Probabilistic Methods

Yang Siwei, Allen
School of Physical and Mathematical Sciences

Assoc. Prof. Wu Guohua
School of Physical and Mathematical Sciences

Abstract - In combinatorics, properties of combinatorial structures are often proved via constructive methods. To show that a class of objects exhibits a certain property, or specifically the existence of a property, we need only to construct an object with said property. This proves its existence. However, often such an approach is either infeasible or computationally tedious. In this project, we investigate an alternative to such methods, known as the probabilistic method. Here, a survey of combinatorial methods and proofs is conducted. We look through content of a multitude of graduate-level texts and survey papers. Thereafter, we delve into standard principles of classical information theory. Here we formally explain principles such as Shannon Entropy as well as the intuition behind said principles. We also analyse various combinatorial structures in information theory, serving as a natural progression of the survey. This is then followed through with Quantum Information Theory.

1 INTRODUCTION

The probabilistic method, popularised by Erdos[1], is a combinatorial method with applications to areas such as graph theory[3] and number theory[4,5]. It provides an alternative to constructive proofs in combinatorics. To prove that a combinatorial object exhibits a certain property, we could construct such an object to prove the existence of said property. Alternatively, we could create an appropriate probability space and then prove that a random element of the space exhibits the property with non-zero probability. This would be the essence of the probabilistic method[1].

Through surveying a collection of notes and papers, we here present instances of the probabilistic method as well as other well-known combinatorial proofs. This will cover topics of Ramsey Theory[1,2,5,7] and van de Waerden numbers[4,5].

From here, we then survey classical information theory, explaining and providing intuition for standard concepts such as Entropy[6]. This will then lead into Quantum Information theory towards the end[8].

2 RAMSEY THEORY

2.1 RAMSEY NUMBERS

We explore the two-coloured version of Ramsey's Theory, with its respective definitions, theorems, and proofs. In this section, 2.1, edges are also defined to consist only of two vertices.

Definition 2.1.1

A graph G = (V,E) is a set of vertices and edges where V(G) and E(G) are the sets of vertices and edges in G respectively.[2]

Definition 2.1.2

A complete graph on n vertices, denoted K_n, is a graph in which every vertex is adjacent, or connected by an edge, to every other vertex in G.[2]

Definition 2.1.3

A clique is a subset of vertices such that there exists an edge between any pair of vertices in that subset of vertices. i.e having a complete subgraph.[2]

Definition 2.1.4

An independent set of a graph is a subset of vertices such that there exists no edges between any pair of vertices in that subset.[2]

Definition 2.1.5

Let C be a set of colours $\{c_1,c_2,\ldots,c_m\}$ and E(G) be the edges of a graph G. An edge colouring $f:E\rightarrow C$ assigns each edge in E(G) to a colour in C. If an edge colouring uses k colours on a graph, it is referred to as a k-coloured graph.[2]

Definition 2.1.6

A Ramsey Number, n = R(r,b), is the smallest integer n such that the 2-coloured graph K_n, using colours red and blue for edges, implies a red monochromatic subgraph K_r or a blue monochromatic subgraph K_b.[2]

We now state Ramsey's Theorem without proof for the sake of brevity.

Theorem 2.1.1

Let $r \in \mathbb{N}$. Then there exists an $n \in \mathbb{N}$ such that any 2-coloured K_n graph contains a monochromatic (1-coloured) subgraph K_r of K_n.[2]

We now present a theorem on the symmetry of Ramsey numbers, followed by proofs of known Ramsey numbers.

Theorem 2.1.2

$\forall$ r, b $\in \mathbb{N}$, R(r,b) = R(b,r).[2]

Proof:

This proof is intuitive and arises naturally from the symmetry of graphs. Given a two-colouring of red(r) and blue(b), it is always possible to invert the colouring such that the red edges are re-coloured as blue edges and vice versa.

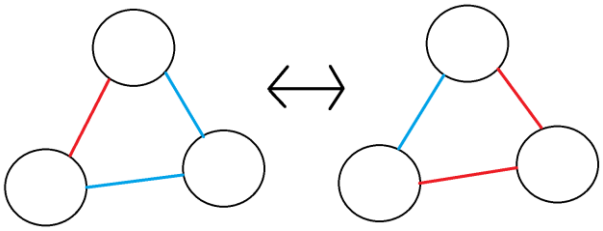


Figure 1

Theorem 2.1.3

R(1,k) = 1.[2]

Proof:

A monochromatic K_1, by definition, would be a single vertex, requiring no edges. A "red" K_1 would therefore only require 1 vertex.

Theorem 2.1.4

R(2,k) = k. [2]

Proof:

A K_2 monochromatic subgraph coloured red would be an edge between two vertices, with the edge coloured red. Give a two-colouring on K_n, we split this into two cases.

Case 1:

If there exist an edge coloured red, then there exists a red K_2 subgraph of K_n. Hence, we would be done. There requirement on n is n>=2 here.

