7, 8, 9, 10\}$ etc.

There are $2^5 = 32$ possible ways to color any block with 2 colors. Thus, amongst the first 33 blocks, there must be two which are colored in exactly the same pattern. Pick any two such blocks, say B_i and B_{i+j}. Since $i + j \leq 33$, it follows that $i + 2j \leq 65$. Hence the block B_{i+2j} exists. We now focus our attention on the three blocks B_i, B_{i+j} and B_{i+2j}.

The rest of the proof is most easily understood with the help of pictures. I am not going to draw any pictures here, so I recommend that you look in [2].

Note that

$$\begin{aligned} B_i &= \{5i-4, 5i-3, 5i-2, 5i-1, 5i\}, \\ B_{i+j} &= \{5(i+j)-4, 5(i+j)-3, 5(i+j)-2, 5(i+j)-1, 5(i+j)\}, \\ B_{i+2j} &= \{5(i+2j)-4, 5(i+2j)-3, 5(i+2j)-2, 5(i+2j)-1, 5(i+2j)\}. \end{aligned}$$

Amongst the first three elements of the block B_i, at least two must get the same color. Let's suppose that $5i - 4$ and $5i - 2$ are both colored red and complete the proof in this case. The argument is similar in the other five cases and I leave it to yourselves to become convinced of that.

If now $5i$ was also colored red, then we'd have a red 3-AP, namely $\{5i - 4, 5i - 2, 5i\}$. So we may assume $5i$ is colored blue. Next, we turn to the block B_{i+j}. Since it has exactly the same color pattern as B_i, we conclude that $5(i + j) - 4$ and $5(i + j) - 2$ are both colored red, whereas $5(i + j)$ is colored blue.

Finally, now, we focus on B_{i+2j} and, in particular, zone in on the number $5(i + 2j)$. I claim that, no matter what color we give it, we can't avoid having a monochromatic 3-AP. For if this number is colored red, then $\{5i - 4, 5(i + j) - 2, 5(i + 2j)\}$ is a red 3-AP. But if is colored blue, then $\{5i, 5(i + j), 5(i + 2j)\}$ is a blue 3-AP. This completes the proof of (12).

The bounds on Van der Waerden numbers obtained by this kind of color focusing method are eeeeeeeenooooorrrrrmooooouuusssss[4]. We can see that the method is not optimal even for the example of $W(3, 2)$. Our method gives that $W(3, 2) \leq 325$. But, in fact, $W(3, 2) = 9$. To see this, first check by hand that for every partition of $\{1, 2, ..., 9\}$ into two subsets, at least one contains a 3-AP. On the other hand, we can 2-color the integers 1,...,8 so that there are no monochromatic 3-APs. For example, let 1,3,6,8 be red and 2,4,5,7 be blue.

Even the best-known upper bounds on van der Waerden numbers (obtained by quite different and, I think it is safe to say, more sophisticated methods)

[4]more precisely, they are not *primitive recursive*, for those of you who know what that means

are really, really big. We know that[5]

$$W(k,l) \leq 2^{2^{l^{2^{2^{k+9}}}}}. \tag{13}$$

We finish our discussion by instead obtaining lower bounds for the numbers $W(k,l)$ via a probabilistic argument.

Theorem 8

$$W(k,l) > \sqrt{2(k-1)}\, l^{(k-1)/2}. \tag{14}$$

PROOF : We need to show that if $n \leq \sqrt{2(k-1)}l^{(k-1)/2}$, then there exists an l-coloring of $\{1,...,n\}$ which yields no monochromatic k-AP. Fix n and consider a random l-coloring of $\{1,...,n\}$, i.e.: each number is independently assigned a color by tossing a fair l-sided die. Now fix a color $\mathcal{C}$ and a k-AP. The probability that this k-AP is monochromatic in color $\mathcal{C}$ is l^{-k}. Hence, the probability that this k-AP is monochromatic, in some color, is $l^{-(k-1)}$. If $f(n)$ is the total number of k-APs in $\{1,...,n\}$, then (10) implies that the probability of there being some monochromatic k-AP is at most $f(n) \cdot l^{-(k-1)}$. To estimate $f(n)$ we observe that a k-AP is determined by its first term and common difference. If the first term is $x \in [1,n]$, then the common difference cannot exceed $\frac{n-x}{k-1}$. This gives us the estimate

$$f(n) \leq \sum_{x=1}^{n-1} \frac{n-x}{k-1} = \frac{n(n-1)}{2(k-1)}. \tag{15}$$

Hence the probability of a random l-coloring of $[1,n]$ yielding a monochromatic k-AP is at most $\frac{n(n-1)}{2(k-1)l^{k-1}}$. We wish this quantity to be strictly less than one, and it is easy to see that this is the case when $n \leq \sqrt{2(k-1)}l^{(k-1)/2}$. This completes the proof of Theorem 8.

