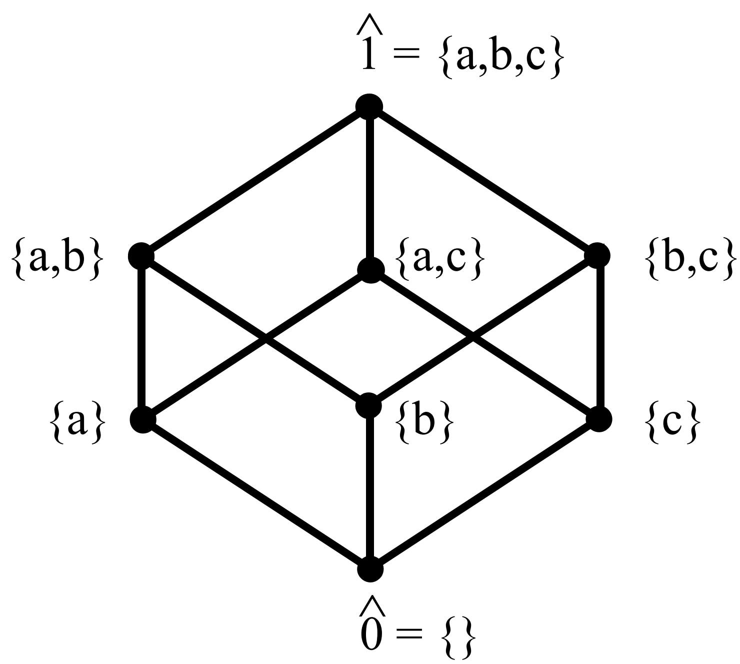
Combinatorial Theory



John N. Guidi

Lecture Notes – Fall 1998
MIT Course 18.315
Professor Gian-Carlo Rota

This publication is dedicated to John N. Guidi (1954-2012) whose remarkable almost "verbatim" notes in Prof. Gian-Carlo Rota's courses in 1998 at MIT faithfully reproduce both the content and erudition of Rota's famous lectures just before Rota's premature death in 1999.

John N. Guidi
Research Affiliate
Department of Mathematics
Room 2-363A
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
USA
Email: guidi@math.mit.edu

Devil: 74

These *Lecture Notes* originated from the lectures presented by Gian-Carlo Rota, Professor of Applied Mathematics, for graduate course 18.315 - Combinatorial Theory, which he taught at MIT, during the Fall 1998 semester. Topics covered included sets, relations, enumeration, order, matching, matroids, and geometric probability. These *Lecture Notes* were produced from notes I made during class, audio recordings I made of lectures, as well as clarifications and expansions I made of the material presented, after the fact. These *Lecture Notes* were not reviewed by Professor Rota and should not be considered endorsed by him.

I had an ulterior motive for writing these up. I found this a particularly useful way to profoundly understand the material (or, as Professor Rota was fond of saying, "*to really rub it in*"). My goal was not to provide verbatim transcriptions of the lectures, but rather to provide a set of comprehensive notes, including some of the oral commentary, of the material presented in class. I hope to have captured a bit of the spirit of these lectures and to have introduced only a limited number of errors.

I wish to thank a number of people. Richard Stanley, who is the Norman Levinson Professor of Applied Mathematics, is my host at MIT. I am deeply grateful for his interest, efforts on my behalf, and encouragement. Daniel Kleitman, who is the Chairman of Applied Mathematics Committee at MIT, has been most supportive. Jeff Lieberman, who was a student at MIT in this course, graciously provided me with a copy of his notes. His notes were often helpful when I struggled to understand a point and my own were unclear.

Gian-Carlo Rota died around April 19, 1999 (an obituary and other materials about his life and career have been made available by Richard Stanley at http://www-math.mit.edu/~rstan/rota.html). I am grateful to Professor Rota for enthusiastically sharing his wealth of knowledge about combinatorics and mathematics, in general. His many lessons, regarding both content and manner, on education, scholarship, and research were enlightening and enduring. His kindness and generosity are appreciatively acknowledged. He was a superb teacher, in the truest sense of the word. He is sorely missed.

Professor Rota was keen for me to complete these *Lecture Notes*, as he also felt that others might find them useful. I regret that he never saw this volume. I like to think he would have been pleased.

John N. Guidi
March 20, 2002

Note: within the body of the text, pagination is of the form [*lecture.page*]. For example, page [3.5] refers to the fifth page of the third lecture, which was given on September 14, 1998.

Contents

Lecture 3 - [9/14/98]

Lecture 4 - [9/16/98]

Lecture 5 - [9/18/98]

Lecture 6 - [9/21/98]

Lecture 7 - [9/23/98]

Lecture 8 - [9/25/98]

Lecture 9 - [9/28/98]

Chapter II - Order

Lecture 10 - [9/30/98]

Lecture 11 - [10/2/98]

Lecture 12 - [10/5/98]

Lecture 13 - [10/7/98]

Lecture 14 - [10/9/98]

Lecture 15 - [10/13/98]

Lecture 26 - [11/9/98]

Lecture 27 - [11/13/98]

Lecture 28 - [11/16/98]

Lecture 29 - [11/18/98]

Note: within the body of the text, pagination is of the form [*lecture.page*]. For example, page [3.5] refers to the fifth page of the third lecture, which was given on September 14, 1998.

Exercises

Stars are used to rank the exercises in the following manner:

unstarred	Ordinary exercise, as you might expect in an introductory course.
*	Difficult exercise that requires some serious thinking.
**	If worked out, the exercise might make a publishable paper.
***	Possible topic for a Ph.D. thesis.