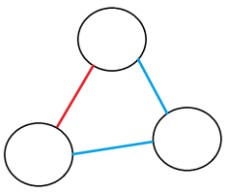


Figure 2

Case 2:

If there does not exist an edge coloured red, then all edges are coloured blue. To satisfy R(2,k), there have to be at least k blue edges. The requirement on n is n>=k here.

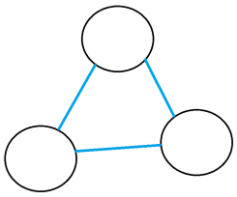


Figure 3

Therefore, to satisfy both cases, R(2,k) = k.

The next theorem is known as the party problem. It refers to finding the minimum number of people(R(3,3)) invited to a party such that there are at least 3 friends(r) or three strangers(b).

Theorem 2.1.5

R(3,3) = 6.[2]

Proof:

We prove this by showing 5 < R(3,3) ≤ 6.

Consider K_5. We observe that the following colouring given below has no K_3 monochromatic subgraph.

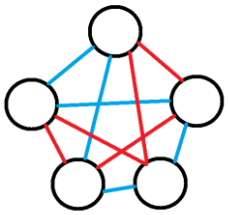


Figure 4

Therefore 5 < R(3,3).

To show R(3,3) ≤ 6, consider K_6. We fix a vertex, labelling it A. There are at least 3 edges emanating from A that are of the same colour. Without loss of generality, suppose they are coloured red.

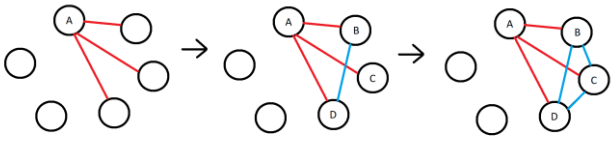


Figure 5

We label the vertices at the ends of those edges as B,C,D. To prevent a monochromatic K_3 subgraph being formed between A,B,D, the edge BD has to be coloured blue. Similarly for edges BC and CD, to prevent A,B,C and A,C,D from forming monochromatic K_3 subgraphs respectively. This

however forms a monochromatic K_3 subgraph coloured blue. Therefore, there will always be a monochromatic K_3 subgraph in K_6, i.e $R(3,3) \leq 6$.

This shows $R(3,3) = 6$.

The next theorem is an identity for the upper bound of Ramsey Numbers. It is required later for the theorems following it.

2.2 BOUNDS

We now proceed to prove generalised bounds of Ramsey Numbers. In this subsection, we also provide an instance of the probabilistic method, as proved by Erdos.[2]

Theorem 2.2.1

$\forall$ r, b $\in \mathbb{N}$, $R(r,b) \leq R(r-1,b) + R(r,b-1)$.[2]

Proof:

Let G be a two-coloured $K_{R(r-1,b) + R(r,b-1)}$. Fix a vertex $v \in G$. Denote the number of vertices adjacent to v via a red edge as n_r and the set S_r to be the set of these vertices. Similarly, n_b and S_b for those via a blue edge.

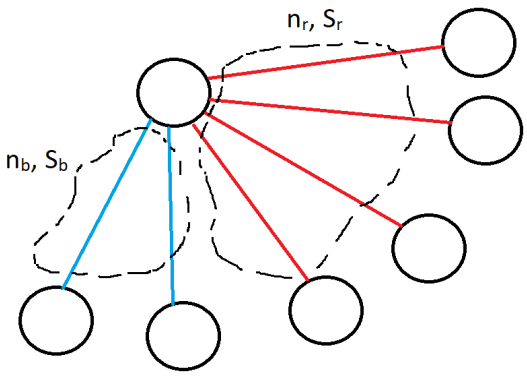


Figure 6

The total number of vertices is therefore n_b+n_r+1. That is, $n_r + n_b + 1 = R(r-1,b) + R(r,b-1)$.

From here, we consider two cases and show that the inequality in Theorem 2.1.6 holds for both.

Case 1:

If $n_r < R(r-1,b)$, then $R(r,b-1) \leq n_b$. This implies that the complete subgraph formed by S_b contains a monochromatic complete subgraph of r vertices, edges coloured red, or b-1 vertices, edges coloured blue. If it is of r vertices, then we are done.

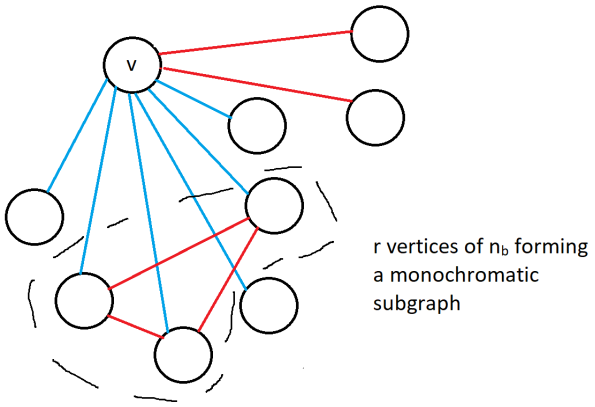


Figure 7

If it is of b-1 vertices, then we can consider complete subgraph induced by the b-1 "blue edged" vertices with v. As v is connected to every vertex in S_b by a blue edge, this subgraph is monochromatic with edges coloured blue. This subgraph also has cardinality b - 1 + 1 = b. There would therefore be a monochromatic subgraph of b vertices coloured blue in this case.