The gap between (13) and (14) is an important open problem in combinatorial number theory/Ramsey theory. The gap is obviously enormous. I think it is fair to say that most people believe that the lower bound (14), which gives exponential growth in k for a fixed l, is closer to the truth. But noone knows ... By the way, a slight improvement on (14) can also be obtained via the Lovasz Local Lemma, as we may see later.

References

[1] D. Conlon, *A new upper bound for diagonal Ramsey numbers*, Ann. of Math. **170** (2009), 941–960.

[2] R.L. Graham, B.L. Rothschild and J.H. Spencer, *Ramsey Theory*, Wiley (1990).

[5] this bound was obtained by Timothy Gowers only a few years ago as a consequence of his proof of what is known as *Szemeredi's theorem*, which is in itself a strengthening of van der Waerden's result. Gowers obtained the Fields Medal for this and other work.

Lecture 2 (Nov. 3, 2011)

We now introduce some basic terminology from probability theory, and explain the general principles behind the probabilistic method. We will then rewrite the proof of Theorem 3 in this more formal language. In subsequent applications we will move freely between less and more formal language as best suits the situation.

DEFINITON 3 : Let Ω be a set and $\mathcal{F} \subseteq 2^{\Omega}$, i.e.: $\mathcal{F}$ is a collection of subsets of Ω. We say that $\mathcal{F}$ is a *σ-algebra* if the following three conditions are satisfied :

(i) $\Omega \in \mathcal{F}$,

(ii) $\mathcal{F}$ is closed under complementation, i.e.: $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,

(iii) $\mathcal{F}$ is closed under countable unions, i.e.: if $A_1, A_2, ... \in \mathcal{F}$ then $\cup_{i \geq 1} A_i \in \mathcal{F}$.

DEFINITION 4 : Let Ω be a set, $\mathcal{F}$ a σ-algebra on Ω and $\mu : \mathcal{F} \to [0,1]$ a function. We call μ a *probability measure* if the following two conditions are satisfied :

(i) $\mu(\Omega) = 1$,

(ii) μ is *countably additive*, i.e.: if $A_1, A_2, ... \in \mathcal{F}$ and these sets are pairwise-disjoint, then

$$\mu\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mu(A_i). \tag{16}$$

DEFINITION 5 : A *probability space* is a triple $(\Omega, \mathcal{F}, \mu)$, where Ω is a set, $\mathcal{F} \subseteq 2^{\Omega}$ is a σ-algebra and $\mu : \mathcal{F} \to [0,1]$ is a probability measure.

If $A \in \mathcal{F}$ then the set A is said to be *μ-measurable*. More informally, A is called an *event*.

In this course, all probability spaces will be *discrete*, i.e.: the set Ω is countable and every subset is measurable, i.e.: $\mathcal{F} = 2^{\Omega}$. A discrete space is thus completely determined by the pair (Ω, μ).

In a discrete space, every singleton set is measurable. We will usually think of Ω as a subset of $\mathbb{N}$ and, for $i \in \Omega$, we will often write $\mu(\{i\}) := p_i$. More generally, for an event A, we usually denote $\mu(A) := \mathbb{P}(A)$, and call this "the probability of the event A".

If Ω is a finite set with n elements, say $\Omega := \{1, ..., n\}$, then the simplest probability measure on Ω is the one that assigns equal probability to each outcome, i.e.: $p_i = 1/n$ for each $i = 1, ..., n$. This is called *uniform measure* and corresponds most intuitively to the notion of the outcome being "random".

DEFINITION 6 : Let (Ω, μ) be a discrete probability space. A function

$X : \Omega \to \mathbb{R}$ is called a *(real-valued) random variable* on Ω[6]. From now on, we will write RV instead of "real-valued random variable".

If $\Omega = \{1, ..., n\}$, then the standard notation is $X(i) := x_i$.

DEFINITION 7 : If X is a random variable on the space (Ω, μ) then the *expected value/expectation/first moment* of X, denoted $\mathbb{E}[X]$, is the quantity

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} x_\omega p_\omega,$$

provided this sum is well-defined, in other words, provided the value of the sum does not depend on the order of summation. This is certainly the case if Ω is a finite set. If Ω is infinite, one has to be more careful, but for example everything is fine if X is non-negative, i.e.: if $x_\omega \geq 0$ for every $\omega \in \Omega$. One usually abbreviates this to "$X \geq 0$". Observe that, if $X \geq 0$, then $\mathbb{E}[X] \in [0, +\infty]$.

By another unfortunate historical accident, the expectation of a RV is often denoted by the letter μ, i.e.: the same letter as that used to denote the underlying probability measure.