John Guidi
guidi@math.mit.edu

Chapter One: Sets and Relations

We want to review in detail the Boolean algebra of sets.

S = set

We denote by $P(S)$ the family of all subsets of S
Such a family is often called The Boolean algebra of subsets.
because, as you will see, there are other Boolean algebras.

Most of you are familiar w/ the elementary operations on sets, but we have to review them carefully, because we will use them in an unusual way.

Operations on sets:

union: $A \cup B$

intersection: $A \cap B$

complement: A^c

where $A, B \subseteq S$ ← the Universal set

I don't need to explain what these mean.
I assume you are familiar with these operations.

You are also familiar w/ some of the results of these operations.
In particular:

$$\phi = \text{empty set}$$

$$\phi^c = S = \hat{1} \longleftarrow \text{"one hat"}$$

The complement of the emptyset is the Universal Set.
And for reasons that we will see later, is sometimes written as $\hat{1}$.

Let's define another operation:

Sheffer stroke:

$$A/B = A^c \cap B^c$$

This was discovered in 1913 by Prof. Sheffer. The Sheffer Stroke has a very peculiar property:

It can be used to define every other operation among sets.

The Sheffer Stroke suffices to define all Boolean operations on sets, to wit:

complement: $A|A = \underbrace{A^c \cap A^c}_{\text{by definition}} = A^c$

So the complement of a set is defined as "A, Sheffer Stroke, itself"

intersection: $(A|A)|(B|B) = A^c | B^c$ ← we've just seen that: $A|A = A^c$

$= A^{cc} \cap B^{cc}$

$= A \cap B$

union: $(A|B)|(A|B) = (A^c \cap B^c)|(A^c \cap B^c)$

From above, we have that this is:

$= (A^c \cap B^c)^c$

$= A \cup B$ (from deMorgan's Laws)

Finally, even the <u>null set</u> can be defined using the Sheffer Stroke:

$$A|(A|A) = A|A^c$$
$$= A^c \cap A^{cc}$$
$$= A^c \cap A$$
$$= \phi$$

<u>Excercise 1.1:</u>
Now you've heard what I just said and I know what you are thinking: "Gee, maybe there are many other operations like that."

The <u>only operations (binary)</u> among sets by which <u>all boolean operations</u> may be defined are:

the Sheffer Stroke $A|B = A^c \cap B^c$ <u>and</u> $A \downarrow B = A^c \cup B^c$

These are the only 2 binary operations for which all Boolean operations among sets are defined.

Prove this. This was a research paper published in 1913,

- Why did we say binary here?

Because the operations of union and intersection are operations that involve 2 variables. Hence the word binary.

Boolean algebra is defined by: 2 binary operations $(\cup, \cap)$; 1 unary operation $({}^c)$ } on subsets

Note: The null set $\emptyset$ can be viewed as an operation of picking a special element. In this sense, it is a zero-ary operation.

{ Boolean algebra is an algebraic system w/ 2 binary operations, 1 unary operation, 1 zero-ary operation }

This idea can be generalized, and we probably will.

Exercise 1.1 shows that the only 2 binary operations that give equivalent algebraic systems are the Sheffer Stroke and $\downarrow$:

Sheffer Stroke

$$\text{Boolean Algebra} = \begin{array}{l}\text{Binary ops:}\\ \text{Unary ops}\\ \text{Zero-ary ops:}\end{array} \begin{Bmatrix} \cup, \cap \\ c \\ \emptyset \end{Bmatrix} = \underbrace{\begin{Bmatrix} \mid \\ c \\ \emptyset \end{Bmatrix} = \begin{Bmatrix} \downarrow \\ c \\ \emptyset \end{Bmatrix}}$$

and further, these are the only two single binary operations $(\mid, \downarrow)$ that give you Boolean Algebra.

Now, I know what you are thinking.
What about ternary operations? There's an enormous literature defining ternary operations that give Boolean algebra.
Let's see one. The second most famous one.

- Conditioned Disjunction: $A, B, C \subseteq S$

$$\cancel{[A,B,C] = (A \cup B) \cap (A \cup C) \cap (B \cup C)}$$

this is actually the median. See [7.1 - 7.3]

$$[A,B,C] = (B^c \cap A) \cup (B \cap C)$$

Exercise 1.2 :

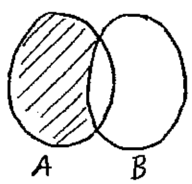
Conditional disjunction may be used to define all Boolean operations.
This is a very interesting remark. Prove it.

Are there any other operations on sets worth talking about?
Yes.
Another famous operation (perhaps the most famous one).

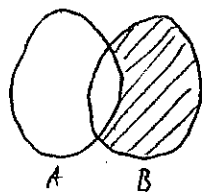
Symmetric Difference of sets :

$$A + B = (A \cap B^c) \cup (A^c \cap B)$$

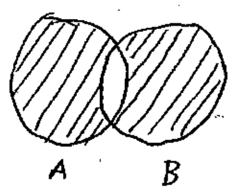
We can visualize this as follows:



$(A \cap B^c)$



$(A^c \cap B)$



$(A \cap B^c) \cup (A^c \cap B)$

The symmetric difference (+) are the elements that belong to either one of A or B, but not both.

This was discovered fairly late in the game, by an American mathimatican Marshall Harvey Stone.

Why is this operation important?
There is an extremely important property, which I want to emphasize.