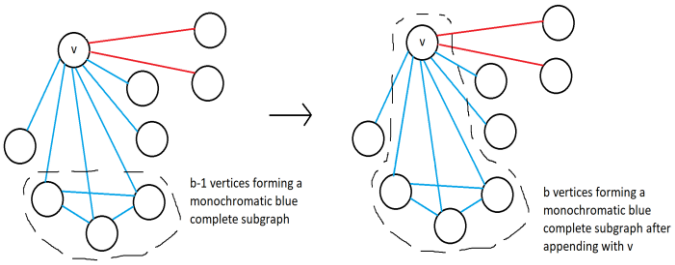


Figure 8

Case 2:

If $R(r-1,b) \leq n_r$, the complete subgraph formed by S_r contains a monochromatic complete subgraph of r-1 vertices, with edges coloured red, or b vertices, with edges coloured blue. If it is of b vertices then we are done as R(r,b) will be satisfied.

If it is of r-1 vertices, then we append v to this set of vertices. Since v is connected to every element of S_r by a red edge, this set will for a monochromatic red complete subgraph of r - 1 + 1 = r vertices. R(r,b) will then also be satisfied.

From this, we proceed to prove another combinatorial identity of Ramsey Numbers.

Theorem 2.2.2

$$R(r,b) \leq \binom{r+b-2}{r-1}$$ [2]

Proof:

We proceed by a double induction on variables r, b.

Base case:

When r, b = 2,

$$R(2,2) = 2 \leq 2 = \binom{2+2-2}{2-1}$$

Induction step:

Suppose the inequality holds for r = x-1, b = y and r = x, b = y - 1. Then,

$$R(r,b) \leq R(r-1,b) + R(r,b-1) \leq \binom{(r-1)+b-2}{(r-1)-1} + \binom{r+(b-1)-2}{r-1} = \binom{r+b-2}{r-1}$$

This proves the inequality.

The next theorem demonstrates the efficacy of the probabilistic method.

Theorem 2.2.3

If

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1,$$

then R(k,k) > n.[7]

It should be noted that this generalises to $2^{k/2}$ < R(k,k), for 3 ≤ k.

Proof:

The idea for the proof is that we condition on the existence of a monochromatic subgraph. To be precise, we compute the probability that there exists at least 1 subgraph of size k that is monochromatic. We then show that this probability is less than 1. That is, there is non-zero probability that there exists no monochromatic subgraph of size k.

As there are 2 possible colourings for each edge, we multiply by 2.

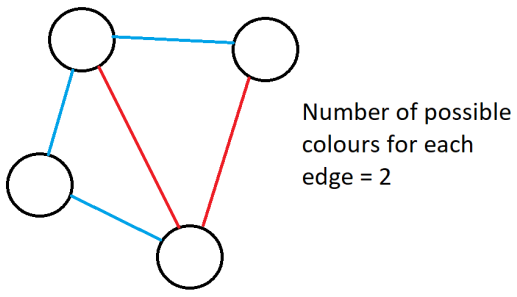


Figure 9

As we have n choose k ways to select k vertices for the complete subgraph, we multiply by n choose k.

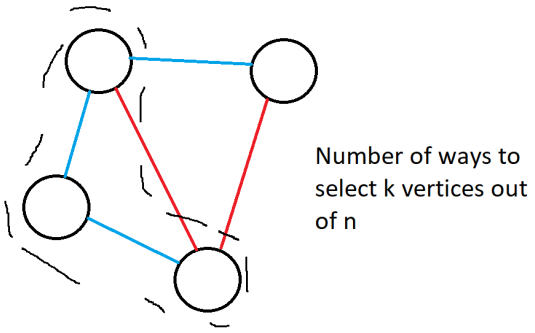


Figure 10

As we require that it is monochromatic and there are k choose 2 edges, we multiply by,

$$0.5^{\binom{k}{2}}.$$

As a result, we have

$$P(K_k \text{ is monochromatic}) = 2\binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}}.$$

If this is less than 1, we then have that there is non-zero probability that there exists no monochromatic subgraph of size k. This is equivalent to the inequality in the theorem. There therefore exists a monochromatic subgraph of size k.

