Math 116	Professor: Padraic Bartlett

Lecture 6: Latin Squares

Week 10 *UCSB 2015*

I want to end this class with a brief introduction to an object that I spent the bulk of my Ph.D studying: **Latin squares**!

1 Latin Squares

1.1 Definitions/Basics

Definition. A **Latin square** of order n is a $n \times n$ array filled with n distinct symbols such that no symbol is repeated twice in any row or column. Usually it's convenient to use the symbols $\{1, 2, \ldots n\}$, but in other situations we may use different symbol sets.

Example. Here are all of the Latin squares of order 2 on the symbols $\{1, 2\}$:

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 2 & 1 \\ \hline\end{array} \quad \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 1 & 2 \\ \hline\end{array}.$$

A quick observation we can make is the following:

Proposition. Latin squares exist for all n.

Proof. Behold!

$$\begin{array}{|c|c|c|c|c|}\hline 1 & 2 & \ldots & n-1 & n \\ \hline 2 & 3 & \ldots & n & 1 \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline n & 1 & \ldots & n-2 & n-1 \\ \hline\end{array}$$

(In general: we made the square above by placing the quantity $1 + (i + j \mod n)$ in each cell (i, j). This square is called the back circulant Latin square; it comes up a lot.) □

Given this observation, a natural question to ask might be "How many Latin squares exist of a given order n?" And indeed, this is an excellent question! So excellent, in fact, that it turns out that we have no idea what the answer to it is; indeed, we only know the exact number of Latin squares of any given order up to 11.

n	$L(n)$, the number of Latin squares of size n
1	1
2	2
3	12
4	576
5	161280
6	812851200
7	61479419904000
8	108776032459082956800
9	5524751496156892842531225600
10	9982437658213039871725064756920320000
11	776966836171770144107444346734230682311065600000
12	?

Asymptotically, the best we know (and you could show, given a lot of linear algebra tools) that

$$L(n) \sim \left(\frac{n}{e^2}\right)^{n^2}.$$

Instead of doing this, however, I'd like to focus on what I do research in: **partial Latin squares**!

Definition. A **partial latin square** of order n is a $n \times n$ array where each cell is filled with either blanks or symbols $\{1, \ldots n\}$, such that no symbol is repeated twice in any row or column.

Example. Here are a pair of partial 4×4 latin squares:

			4
2			
3	4		
4	1	2	

1			
	1		
		1	
			2

The most obvious question we can ask about partial latin squares is the following: when can we complete them into filled-in latin squares? There are clearly cases where this is possible: the first array above, for example, can be completed as illustrated below.

			4
2			
3	4		
4	1	2	

$\mapsto$

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

However, there are also clearly partial Latin squares that cannot be completed. For example, if we look at the second array

$$\begin{array}{|c|c|c|c|}\hline 1 & & & \\ \hline & 1 & & \\ \hline & & 1 & \\ \hline & & & 2 \\ \hline \end{array},$$

we can pretty quickly see that there is no way to complete this array to a Latin square: any 4×4 Latin square will have to have a 1 in its last column somewhere, yet it cannot be in any of the three available slots in that last column, because there's already a 1 in those three rows.

Deciding whether a given partial Latin square is completeable to a Latin square is, practically speaking, a useful thing to be able to do. Consider the following simplistic model of a **router**:

- Setup: suppose you have a box with n fiber-optic cables entering it and n fiber-optic cables leaving it. On any of these cables, you have at most n distinct possible wavelengths of light that can be transmitted through that cable simultaneously. As well, you have some sort of magical/electronical device that is capable of "routing" signals from incoming cables to outgoing cables: i.e. it's a list of rules of the form (r, c, s), each of which send mean "send all signals of wavelength s from incoming cable r to outgoing cable s." These rules cannot conflict: i.e. if we're sending wavelength s from incoming cable r to outgoing cable s, we cannot also send s from r to t, for some other outgoing cable t. (Similarly, we cannot have two transmits of the form $\{(r, c, s), (r, t, s)\}$ or $\{(r, c, s), (t, c, s)\}$.)

- Now, suppose that your box currently has some predefined set of rules it would like to keep preserving: i.e. it already has some set of rules $\{(r_1, c_1, s_1), \ldots\}$. We can model this as a **partial Latin square**, by simply interpreting each rule (r, c, s) as "fill entry (r, c) of our partial Latin square with symbol s."

- With this analogy made, adding more symbols to our partial Latin square is equivalent to increasing the amount of traffic being handled by our router.

This example hopefully motivates why we care about completing partial Latin squares, which lets us turn to the "what" part of our mathematical question: what kinds of partial Latin squares have completions?

Let's start simple.

Question. Suppose that we have a $n \times n$ partial Latin square L in which we've already filled in the first $n - 1$ rows. Can we always complete this partial Latin square to a Latin square: i.e. can we always fill in the last row?

Proof. The answer to this question is yes (as trying a few examples may lead you to believe.) To prove this, consider the following simple algorithm for filling in our partial Latin square:

- Look at row n.

- For each cell (n, i) in row n, look at the column i.
- There are $n-1$ distinct symbols in column i, and therefore precisely one symbol s that is **not** present in column i. Write symbol s in the cell (n, i).