Properties of Symmetric Difference :

Commutative : $A + B = B + A$

Associative : $A + (B + C) = (A + B) + C$ ← {work this out. It's not so obvious}

So, it behaves like addition. But, not completely:

$$A + A = (A \cap A^c) \cup (A^c \cap A)$$
$$= \phi \cup \phi$$
$$A + A = \phi \longleftarrow \left\{ \begin{array}{l} \text{so symmetric difference does } \underline{\text{not}} \\ \text{behave just like addition} \end{array} \right\}$$

implies some advanced mathematics

• Remark:

The family of subsets $P(S)$ with $+$ (symmetric difference) and $\cdot$ (intersection) gives you a <u>commutative ring</u>, where every element is <u>idempotent</u>

$A \cdot A = A$

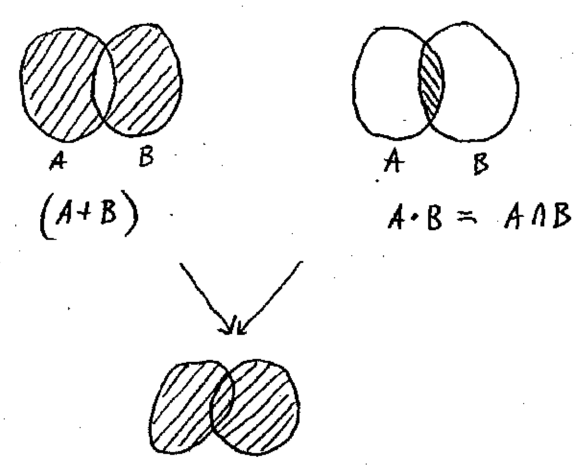
Furthermore, from $+$ (symmetric difference) and $\cdot$ (intersection), you can derive the Boolean operations.

definition of $\cdot$

intersection : $A \cap B = A \cdot B$

union : $A \cup B = A + B + A \cdot B$
$$= (A + B) + A \cdot B$$
$$= (A + B) + (A \cap B)$$

Proof by Picture:



$A \cup B = A + B + A \cdot B$

complement: $A^c = \mathbb{1} + A$
$$= (\mathbb{1} \cap A^c) \cup (\mathbb{1}^c \cap A)$$
$$= A^c$$

- <u>Boolean function :</u>

  Anything you can obtain by iterated applications of the Boolean operations.

  Ex: $\varphi(A_1, A_2, A_3) = ((A_1 \cup A_2) \cap A_3^c) \cup (A_1 \cap A_3)$

  {These kinds of functions are very common, for example, in switching theory.}

  We can simplify, using the <u>distributive law</u>, to get:

  $$= (A_1 \cap A_3^c) \cup (A_2 \cap A_3^c) \cup (A_1 \cap A_3)$$

  In a similar way, <u>any</u> Boolean function can be simplified as a union of intersections of sets and of complements, by using the distributive law.
  This standard form is known as:

  Disjunctive Normal Form

- <u>Disjunctive Normal Form</u> of a Boolean function $\varphi(A_1, A_2, \dots, A_n)$ is the <u>irredundant</u> union of expressions of the form:

  $$A_1^{\pm} \cap A_2^{\pm} \cap \dots \cap A_n^{\pm}$$

  each A_i appears once and only once

  $$A_i^+ = A_i$$
  $$A_i^- = A_i^c$$

Note: Assume you have simplified to a union of intersections.
If some of the intersections consist of less than n terms, each intersection can be placed in standard form, as per the following example:

* = don't care whether term

no A_1 or complement

$$
\begin{aligned}
A_2^* \cap \dots \cap A_n^* &= (A_2^* \cap \dots \cap A_n^*) \cap \mathbb{1} \\
&= (A_2^* \cap \dots \cap A_n^*) \cap (A_1^+ \cup A_1^-) \\
&= \underbrace{(A_1^+ \cap A_2^* \cap \dots \cap A_n^*) \cup (A_1^- \cap A_2^* \cap \dots \cap A_n^*)}_{\text{Standard form}}
\end{aligned}
$$

Example:

Disjunctive Normal Form of the Boolean function:

$$\varphi(A_1, A_2, A_3) = \left((A_1 \cup A_2) \cap A_3^c\right) \cup (A_1 \cap A_3)$$

$$= (A_1 \cap A_3^c) \cup (A_2 \cap A_3^c) \cup (A_1 \cap A_3)$$

$$= (A_1 \cap A_2 \cap A_3^c) \cup (A_1 \cap A_2^c \cap A_3^c) \cup$$

same

$$(A_1 \cap A_2 \cap A_3^c) \cup (A_1^c \cap A_2 \cap A_3^c) \cup$$

$$(A_1 \cap A_2 \cap A_3) \cup (A_1 \cap A_2^c \cap A_3)$$

$$= (A_1 \cap A_2 \cap A_3) \cup (A_1 \cap A_2 \cap A_3^c) \cup$$

$$(A_1 \cap A_2^c \cap A_3) \cup (A_1 \cap A_2^c \cap A_3^c) \cup$$

$$(A_1^c \cap A_2 \cap A_3^c)$$

- Exercise 1.3

  Show that every Boolean function can be expressed in Disjunctive Normal Form
  (kind of easy)

- Historical Digression

  When Boolean algebra was being developed in the first half of the century, people often did things like this.
  One of the most remarkable feats that was performed was an achievment of the mathematician E. L. Post.
  Let me tell you informally what he did.