Theorem 2.2.4

R(k,k) < 4^k.[2]

Proof:

To show this, we use a combinatorial result as stated below.

$$R(k,k) \leq \binom{2k-4}{k-2}$$

We proceed to show that the right-hand side is less than 4^k.

$$\begin{aligned}
\binom{2k-4}{k-2} &= \frac{(2k-4)!}{((2k-4)-(k-2))!(k-2)!} \\
&= \frac{((2k-4)(2k-6)\ldots(2))((2k-5)(2k-7)\ldots(1))}{(k-2)!(k-2)!} \\
&= \frac{2^k(k-2)!((2k-5)(2k-7)\ldots(1))}{(k-2)!(k-2)!} \\
&= \frac{2^k(2k-5)(2k-7)\ldots(1)}{(k-2)!} \\
&= 2^k \frac{2k-5}{k-2}\frac{2k-7}{k-3}\cdots\frac{1}{1} \\
&< (2^k)(2^k) = 4^k
\end{aligned}$$

3 VAN DER WAERDEN NUMBERS

3.1 VAN DER WAERDEN'S THEOREM

In this section, we present Van Der Waerden's Theorem, which discusses the existence of monochromatic l-term arithmetic progressions of the positive integers, for ranges [1,N]. We extend this in the next section to discuss various bounds of N,

Definition 3.1.1

An *l*-term arithmetic progression is any set of form: a, a+d, a+2d, …, a+(*l*-1)d, where a,d $\in \mathbb{R}$, d≠0.[5]

Definition 3.1.2

A k-colouring of a set A is any function

C: A→{1,2,…,k} = [1,k]. [5]

Definition 3.1.3

If c is a k-colouring of set A and if B⊆A is such that for any x,y $\in$ B, c(x) = c(y), then we describe set B as monochromatic.[5]

Definition 3.1.4

Colour-focused arithmetic progressions and spikes: Let c be a finite colouring of an interval of positive integers [1,m] and l,r $\in \mathbb{N}$. A set of of l-term arithmetic progressions $A_1, A_2,\ldots,A_r$, where $A_i = \{a_i+jd_i: j \in [0, l-1], a_i,d_i \in \mathbb{N}\}$ is colour-focused at f $\in \mathbb{N}$ if

- $A_i \subseteq [1, m]$ for each i $\in$[1,r].
- Each A_i is monochromatic.
- If i ≠ j then A_i and A_j are of different colour,
- $a_1 + ld_1 = a_2 + ld_2 = \ldots = a_r + ld_r = f$.

The elements of a colour-focused set are referred to as spikes. [5]

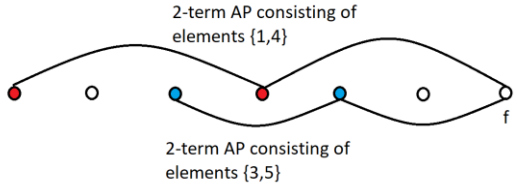


Figure 11

Theorem 3.1.1

Van der Waerden's Theorem – Let l,k $\in \mathbb{N}$. Any k-colouring of positive integers contains a monochromatic l-term arithmetic progression.

Moreover, there is a natural number N denoted as W(l,k), the Van der Waerden number, such that any k colouring of the segment of positive integers [1,N] contains a monochromatic l-term arithmetic progression. [5]

For brevity, we provide only the proof of the case for l = 3.

Proof:

We proceed to show the existence of W(3,k). When k = 1, this is trivial. We proceed to work with 2 ≤ k.

To continue the proof, we prove the following statement:

For all r ≤ k, $\exists$ n $\in \mathbb{N}$ such that when [1,n] is k-coloured, there exists a monochromatic 3-term arithmetic progression or there exists r colour-focused arithmetic progressions of length 2.

We prove this via induction on r.

Base case:

For r = 1, we choose n = k + 1. By choosing n = k + 1, we have that there will be at least 2 elements of [1,n] that are of the same colour. We label them a_1, a_2.

Case 1:

If $(a_2 – a_1) + a_2 \notin$ [1,n], this forms a 1 element colour-focused set of arithmetic progressions.

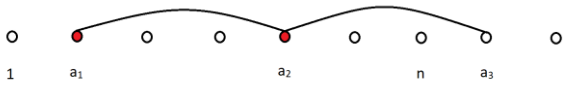


Figure 12

Case 2:

If (a2 – a1) + a2 $\in$ [1,n], then this forms a monochromatic 3-term arithmetic progression.

Figure 13

Inductive step:

Suppose that for r $\in$ [2,k], $\exists$ n $\in \mathbb{N}$ such that when [1,n] is k-coloured, there exists a monochromatic 3-term arithmetic progression or there exists r-1 colour-focused arithmetic progressions of length 2.

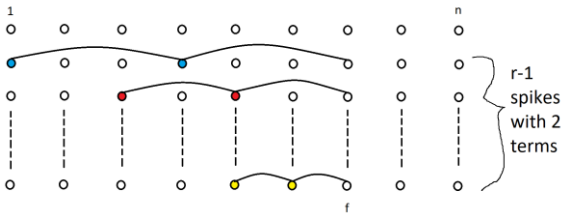


Figure 14

Here, we consider blocks of size [1,2n] which will make up the interval we build. There are also k^{2n} possible colourings of the interval [1,2n].

From here, we consider the interval $[1,(2n)(k^{2n}+1)]$. We divide this interval into $k^{2n}+1$ blocks, each of length 2n. Denote each block as B_i, $1 \leq i \leq k^{2n}+1$.