We claim that this algorithm creates a Latin square. To check this, it suffices to check whether any symbols are repeated in any row or column. By construction, we know that our choice of symbol s does not cause any repetition of symbols in any column; as well, we know that no symbol is repeated in any row other than possibly the n-th row, because we started with a partial Latin square. Therefore, it suffices to check the n-th row for any repeated symbols.

To do this, proceed by contradiction: i.e. suppose not, that there are two cells in the bottom row such that we've placed the same symbols in those two cells. This means that there is some symbol s that we've **never** written in our last row. But this means that this symbol s is used somewhere in all n columns within the first $n-1$ rows, which forces some row in those first $n-1$ to contain two copies of s, a contradiction.

Therefore, our algorithm works! $\square$

(As an aside, the above tactic of "make an algorithm" is ridiculously useful, and is something we should remember and make frequent use of.)

Given our success above, it's tempting to ask whether we can extend this result to partial Latin squares where we've (say) filled in just the first $n-2$ rows, or even in general a partial Latin square where we've filled in the first k rows, for any value of k. As it turns out, this is also possible! Consider the following definition and theorem:

Definition. A $n \times n$ **Latin rectangle** is a $n \times n$ partial Latin square in which the first k rows are completely filled and the remaining $n-k$ rows are completely empty, for some value of k.

Theorem 1. *Every Latin rectangle can be completed to a Latin square.*

Before we can prove this result, we need the concept of a **system of distinct representatives**, a fairly useful concept in mathematics:

Definition. Suppose that you take m sets $A_1, \dots A_m$. A **system of distinct representatives** for these sets is a collection of distinct "representative" elements $a_1, \dots a_m$ such that $a_i \in A_i$, for each i.

Not all collections of sets have systems of distinct representatives; consider $\{1,2,3\}, \{1,2\}, \{1,3\}, \{2,3\}$. We have four sets, but clearly cannot pick out four distinct elements from them as there are only three distinct elements in total across all of our sets!

This generalizes nicely:

Observation. Suppose that $\mathcal{A} = \{A_1, \dots A_m\}$ is some collection of sets. Then if $\mathcal{A}$ has a system of distinct representatives, then for any $k \leq m$ the union of any k of the sets in $\mathcal{A}$ must contain at least k elements.

Proof. This is easy to see: for any $A_{i_1}, \dots A_{i_k}$, let $a_{i_1}, \dots a_{i_k}$ be the representatives of these sets. These are all distinct by definition. Therefore, the union of the $A_{i_1}, \dots A_{i_k}$ must contain these $a_{i_1}, \dots a_{i_k}$ and in particular contain at least k distinct elements, as claimed. □

More interesting is the converse:

Observation. (Hall's theorem.) Suppose that $\mathcal{A} = \{A_1, \dots A_m\}$ is some collection of sets with the following property: for any $k \leq m$, the union of any k of the sets in $\mathcal{A}$ contains at least k elements. (We call this Hall's property, after the mathematician that first noticed its utility.)

Then $\mathcal{A}$ has a system of distinct representatives (SDR, for short.)

Proof. We proceed via a very useful mathematical idea: to construct something useful, we should get rid of everything that's not useful!

Which is to say: take any such $\mathcal{A} = \{A_1, \dots A_m\}$. Suppose that there is some "redundant" element $x \in A_i$ such that deleting x from A_i doesn't break our property. Then delete it! If we can still succeed at finding a SDR for these smaller sets, then we can simply use it for the larger sets, and we will have proven our claim.

Repeatedly delete "redundant" elements x as described above, until there are no more such elements left. This leaves us with a collection $\mathcal{A}$ with the following two properties:

1. (Hall's property.) For any $k \leq m$, the union of any k of the sets in $\mathcal{A}$ contains at least k elements.

2. If you delete any element from any A_i, the above property becomes false.

Here's a thing that would make finding a SDR very easy: if every A_i was a singleton set $\{a_i\}$. If this held, then we would have that all of our a_i-elements are distinct. (Otherwise, we could take the union of two non-distinct $\{a_i\}$ sets, which would give us a union of two things with size 1; this would break the Hall property and yield a contradiction.) Consequently, we can just use the individual a_i contents of each set as a SDR!

Here's a thing that is true: every A_i set is a singleton set $\{a_i\}$! We can prove this by contradiction: assume that some A_k contains two different elements. Call them x and y.

If you delete either we must violate the Hall property, by assumption. Therefore there must be some sets I_x, I_y of indices, neither containing k, such that

$$\left| A_k \cup \bigcup_{i \in I_x} A_i \setminus \{x\} \right| \leq |I_x|,$$

$$\left| A_k \cup \bigcup_{i \in I_y} A_i \setminus \{y\} \right| \leq |I_y|.$$

This is because if deleting one of x, y causes a problem, then it must cause a problem with the set A_k we deleted x, y from; so we can assume that l must be one of the indices chosen to break the Hall property. A contradiction to Hall's property would mean that the size of that set of indices including k is strictly greater than the size of the union of our sets; therefore, the size of that set of indices without k is greater than or equal to the size of this union, which is precisely what we wrote above.