  Take a finite number of Boolean functions.

  $$\varphi_1(A_1, \ldots, A_k), \ \varphi_2(A_1, \ldots, A_k), \ \ldots, \ \varphi_n(A_1, \ldots, A_k)$$

  Then you allow functional composition of these Boolean functions, in arbitrary ways.
  When is it true that, by taking functional compositions of these Boolean functions, you can express any Boolean function, whatsoever.

To that end, when is it true that I take functional compositions of Boolean functions and get union, intersection, and complement.

Post computed all possible sets of Boolean functions and there are 86 of them.
Since he did this, his work can be greatly simplified.
But at his time, this was a great achievement.

For examples: (1) the Sheffer Stroke is a binary Boolean function that generates all Boolean functions.

(2) the $\downarrow$ is a binary Boolean function that generates all Boolean functions.

(3) The Conditional Disjunction is a ternary Boolean function that generates all Boolean functions

$\vdots$

(86) ...

Post worked this all out in 200 pages.

- Infinite Distributive Laws for Sets ← (this is something you probably haven't encountered.)

The distributive laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

We then generalize this to an infinite family of sets A_i:

$$A \cap \left(\bigcup_{i \in I} A_i \right) = \bigcup_{i \in I} (A \cap A_i)$$ } this is true for any family of sets A_i, infinite, or not.

- Exercise 1.4:
Now, let's jazz it up.
Suppose we have a doubly infinite family of sets.
And we have:

$$\bigcup_{i \in I} \left(\bigcap_{j \in J} A_{ij} \right) \quad , \text{ where } A_{ij} \subseteq \underset{\text{probably infinite}}{S}$$

How can we interchange union and intersection in this equation?

I'll give you the formula first and then we'll discover it.

Let's denote by:

$$J^I = \text{the set of all functions from } I \text{ to } J$$

then:

$$\bigcup_{i \in I} \left(\bigcap_{j \in J} A_{ij} \right) = \bigcap_{\varphi \in J^I} \left(\bigcup_{i \in I} A_{i,\varphi(i)} \right)$$

φ ranges over all functions from I to J

This is not so easy to see at first.
Let's look at an example:

$$\bigcup_{i \in \{1,2\}} \left(\bigcap_{j \in \{1,2,3\}} A_{ij} \right) = (A_{11} \cap A_{12} \cap A_{13}) \cup (A_{21} \cap A_{22} \cap A_{23})$$

This is already in disjunctive normal form.
We rewrite this w/ our goal of exchanging union and intersection.

$$= \big((A_{11} \cap A_{12} \cap A_{13}) \cup A_{21}\big) \cap$$
$$\big((A_{11} \cap A_{12} \cap A_{13}) \cup A_{22}\big) \cap$$
$$\big((A_{11} \cap A_{12} \cap A_{13}) \cup A_{23}\big)$$

I	$\to$	J
1		1
2	$\to$	2
		3

$\varphi(1) = \{1,2,3\}$
$\varphi(2) = \{1,2,3\}$

$$= (A_{11} \cup A_{21}) \cap (A_{12} \cup A_{21}) \cap (A_{13} \cup A_{21}) \cap$$
$$(A_{11} \cup A_{22}) \cap (A_{12} \cup A_{22}) \cap (A_{13} \cup A_{22}) \cap$$
$$(A_{11} \cup A_{23}) \cap (A_{12} \cup A_{23}) \cap (A_{13} \cup A_{23})$$

$$= \bigcap_{\varphi \in J^I} \left(\bigcup_{i \in I} A_{i,\varphi(i)} \right)$$

John Guidi
guidi@math.mit.edu

The Theory of Relations (beginning)

Last time, we studied some of the classical properties of the algebra of sets.
Boolean algebras of all subsets of a set (finite or infinite)

{ We will later see that the algebraic properties of union, intersection, complement actually can be used to abstractly characterize the Boolean algebra. }

This is a tremendous discovery in mathematics

We begin w/ one of the most important notions of combinatorics.
A notion that is given about 16 different names.
I have chosen the term relation, because it is the oldest – going back to Aristotle.
This is one of the oldest concepts of mathematics. Even older than the concept of function.

- What is a relation?

S = a set

T = another set

A relation R is a subset of $S \times T$

relation $R \subseteq S \times T$

Now you say "What's the big deal?"
The big deal is this.

$a \in S$, $b \in T$, $(a,b) \in R$ ← { $a \in S$ and $b \in T$ and the pair belongs to R }

↑ ordered pair

We also write:

$a \in S$, $b \in T$, $a \, R \, b$ ← { a is related to b by relation R }

- Bipartite Graph of R

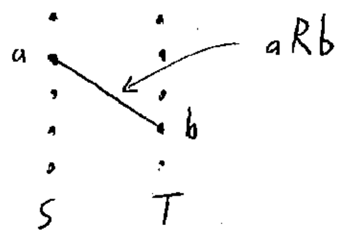
Corresponding to a relation is a bipartite graph.
We have, here, the situation where we are describing concepts that are mathematically identical, yet psychologically different.

The bipartite graph is, strictly speaking, the same as a relation.
But you visualize the bipartite graph differently.

↑ that's the difference

You draw a set of vertices, corresponding to the set S.
You draw a set of vertices, corresponding to the set T.
If an element $a \in S$ and an element $b \in T$ belong to the relation $\underbrace{(a,b) \in R}_{\text{or, } aRb}$
then we draw an edge connecting a and b.

In this way you <u>visualize</u> the relation as a bipartite graph.