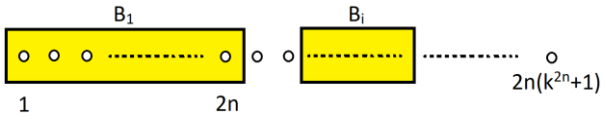


Figure 15

Let a k-colouring be given on the interval $[1,(2n)(k^{2n}+1)]$. If it contains a 3-term arithmetic progression, then we are done. Else, suppose it does not. We attempt to show that there exists r colour-focused arithmetic progressions of length 2.

By the inductive hypothesis, each B_i block has r-1 spikes as well as their focus, f, due to the length of each B_i being 2n.

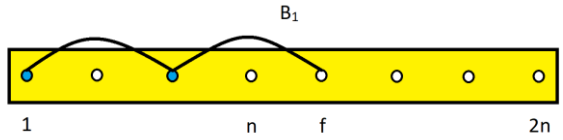


Figure 16

Furthermore, as there are only k^{2n} possible colourings for each block, having $k^{2n}+1$ blocks guarantees that at least two blocks have the same colouring. Label them B_i and B_j.

Figure 17

Consider B_i+B_j. To generate each of the r-1 spikes, we consider the following arithmetic progression. Let $\{a_{i1}, a_{i2}\}$ be any of the r-1 spikes of B_i, with a_{i3} being the focus corresponding to this spike, also in B_i. Let the difference between the first terms of B_i and B_j be d. Choose $a_{j1} = a_{i1} + d$ and $a_{j2} = a_{i2} + d$. Then $\{a_{j1}, a_{j2}\}$ will be the corresponding spike in B_j with a_{j3} being the focus of this spike, also in B_j. It can then be observed that $\{a_{i1}, a_{j2}\}$ forms a spike in B_i+B_j. Repeating this for each of the r-1 spikes in B_i, we attain r-1 spikes in B_i+B_j.

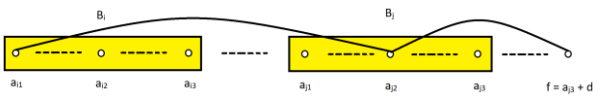


Figure 18

At the same time we use each focus, $\{a_{i3},a_{j3}\}$, which forms another spike. Each of these spikes intersect at $a_{j3}+d$, which is the new focus.

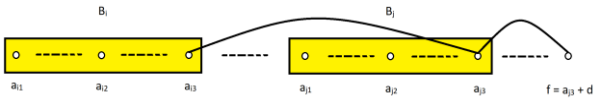


Figure 19

We therefore have r colour-focused 2-term arithmetic progressions in the interval $[1,(2n)(k^{2n}+1)]$. Setting r = k, we have that there either exists a monochromatic 3-term arithmetic progression or there exists k colour-focused arithmetic progressions of length 2.

If there exists a monochromatic 3-term arithmetic progression, we are done. Else, we need only consider $N = 2(2n)(k^{2n}+1)$, which will include the focus of the k spikes, resulting in a monochromatic 3-term arithmetic progression.

3.2 BOUNDS

We now proceed to introduce bounds on Van der Waerden Numbers. To do so, we first introduce two theorems on arithmetic progressions which shall be of use later.

Theorem 3.2.1

Let k, n be positive integers. Given a k-arithmetic progression of [n], the number of k-arithmetic progressions that intersect it is less than kn.[4]

Proof:

Let $x \in [n]$. We first produce a bound on the number of k-arithmetic progressions containing it. Suppose x is the i^{th} element of some k-arithmetic progression with step d, with $i \in [1,k]$. Then $1 \leq x-(i-1)d$, since the arithmetic progression has to begin at 1. Also, $x+(k-i)d \leq n$ since the k-arithmetic progression is an arithmetic progression of [n]. We proceed to use these lower and upper bounds.

We fix a k-arithmetic progression and suppose other arithmetic progressions intersect it at a fixed x. That is, i and d are fixed. After determining the number that intersect at this x, we then sum over all possible x's, thereby establishing the number of arithmetic progressions which intersect it.

We suppose that k is even as the odd case would be similar. For a fixed x, we apply the upper bound, for d when $i \leq k/2$,

$$d \leq \frac{n-x}{k-i}$$

Similarly, we apply the following upper bound for d when $i > k/2$.

$$d \leq \frac{x-1}{i-1}$$

Notice that for a position i of a k-arithmetic progression, the number possible such progressions would correspond to the number of possible values of d.

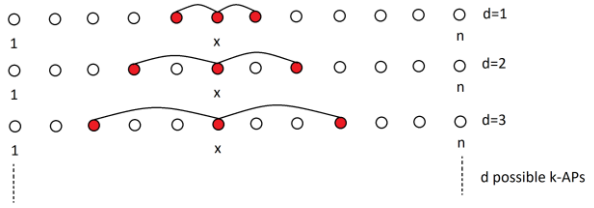


Figure 20

We therefore establish a bound for d and thereby a bound for the number of k-arithmetic progressions for a fixed x. Fixing x, we sum over possible values of i to do so.

$$\sum_{i=1}^{\frac{k}{2}} \frac{n-x}{k-i} + \sum_{i=\frac{k}{2}+1}^{k} \frac{x-1}{i-1} = (n-x)\sum_{i=1}^{\frac{k}{2}} \frac{1}{k-i} + (x-1)\sum_{i=\frac{k}{2}+1}^{k} \frac{1}{i-1}$$

$$= (n-x+x-1)\sum_{i=\frac{k}{2}+1}^{k} \frac{1}{i-1} \leq n-1 < n$$

We therefore have that for a fixed x of a k-arithmetic progression, there are at most n k-arithmetic progressions passing through x. As there are k positions in this k-arithmetic progression, there are k possible x's. The number of possible k-arithmetic progressions that intersect is hence kn.