In very general terms, an application of the probabilistic method to a problem in combinatorics involves understanding the moments[7] of some non-negative RV on some discrete probability space. Often, though by no means always, the space is finite and the measure uniform. In the simplest applications[8] only the first moment is needed. Note that non-negativity at least guarantees that the first moment is well-defined, even when the space is infinite. Henceforth, unless otherwise stated, all RVs are assumed to be non-negative.

We now state some simple facts about first moments which are used over and over again. The first property is informally referred to as *linearity of expectation* :

Proposition 9 (Linearity of expectation) *Let $X_1, ..., X_k$ be RVs on the same probability space (Ω, μ). Then*

$$\mathbb{E}[X_1 + \cdots + X_k] = \sum_{i=1}^{k} \mathbb{E}[X_i]. \tag{17}$$

[6]the term *random variable* is an unfortunate historical accident. A more accurate term would be *random function*. But the former term has become so conventional that no-one dares change it. It also explains why the letter X, rather than say f, is used to denote a random variable.

[7]We postpone a definition of the higher moments of a RV to later.

[8]Here I mean "simplest" in theoretical terms, though not necessarily in terms of the ingenuity of the application.

PROOF : Note that the sum of RVs on the left of (17) means just what one would expect, namely the pointwise sum of functions. The proof relies on being able to interchange the order of a double summation, which one can certainly do when the X_i are non-negative. Indeed, the proof then extends to an infinite sum of RVs. The details are left to the reader.

The next proposition goes under the informal title of the *averaging principle*. It is a formalisation of the simple idea that, given the average of something, there must be at least one instance which is no worse than average and at least one instance which is no better.

Proposition 10 *Let X be a RV on a space (Ω, μ). Assume that $\mathbb{E}[X]$ is finite. Then*
(i) $\mathbb{P}(X \geq \mathbb{E}[X]) > 0$ and
(ii) $\mathbb{P}(X \leq \mathbb{E}[X]) > 0$.

PROOF : The proof is trivial once one understands the notation. First, it is common to write $\mathbb{P}(\cdots)$ instead of $\mu(\cdots)$ when there can be no confusion as to what probability measure is being used. Second, the expression '$X \geq \mathbb{E}[X]$' is shorthand for the event $\{\omega \in \Omega : X(\omega) \geq \mathbb{E}[X]\}$. This kind of sloppy notation has become standard, so we will use it from now on without further comment.

The following corollary of Proposition 10(ii) is particularly useful :

Corollary 11 *If X is a non-negative integer-valued RV and $\mathbb{E}[X] < 1$, then $\mathbb{P}(X = 0) > 0$.*

One particular class of RVs which is especially useful in applications is the class of so-called *indicator* variables.

DEFINITION 8 : Let (Ω, μ) be a probability space and $A \subseteq \Omega$. The *indicator random variable of the event A*, denoted $\mathcal{X}_A$, is the random variable given by

$$\mathcal{X}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

Note that it is an immediate consequence of the definition that

$$\mathbb{E}[\mathcal{X}_A] = \mathbb{P}(A). \tag{18}$$

More generally, let $f : \Omega \to \Omega$ be any function. The *indicator random variable of the event '$f \in A$'*, denoted $\mathcal{X}_{f,A}$, is the random variable given by

$$\mathcal{X}_{f,A}(\omega) = \begin{cases} 1, & \text{if } f(\omega) \in A, \\ 0, & \text{if } f(\omega) \notin A. \end{cases}$$

The analogue of (18) is then

$$\mathbb{E}[\mathcal{X}_{f,A}] = \mathbb{P}[f^{-1}(A)], \tag{19}$$

where $f^{-1}(A) = \{\omega : f(\omega) \in A\}$. Note that (18) is the special case where f is the *identity map* on Ω, i.e.: $f(\omega) = \omega \; \forall \; \omega$. Also note that if f is a 1-1 mapping and μ is uniform measure, then

$$\mathbb{E}[\mathcal{X}_{f,A}] = \mu(A) = \frac{|A|}{|\Omega|}. \tag{20}$$

We are now ready to repeat the proof of Theorem 3 in more formal language. Let k, n be positive integers such that $\left(\begin{array}{c} n \\ k \end{array} \right) 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)} < 1$. Let Ω be the set of all possible 2-colorings of the edges of K_n, hence $|\Omega| = 2^{\left(\begin{array}{c} n \\ 2 \end{array} \right)}$. Let μ be uniform measure on Ω - this corresponds to 2-coloring the edges independently and fairly at random. We order the $\left(\begin{array}{c} n \\ k \end{array} \right)$ K_k-subgraphs of K_n in some way, and consider the random variables X_i, $i = 1, ..., \left(\begin{array}{c} n \\ k \end{array} \right)$, where $X_i = \mathcal{X}_{A_i}$ and A_i is the event that the i:th K_k-subgraph is monochromatic. As shown before, for each i,

$$\mathbb{E}[X_i] = \mathbb{P}(A_i) = 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}. \tag{21}$$