So, the <u>theory of bipartite graphs</u> is the <u>same</u> as the <u>theory of relations</u>.

- Another name used for relation, especially by geometers, is:
  "correspondence"

What are examples of relations? A Mickey Mouse example is worthwhile to consider.

- <u>Example 1</u>

T = a set
$S \subseteq P(T)$
↖ S is some family of subsets of T

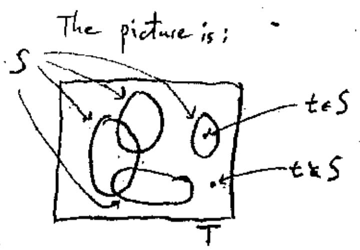
<u>Any</u> $\underbrace{\text{family of subsets of } T}_{S}$ defines a relation R as follows:

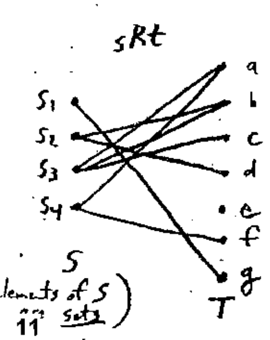
$s \in S, \; t \in T$

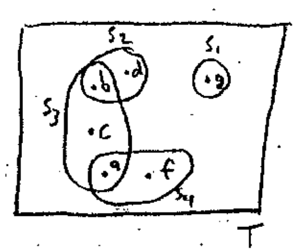
$(s,t) \in R \iff t \in S$

↑ s is related to t by R whenever t belongs in S

The picture is:

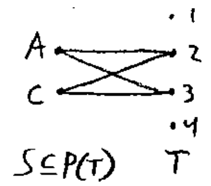






- Can every relation be represented in this way?
  Answer – <u>No</u>

Assume we have subsets A and C that are related to the <u>same</u> elements



A is related to a set of elements of T $(\{2,3\})$
C is related to a set of elements of T $(\{2,3\})$

exactly the same

Thus $A + C$ have to be the <u>same subset</u>

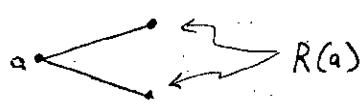
$A = C$

- This dispute goes back 2000 years.
  Allow me this philosophical digression.

  Not every relation can be represented as in example 1.
  Because in a relation, 2 elements ($S \subseteq P(T)$) may be related to the same things.
  In which case you are forced to call these 2 elements the <u>same set</u>.

  Are we in the presence of a generalization of the notion of set?
  Not quite.
  Let's see what happens.

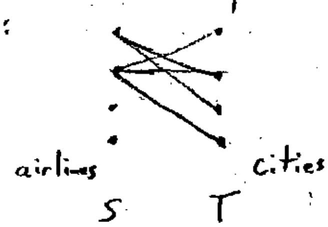
  For relation $R \subseteq S \times T$,
  set $R(a) = \{b \in T : (a,b) \in R\}$



A relation may be represented as a family of subsets of T $\iff$ $R(a) = R(c)$ for $a, c \in S \overset{\text{implies}}{\Longrightarrow} a = c$

This is what we just said at the top of the page, but in more formal language.

- Relations arise in the most disparate circumstances.
  Recently, computer scientists got hold of the theory of relations.
  Why? Because relations express the most primitive notions we can think of.
  Example:



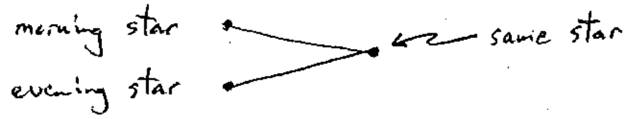
Other examples: People $\times$ Jobs people can do
Boys $\times$ Girls
etc.

Relations are a <u>universal concept</u>.

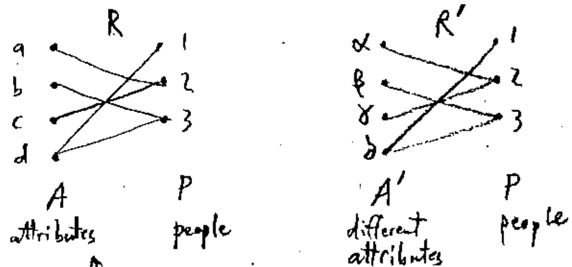
- Where does Aristotle come in?
  (This is more philosophy than mathematics)

Aristotle comes in as a <u>philosophical dispute</u> over defining a set by the <u>extent</u> -vs- the <u>intent</u>

classical example from Frege



Another example:
You can take a number of attributes that determine some set of people.
Then you can take a number of <u>completely different</u> attributes that determines <u>exactly</u> the <u>same set of people.</u>



Even though <u>as sets</u>, these Attributes are <u>equal</u> $\{a,b,c,d\} = \{\alpha,\beta,\gamma,\delta\}$, their <u>properties</u> may be different.
Therefore, you <u>supplement</u> the <u>concept</u> of a <u>set</u> by the concept of a <u>relation</u>.

$R \neq R'$

The concept of relation has the advantage that one can define:

- <u>Inverse relation</u>

$$R^{-1} \subseteq T \times S \ : \ R^{-1} = \{(b,a) : (a,b) \in R\}$$

R^{-1} is a relation between T and S

R^* notation is preferred to R^{-1} for the inverse relation. See [23.6] for discussion.