Let $\bigcup_{i\in I_x} A_i \setminus \{x\} = X$ and $\bigcup_{i\in I_y} A_i \setminus \{y\} = Y$ for short. Consider the two sets $X \cup Y$, $X \cap Y$. On one hand, we can see that

$$X \cup Y = A_k \cup \bigcup_{i\in I_x \cup I_y} A_i,$$

because $x \in A_k \setminus \{y\}$ and $y \in A_k \setminus \{x\}$, so when we union X and Y we don't have to worry about the set-subtraction.

Similarly, when we form $X \cap Y$, we get that

$$X \cap Y \supseteq \bigcup_{i\in I_x \cap I_y} A_i,$$

by just ignoring the A_k set and only considering everything else they have in common.

But $X \cup Y, X \cap Y$ are both unions of sets in $\mathcal{A}$; so Hall applies, and we get that $|X \cup Y| \geq |I_x \cup I_y| + 1, |X \cap Y| \geq |I_x \cap I_y|$.

From our earlier work, we know that

$$|I_x| + |I_y| \geq |X| + |Y|.$$

However, we know that for any two sets X, Y, we have $|X| + |Y| = |X \cup Y| + |X \cap Y|$; so we actually have

$$\begin{aligned} |I_x| + |I_y| &\geq |X \cup Y| + |X \cap Y| \\ &\geq |I_x \cup I_y| + |I_x \cap I_y| + 1 \\ &= |I_x| + |I_y| + 1, \end{aligned}$$

by applying the same set-size trick (i.e. $|I_x| + |I_y| = |I_x \cup I_y| + |I_x \cap I_y|$) on the second-to-last line.

This is a contradiction! Therefore each set A_i must have one element; so we have a SDR, as claimed. $\square$

This result applies nicely to Latin rectangles. To see this: take any $n \times n$ partial Latin square P that's a Latin rectangle with k filled rows. For each column j, associate a set A_j consisting of all of the symbols s that are not used in column j in these k filled rows.

This gives us a collection $\mathcal{A} = \{A_1, \ldots A_n\}$ of sets. Notice that any way to fill in the $k+1$-th row of P is precisely any way to pick out a SDR from $\mathcal{A}$:

- The Latin property for each column means that for each j, the entry $(k+1, j)$ needs to have not been used in earlier rows in that column; i.e. that the symbol for $(k+1, j)$ comes from A_j.
- The Latin property for the $k+1$-th row just asks that all of these chosen symbols are distinct; i.e. that when we pick our representatives, they're all distinct!

So, if we want to fill in the $k+1$-th row, we just need to find a system of distinct representatives, or equivalently (by the theorem above) establish that the collection $\mathcal{A}$ satisfies Hall's property.

This is not too hard to do. Take any m of our sets $A_{j_1}, \ldots A_{j_m}$. If we count repeated elements multiple times, we have $m(n-k)$ elements in each set.

However, because each symbol must occur once in each filled row of anything with the Latin property, every symbol in our Latin rectangle P must be used exactly k times, once for each row. Therefore, for any symbol s there must be k columns in which s occurs, and therefore $n-k$ in which it does not (and thus $n-k$ values of k for which $s \in A_j$.)

So, in particular, no element in $A_{j_1}, \ldots A_{j_m}$ can be repeated more than $n-k$ many times, as it occurs in a maximum of $n-k$ of these sets! So, if we're in the worst-case scenario where every element of $A_{j_1}, \ldots A_{j_m}$ is repeated the maximum $(n-k)$ number of times, we have that the total number of **distinct** cells is $\frac{m(n-k)}{n-k} = m$, i.e. the number of sets! So we have proven that for any $A_{j_1}, \ldots A_{j_m}$,

$$|A_{j_1} \cup \ldots \cup A_{j_m}| \geq m,$$

and therefore that we satisfy Hall's property.

This gives us our claimed result on Latin rectangles for free:

Corollary. Any Latin rectangle can be completed to a Latin square.

Proof. Take any $n \times n$ partial Latin square P that is a Latin rectangle on k filled rows. Use the result above to fill in rows of P one-by-one, until you get a filled Latin square! □

Determining what kinds of partial Latin squares have completions is an open question in many situations! Here are some known results:

- (Hall, 1949:) A **Latin rectangle** is a partial Latin square P where the first k rows of P are filled and the rest are blank. All Latin rectangles can be completed.
- (Smetaniuk, 1981:) If P is a partial latin square with $\leq n-1$ filled cells, P can be completed.
- (Buchanan, 2007:) If P is a $n \times n$ partial Latin square where precisely 2 rows and columns of P are filled, P can be completed.

A question I spent part of my dissertation studying is the following: Call a $n \times n$ partial Latin square **ϵ-sparse** if at most ϵn of the entries in any row, column, or symbol are not blank. It is conjectured that all 1/4-sparse partial Latin squares are completable; in my dissertation I proved that all 10^{-4}-sparse partial Latin squares are completable.

Somewhat frustratingly, 10^{-4} is not 1/4. If you can improve this, I'd love to see it!