In the next theorem, we provide another instance of the probabilistic method.

Theorem 3.2.2

The number of k-arithmetic progressions of [n] is less than n^2/k.[4]

Proof:

Fixing a∈[n] starting point of a k-arithmetic progression, we establish a bound on d. Similar to the theorem above, as d counts the number of possible k-arithmetic progressions, the bound on d corresponds to the bound on such possible progressions. We then sum over possible values of a for each of the possible start points.

$$\sum_{a=1}^{n-1} \frac{n-a}{k-1} = \frac{1}{k-1}\sum_{a=1}^{n-1}(n-a) = \frac{n(n-1)}{2(k-1)} < \frac{n^2}{k}$$

Theorem 3.2.3

$$W(k,2) \geq \sqrt{\frac{k}{3}}\,2^{\frac{k-1}{2}}$$ [4]

Proof:

$$\text{Let } n = \sqrt{\frac{k}{3}}\,2^{\frac{k-1}{2}}$$

Let there be an induced colouring on [n], where ∀x∈[n], there is 0.5 probability that x is coloured red and 0.5 probability that x is coloured blue.

Here, we employ the probabilistic method. We calculate the probability p that there exists a monochromatic k-arithmetic progression and then show that it is less than 1. This would imply that there is at least 1 instance of colouring where there is no monochromatic k-arithmetic progression, thereby establishing a lower-bound for the van der Waerden number.

As each x∈[n] has probability 0.5 of being a specified colour, the probability that all elements of a k-arithmetic progression are a specified colour is 0.5^k. We therefore multiply by 0.5^k.

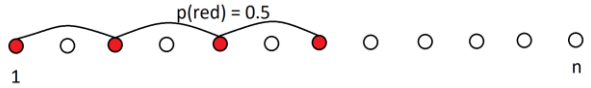


Figure 21

Since there are 2 such possible colours, red and blue, we multiply by 2.

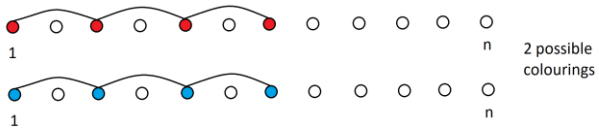


Figure 22

As the number of possible k-arithmetic progressions are bounded by n^2/k we multiply by n^2/k.

We therefore have,

$$p \leq \left(\frac{n^2}{k}\right)0.5^{(k-1)} = \frac{n^2}{k2^{(k-1)}} = \frac{1}{3} < 1,$$

which completes the proof.

4 CLASSICAL INFORMATION THEORY

4.1 ENTROPY

We now proceed to delve into information theory, beginning with Entropy. We will consider its motivations, definition as well as intuition.

Definition 4.1.1

An ensemble is the set of outcomes of one or more random variables.[6]

Entropy provides a representation of the information content of an ensemble. As it draws from random variables, it can be observed later on that it exhibits properties that have strong connections to various properties in probability.

Definition 4.1.2

Entropy H of an ensemble is defined as

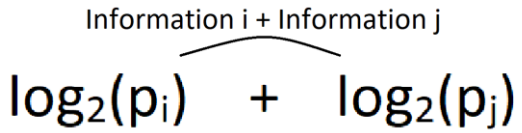
$$H = -\sum_i p_i \log p_i$$

where p_i is the probability of the event i of the ensemble.[6]

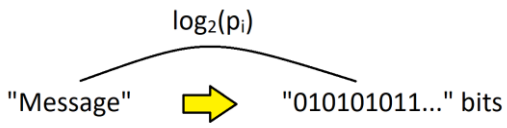
Intuition:

We provide intuition by first considering the case when the ensemble consists only of 1 element. That is, we consider the entropy of a single event. When $p_i = 1$, we can see that H = 0. This corresponds to gaining 0 information or there being 0 uncertainty from an event that we know was completely certain. This provides a justification for the log function in the equation as it reduces H to 0 in this scenario.

Another justification for the log function arises from its additive property. Given multiple receipts of information, it is desirable to quantify the total information by summing the information of each receipt. This contrasts with multiplying probabilities to find the probability of a given event.



We now consider the case of a general ensemble. An alternative to taking the definition at face value would be to interpret H as the expected number of bits required to encode a message. Here p_i represents the probability of a given message being sent. Taking negative log with base 2, we then encode the probability of the message into bits.