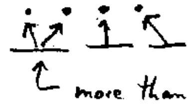
- <u>Function</u>

You can view a <u>function</u> <u>as</u> a <u>special</u> kind of <u>relation</u>

A <u>function</u> is a relation R s.t. if $c, d \in R(a)$ then $c = d$ and
$R \subseteq S \times T$ for every $a \in S$, $R(a) \neq \phi$

In other words: for every vertex of S, there issues <u>exactly</u> <u>one</u> <u>edge</u>. } balls (S) into boxes (T)

- The inverse of a function is not a function.
  It is a relation

function f:  f^{-1}: 

more than 1 edge issues
Not a function
It is a relation.

The next concept we meet:

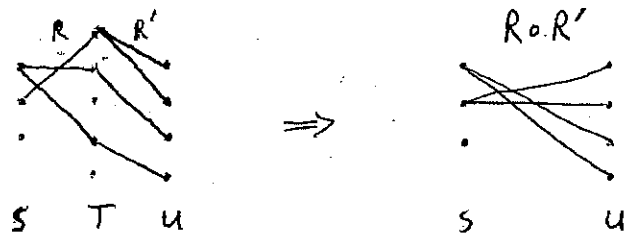
- Composition of relations

Given relation $R \subseteq S \times T$ and another relation $R' \subseteq T \times U$

Then we define:

$$R \circ R' = \{(a,c) : \text{there is a } b \in T \text{ s.t. } (a,b) \in R \text{ and } (b,c) \in R'\}$$

We can visualize the composition of relations as:



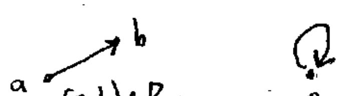
An important class of relations consists of the relation of a set with itself.

- Special Case:

$R \subseteq S \times S$ ← {aka "relation R on a set S," contrasted with $R \subseteq S \times T$ aka "relation R from set S to set T"}

In this special case, we can represent the relation not only as a bipartite graph, but as an oriented graph.

Namely, you visualize the relation as follows:



$(a,b) \in R$ {an oriented arrow from a to b}

$(c,c) \in R$ {loop}

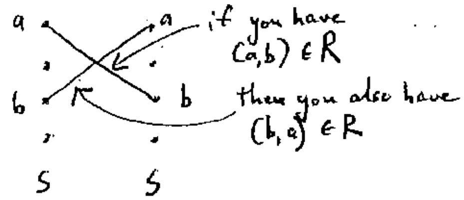
The Theory of Oriented Graphs is the same as the Theory of Relations of Sets w/ themselves. It's just a matter of wording.

Some people like to talk about oriented graphs – good.
Some people like to talk about relations of sets w/ themselves – good. } They are the same

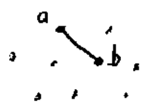
- When $R \subseteq S \times S$, the relation is <u>symmetric</u> when:

$$R = R^{-1}$$

In terms of bipartite graphs:



- A <u>symmetric relation</u> $R \subseteq S \times S$ has a <u>simpler graphical representation</u>.



$(a,b) \in R$

S {an <u>unoriented</u> edge between a and b.}

it doesn't matter in which orientation you are, because the relation is symmetric.

The <u>Theory of Unoriented Graphs</u> is the <u>Theory of Symmetric relations</u> of sets with themselves.

They are one and the same.
I don't like to talk about graphs, I like to talk about relations.
So you are stuck w/ it.

- Furthermore, if the relation ($R \subseteq S \times S$) has the property that it is <u>anti-reflexive</u>:

$$(a,a) \notin R \quad \text{for any } a \in S.$$

← no loops

The graph is called a <u>linear graph</u>.

- Associated w/ a relation, we have the <u>incidence matrix of a relation</u>.

Given $R \subseteq S \times T$, $|S| < \infty$, $|T| < \infty$ ← cardinality of set $< \infty$ = finite set

then the <u>incidence matrix</u> of R is a matrix of 0 and 1

$$\begin{array}{cc} & T \\ S & \begin{bmatrix} & & \\ & x_{ab} & \\ & & \end{bmatrix} \end{array}$$

if $(a,b) \in R$, set $x_{ab} = 1$
$(a,b) \notin R$, set $x_{ab} = 0$

There is a whole school that talks about nothing except matrices of 0's and 1's.

The <u>theory of matrices of 0's and 1's</u> is <u>cryptomorphic</u> to the <u>theory of relations</u>.

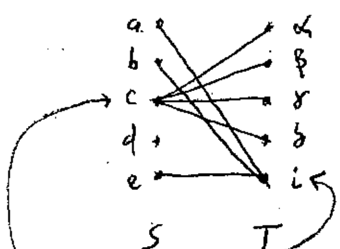
- If you look at a matrix from the point of view of its incident matrix, then it becomes natural to associate with the matrix its <u>marginals.</u>

<u>Marginals</u> of R

└ the term originated in statistics

You take the sum of all the 1's in a row and write it to the left.
" " " " " " " " " " column " " " at the top.
So, it's the <u>row sums</u> and the <u>column sums</u> of the <u>incidence matrix.</u>

From the point of view of graph theory, how do we view the marginals?



$$
\begin{array}{cc|ccccc|}
 & & 1 & 1 & 1 & 1 & 3 \\
 & & \alpha & \beta & \gamma & \delta & i \\
\hline
1 & a & & & & & 1 \\
1 & b & & & & & 1 \\
4 & c & 1 & 1 & 1 & 1 & \\
0 & d & & & & & \\
1 & e & & & & & 1 \\
\hline
\end{array}
$$

column sums

incidence matrix $R \subseteq S \times T$

row sums

A <u>row sum</u> is the number of edges emanating from a given vertex of S.

row_sum (c) = 4 ← degree of vertex c

A <u>column sum</u> is the number of edges incident on a given vertex of T.

column_sum (i) = 3

$$\sum_{s \in S} \text{row\_sum}(s) = \sum_{t \in T} \text{column\_sum}(t)$$

- Next Monday, we will answer the following interesting question:

Given 2 sets of Integers, when is there a relation that has those sets of Integers as marginals?