Let $X := \sum_i X_i$. Then X is the total number of monochromatic K_k-subgraphs. By (21) and Proposition 9 we have that

$$\mathbb{E}[X] = \left(\begin{array}{c} n \\ k \end{array} \right) \cdot 2^{1-\left(\begin{array}{c} k \\ 2 \end{array} \right)}. \tag{22}$$

Hence, by assumption, $\mathbb{E}[X] < 1$. Since X is non-negative integer-valued, Corollary 11 implies that $X = 0$ with positive probability. In other words, with positive probability, a random 2-coloring yields no monochromatic K_k-subgraph, Q.E.D.

We now give four examples where part (i) of Proposition 10 will be used.

Example 1 : MAXSAT problem

I took this material directly from Chapter 6 of [MU]. Please see the supplementary scanned document on the homepage.

Example 2 : Turán's Theorem

Let's first go back to van der Waerden's theorem. It is natural to conjecture, but apparently much harder to prove, the following stronger result :

Theorem 12 (Szemerédi's Theorem) *Let $k \geq 3$ and $\epsilon > 0$. Then for all sufficiently large n, depending on k and ϵ, if A is any subset of $\{1, ..., n\}$ such that $|A| > \epsilon n$, then A contains a k-AP.*

This theorem was first proven by the Hungarian mathematician Endre Szemerédi in 1975, in a 50-page paper which is generally considered "a masterpiece of combinatorial reasoning". The theorem had been conjectured by Erdős and Turán in the 1930s already when they worked (more or less unsuccessfully) on strengthening van der Waerden's result. The special case $k = 3$ was proven by Roth in 1952 using Fourier analysis, and this work was cited when Roth received the Fields Medal in 1956.

For our present purposes, what is of interest to us is the comparison with the situation for graphs. Ramsey's theorem (for an arbitrary number of colors - we just stated it for 2 colors earlier) may be considered the analogue of van der Waerden's theorem for graphs. The analogue of Szemerédi's theorem would then be the following :

Let $k \geq 3$ and $\epsilon > 0$. For all sufficiently large n, depending on k and ϵ, if G is a graph on n vertices and with more than $\epsilon \cdot \binom{n}{2}$ edges, then G must contain a K_k.

It is pretty easy to see, however, that this statement is false. In fact it is already false for $k = 3$ and $\epsilon = 1/2$. For let n be any even integer. Consider a graph on n vertices in which the vertices are partitioned into two subsets of size $n/2$ and in which two vertices are joined by an edge if and only if they lie in opposite halves of the partition. Such a graph is called *bipartite*. Now G has $\frac{n}{2} \cdot \frac{n}{2} = \frac{n^2}{4}$ edges, which is more than half of $\binom{n}{2} = \frac{n(n-1)}{2}$. But G contains no K_3, indeed no cycle of any odd length, since any path of odd length takes one from one side of the partition to the other.

We can generalise this example to any $k \geq 3$. For simplicity suppose that n is a multiple of $k-1$. A $(k-1)$-*partite* graph on n vertices is obtained by partitioning the vertices into $k-1$ subsets of equal size, and joining two vertices by an edge if and only if they do not lie in the same part. The total number of edges in this graph is

$$\binom{k-1}{2} \cdot \left(\frac{n}{k-1}\right)^2 = \frac{k-2}{k-1} \cdot \frac{n^2}{2},$$

which, as k gets bigger, heads towards 100 procent of all possible edges ! But the graph has no K_k since, at the very least, a K_k contains k vertices, hence (by the pigeonhole principle) at least two would have to come from the same part of the graph. But then they are not joined to one another - contradiction !

Turán's theorem, proven in 1941, is the statement that the above examples can't be improved upon.

Theorem 13 (Turán's Theorem) *Let $k \geq 3$ and n be a multiple*[9] *of $k-1$. Then any graph with n vertices and strictly more than $\frac{k-2}{k-1} \cdot \frac{n^2}{2}$ edges contains a K_k.*

This theorem can be proven in a number of ways. Next day, we will present a beautiful proof which uses a probabilistic method.

[9] If n is not a multiple of k then one can prove a correspoding result anyway, but I wish to avoid the associated technicalities in this presentation. If $n = (k-1)q + r$ say, where $0 < r < k-1$, then the optimal way to avoid a K_k is to take a $(k-1)$-partite graph, where r of the parts have $q+1$ vertices each and the remaining $k-1-r$ parts have q vertices each.