Multiplying this by p_i, we are then computing the expected value or weighted cost of bits required to encode said message. Summing over the i's of the ensemble, we therefore attain the expected number of bits required to encode a message.

Definition 4.1.3

Joint ensemble XY denotes an ensemble whose outcomes are ordered pairs x,y where x is from the set of outcomes corresponding to random variable X and y for random variable Y.[6]

Definition 4.1.4

Joint entropy H of XY is defined as

$$H(X,Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)}$$ [6]

This mirrors the definition of joint probability.

Theorem 4.1.1

Joint entropy is additive if X and Y are independent random variables. That is,

H(X,Y) = H(X) + H(Y) iff p(x,y) = p(x)p(y). [6]

Proof

For brevity, we prove only the backwards direction. Suppose p(x,y) = p(x)p(y). Then,

$$\begin{aligned} H(X,Y) &= \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)} \\ &= \sum_{x,y} p(x,y)(-\log(p(x)p(y))) \\ &= \sum_{x,y} p(x)p(y)(-\log p(x) - \log p(y)) \\ &= \sum_{x,y} p(x)p(y)(-\log p(x)) + \sum_{x,y} p(x)p(y)(-\log p(y)) \\ &= \sum_{x} p(x)(-\log p(x)) + \sum_{y} p(y)(-\log p(y)) \\ &= H(X) + H(Y) \end{aligned}$$

Definition 4.1.5

Conditional entropy of an ensemble X, given an ensemble Y is defined as

$$H(X|Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x|y)}$$ [6]

This can be interpreted as measuring the average uncertainty about X when Y is known.

Theorem 4.1.2

H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y). [6]

Intuitively, this can be regarded as the equivalent of the chain-rule for Entropy.

Proof:

We show only H(X,Y) = H(X) + H(Y|X) as the other will only require a reparameterization of this proof.

$$
\begin{aligned}
H(X) + H(Y|X) &= \sum_x p(x) \log \frac{1}{p(x)} + \sum_{x,y} p(x,y) \log \frac{1}{p(y|x)} \\
&= \sum_{x,y} p(x,y) \left(-\log p(x)\right) - p(x,y) \log\left(p(y|x)\right) \\
&= \sum_{x,y} p(x,y) \left(-\log p(x) p(y|x)\right) \\
&= \sum_{x,y} p(x,y) \left(-\log p(x,y)\right) \\
&= H(X,Y)
\end{aligned}
$$

We now conclude the section on entropy by establishing a upper bound. This is referred to as the Independence Bound on Entropy.

Theorem 4.1.3

$$H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^{n} H(X_i)$$ [6]

Note that this upper bound is attained if all random variables are independent.

Proof:

This follows directly from Theorem 4.1.2.

4.2 MUTUAL INFORMATION

We now attempt to characterise the amount of information one random variable conveys about the other. This is known as mutual information.

Definition 4.2.1

Mutual Information is defined as

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$ [6]

This measures the average reduction in uncertainty about X that comes about from learning of Y.

Intuition:

Firstly, it is to be noted that this retains symmetry, I(X;Y) = I(Y;X). This means that the information X provides about Y is as much as the information Y provides about X.

It is also to be noted that when X and Y are independent, the log term is 0, thereby setting I(X;Y) to 0. This is in line with the notion that independence implies one does not affect another. No information is then conveyed about one from another.

Thirdly, it can be observed that I(X;X) = H(X). This shows that when two random variables are perfectly correlated, mutual information reduces to the entropy of a single variable.

Next, we state three properties of mutual information and note the intuition.

I(X;Y) = H(X) – H(X|Y)

I(X;Y) = H(Y) – H(Y|X) = I(Y;X)

I(X;Y) = H(X) + H(Y) – H(X,Y)

These correspond to set theoretic laws, allowing us to interpret I(X;Y) as the intersections of H(X) and H(Y).

5 QUANTUM INFORMATION THEORY

5.1 *CBIT NOTATION*

We first introduce Cbits in quantum notation. Here, intuition and explanations of the respective representations, notations and operations will be provided. This will then be extended to Qbits.

First, we consider a single Cbit. It is a representation of information with two possible states - 0 and 1. To translate this into quantum terminology, this can be thought of as a box, | ⟩, within which contains the state 0 or 1. That is, |0⟩ and |1⟩. This “box” is known as Dirac notation.[8]

We now extend this to strings of bits such as 1101. In this case, it will be of the form |1⟩|1⟩|0⟩|1⟩. For the sake of notational brevity however, this can be simplified to |1101⟩.

Binary Expansion Representation

A third possible notational form involves considering the bit strings as binary expansions of integers. For instance, 1101 would represent $1x2^3 + 1x2^2 + 0x2^1 + 1x2^0 = 13$. Here, we can regard |1101⟩ to be |13⟩. However, it can be observed that such notation raises an issue. Given |13⟩ for instance, we are unable to ascertain as to whether the initial string was a representation of the 4-Cbit state, |1101⟩, or the 5-Cbit state, |01101⟩. To rectify this, a subscript which denotes the Cbit state is added. For instance, $|13\rangle_5$ represents |01101⟩.