This is a very important question, which has a very elegant answer.

*** Exercise 2.1 :

Why don't we define ternary relations?
What we've just defined is a binary relation.

We define a ternary relation as:

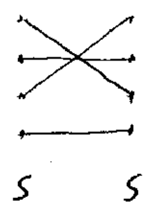
$$R \subseteq S \times T \times U$$

It's a very nice definition.
No one has ever been able to find a non-trivial property of ternary relations.

The situation is even worse than that.
Let me define a special kind of relation.

$R \subseteq S \times S$ is a permutation iff all its marginals equal 1

What does that mean?
It means that every element maps into a unique element of S.



You are permuting the elements.

Part 1: Find a non-trivial property of ternary relations.

Part 2: Find the right (3D) ternary analogue of a permutation.

What do we mean by right?
Well, when you've found it, you will know.

I've known some very good mathematicians who've worked on this for 10 years w/o success.
Mark Mayer, for example, worked very hard.
We will see later, when we do the Birkhoff - von Neumann Theorem that this pattern comes up.

Continuing w/ our laundry list of definitions:

- If we take a symmetric relation, a reflexive relation can be defined by introducing 2 special kinds of relations:

  - Identity Relation

    $I \subseteq S \times S \qquad I = \{(a,a) : a \in S\}$

    $$\begin{array}{cc} & S \\ S & \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \end{array}$$

    incidence matrix of I

  - Universal Relation

    $U_S \subseteq S \times S \qquad U_S = \{(a,b) : a, b \in S\}$

    $$\begin{array}{cc} & S \\ S & \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \\ 1 & \cdots & 1 \end{bmatrix} \end{array}$$

    incidence matrix of U_S

- Algebra of relations

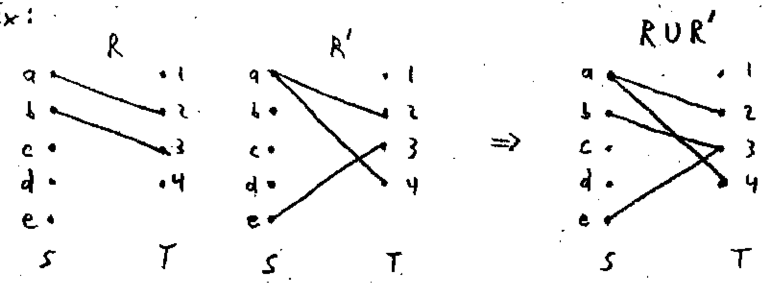
  There is such a thing as an algebra of relations.
  We define this in terms of relations of a set into itself, even though some of these operations can be defined for relations of a set into another set.

if $R, R' \subseteq S \times T$

then you can define: • $R \cup R'$

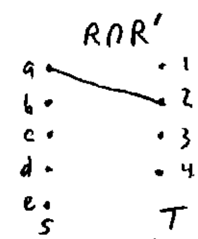
union of the elements of the relations.
You take the edges and join them together. If you have a double edge, you reduce it to a single edge.

Ex:

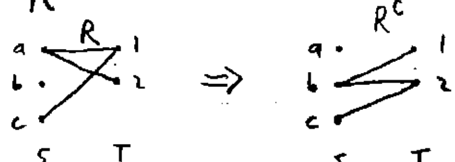


• $R \cap R'$

using $R + R'$ from above $\Rightarrow$



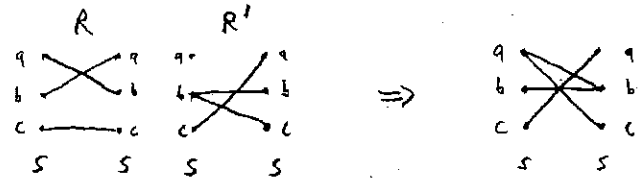
• R^c



- Further,
if $R, R' \subseteq S \times S$
then we have the additional operation:

Composition [p2.5]

$R \circ R'$



The algebra of relations with a set into itself has all the Boolean operations and compositions.

Mathematicans, starting in 1870 and through to 1993, tried to study all the identities that hold with the Boolean operations and composition.
And they thought:
Just as we can characterize the algebra of sets by the Boolean operations [p1.7-8],
Perhaps we can characterize the algebra of relations by the Boolean operations and composition.

$$(\cup, \cap, {}^c, \circ) \overset{?}{\Rightarrow} \text{all identities of relations}$$

{ this effort failed.
It was proved that it is impossible to characterize the algebra of relations w/ the Boolean operations and composition. }

Now, continuing w/ our definitions:

- $R \subseteq S \times S$

R is reflexive if $R \supseteq I$ ← R is contained in the Identity relation, which means for every $a \in S$, the pair $(a,a) \in R$

symmetric if $R = R^{-1}$

transitive if $R \circ R \subseteq R$ ← i.e., if $(a,b) \in R$ and $(b,c) \in R$ then $(a,c) \in R$,
That's what $R \circ R \subseteq R$ says, in a concise and efficient way.