Vector Representation

A final way to represent Cbits would be to represent them as a column vector whose entries are all 0s except for one i^{th} entry. This i^{th} entry will have value 1 and corresponds to the Cbit having state |i⟩. It is to be noted that the first entry is regarded as the 0^{th} entry. The length of the column vector would be of form 2^n, where n is the length of the bit string. For instance,

$$|5\rangle_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$.[8]

There exists also a more formal way of attaining the representation above. This invokes the usage of tensor products.

To do so, first notice that each entry of any Cbit string, |1⟩ or |0⟩ can be represented as a vector in $\mathbb{R}^2$. As convention, these will be represented as per below.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Informally, the tensor product, ⊗, of two vectors, will be a column vector whose entries consist of all possible products of entries of the two vectors. If the two vectors are of length M and N, the tensor product will produce a vector of length MN.

We can now translate the representation of the binary expansion in integer form to that above. We first convert it into the bit-string form and then express it as a tensor product of each bit. This is demonstrated below.

$$|5\rangle_3 = |101\rangle = |1\rangle|0\rangle|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

We therefore attain the column representation as desired.

5.2 *CBIT OPERATIONS*

Operations can be classified into two categories, reversible and irreversible.

Erasure

An example of an irreversible operation would be the erasure operation,

E: |x⟩ → |x'⟩, where E(|1⟩) = |0⟩, E(|0⟩) = |0⟩

This sends every bit to 0, making it impossible to deduce the prior state.[8]

NOT

X: |x⟩ → |x'⟩, where X(|1⟩) = |0⟩, X(|0⟩) = |1⟩

The NOT operation involves flipping string bits. It is reversible as it is its own inverse. That is, XX = 1. It should also be noted that in the case of a single bit string, the NOT operation is the only non-trivial reversible operation.[8]

Due to the vector representation of Cbits, we can express operations as matrices. For the case of the NOT operator and single Cbit, the matrix is as follows:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

SWAP

The SWAP operation, S_{ij}, interchanges the i^{th} and j^{th} positions of the Cbits. For instance, $S_{10}(|xy\rangle) = |yx\rangle$. This operation is also reversible as it is its own inverse. It can also be represented as a permutation matrix.[8]

CNOT

The controlled-NOT operator C_{ij} (CNOT) is a 2-Cbit operator. It works as follows:

If the i^{th} Cbit position has value |1⟩, the NOT operator will be applied to the j^{th} position. Else, nothing will happen.[8]

5.3 *QBITS*

We now introduce Qbits. The state, |ψ⟩, of a Qbit is of the form:

$$|\psi\rangle = a_0|0\rangle + a_1|1\rangle = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$$

where a_0 ,$a_1 \in \mathbb{C}$ such that ||ψ⟩| = 1.[8]

It can be noted that when $a_0 = 1$ and $a_1 = 0$ or $a_0 = 0$ and $a_1 = 1$, the Qbit reduces to a Cbit case.

Unlike the Cbit, whose reversible operation is only the NOT operation, the Qbit is not subject to such restrictions. Operations on the Qbit and Qbits require only that the transformation applied be unitary. That is, for a given operation represented by U,

$$UU^{\dagger} = 1$$.[8]

6 CONCLUSION

To conclude, here we present a survey of different combinatorial methods, structures and applications. The probabilistic method is only one of many in the large field of combinatorics. Further work may attempt to apply the method to arising combinatorial structures.

ACKNOWLEDGMENT

I would like to express my gratitude to Assoc. Prof. Wu Guohua for his continued guidance. He introduced me to and his explanations allowed for me to delve into the large variety of fields in Mathematics.

I would like to acknowledge the funding support from Nanyang Technological University – URECA Undergraduate Research Programme for this research project..

REFERENCES

[1] Alon, N., & Spencer, J. H. (2016). The Probabilistic Method (Wiley Series in Discrete Mathematics and Optimization) (4th ed.). Wiley.
[2] Barton, L. (2016). Ramsey Theory.
[3] Diestel, R. (2010). Graph Theory (Vol. 173). Heidelberg; New York: Springer. ISBN: 9783642142789 3642142788 9783642142796 3642142796
[4] Gasarch, W., & Haeupler, B. (2011). Lower Bounds on van der Waerden Numbers: Randomized- and Deterministic-Constructive. The Electronic Journal of Combinatorics, 18(1). https://doi.org/10.37236/551
[5] Jungic, V. (2020) Introduction to Ramsey Theory Lecture notes for undergraduate course.
[6] J G Daugman Information Theory and Coding, Computer Science Tripos Part 2, Michaelmas Term 11 Lectures
[7] M. Schacht. (2011) Ramsey Theory Lecture Notes
[8] Mermin, D. N. (2007). Quantum Computer Science: An Introduction (1st ed.). Cambridge University Press.

APPENDIX

Please attach all your appendix at the last section of the paper. All content in the paper must be 2 columns with moderate margin including appendix.

The referencing in URECA research paper template is meant for use as a guide only. Please confirm the referencing requirements of your school with your URECA/ FYP-URECA supervising professor.