- Exercise 2.2 There was a research paper, some time ago, from UNC that studied this.

Study properties of relations satisfying $R \circ R \circ R \subseteq R$

There are some remarkable properties.

- A relation R on a set S that is reflexive, symmetric, and transitive is an equivalence relation.

$$R \subseteq S \times S \quad \text{where} \quad R \supseteq I,$$
$$R = R^{-1},$$
$$R \circ R \subseteq R$$

R is an equivalence relation.

- Equivalence class of an equivalence relation

equivalence classes = maximal subsets B of S s.t. for $a, c \in B$, we have aRc

- An equivalence class is always non-empty
- Any two equivalence classes are disjoint:

$$\bigcup B = S$$

{ the union of all equivalence classes (i.e., maximal subsets of S) gives S. }

Therefore, the equivalence classes of an equivalence relation define what is known as a partition of a set S.

- Partition of a set

Partition of S

$$\pi = \{ B : B \subseteq S \}$$

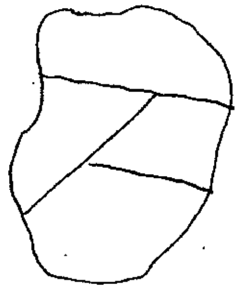
if $B, B' \in \pi$ and $B \neq B'$ then $B \cap B' = \phi$

if $B \in \pi$ then $B \neq \phi$

$$\bigcup_{B \in \pi} B = S$$

Next to a set, the notion of a partition is the next most important notion in combinatorics.

A partition is typically viewed as taking a set and cutting it up.



Notice, again, the same sort of strange phenomenon (i.e., relation : bipartite graph [p 2.1]):

The notion of a partition and the notion of an equivalence relation are mathematically equivalent, though psychologically different.

- Every partition defines an equivalence relation:

  Given a partition π, set $\sim_\pi$ = equivalence relation defined by π

- Every equivalence relation defines a partition:

  Given an equivalence relation R, set π_R = partition defined by R

We're not going to give examples of all of these, as you are going to see hundreds of them. Also, you should be slightly familiar w/ these notions, We're just filling in the gaps.

Now let's go back to Boolean algebra for 5 minutes.

Boolean Algebra (cont'd)

Consider the Boolean algebra of $\overbrace{\text{all subsets of a set } S}^{P(S)}$.

Let's define the notion of Boolean subalgebra of sets.

A Boolean subalgebra $\mathcal{B}$ (script B) of $P(S)$ is a subfamily of $P(S)$ containing ϕ, S, and closed under arbitrary (even infinite) $\cup, \cap, {}^c$.

To stress the fact that you are allowed to take infinite unions and intersections, we sometimes say this is a:

Complete Boolean subalgebra

↳ shorthand for permission to take arbitrary unions + intersections.

- There is a remarkable relationship between the family of all Boolean subalgebras of $P(S)$ and the family of all partitions.
  We now make this explicit.

$$\mathbb{B} = \text{given Boolean subalgebra of } P(S)$$

Let's take:
$a \in S$

$$\bigcap_{\substack{A \in \mathbb{B} \\ a \in A}} A$$

and we also take:
$b \in S$

$$\bigcap_{\substack{A \in \mathbb{B} \\ b \in A}} A$$

these 2 intersections will be either <u>identical</u> or <u>disjoint.</u> ← {I leave it for you to realize that.}

What is a set of this form? A set of this form is the <u>minimal elements</u> of the Boolean subalgebra $\mathbb{B}$. And we ensure that it is <u>non-empty</u> by making it contain a. $\Big\}$ $\bigcap_{\substack{A \in \mathbb{B} \\ a \in A}} A$

So if we take 2 different minimal non-empty elements of the Boolean subalgebra $\mathbb{B}$, they will be disjoint.
That's obvious.
If you don't see it, sit down and realize it.

Therefore, the minimal non-empty members of $\mathbb{B}$ are a partition $\pi_{\mathbb{B}}$ of the set S.

- Therefore, we have the following result:

  Every Boolean subalgebra is completely determined by the partition.

  If you find the partition, you find the Boolean subalgebra.

> There is a 1-1 correspondence between the Boolean subalgebra of $P(S)$ and the partitions of S.

- There is a bijection between the family of all (complete) Boolean subalgebras of $P(S)$ and the family $\Pi[S]$ of all partitions of the set S.

  {To stress fact that we allow arbitrary unions + intersections.}

We will see next time this bijective correspondence is also order inverted.

Above proves the entire result because it's so obvious.
And I hope it's obvious to you, too.

The essence of the result is that:

you have 2 different Boolean subalgebras $\Longleftrightarrow$ you have 2 different partitions

I leave it to you to verify that this is so.

I also leave it to you to prove:

Given $\pi \in \Pi[S]$ we define a Boolean subalgebra of $P(S)$ consisting of all unions of members of π.

aka "blocks"

Reasoned Review ← a phrase the French often use

So far, we have been studying sets and relations.

- S = set

  $P(S)$ = Boolean algebra of all subsets of S

  The most important fact about this Boolean algebra, which I keep insisting upon, is that you can take arbitrary unions and intersections.

  This doesn't happen for other Boolean algebras, so this is an exceptional Boolean algebra.

  For this reason we call it complete ← i.e., arbitrary unions + intersections are allowed.

- Then we studied relations:

  $R \subseteq S \times T$

- And, in particular, we discussed a relation R on the set S:

  $R \subseteq S \times S$

For the following, assume $R \subseteq S \times S$:

- The family of relations on S (say) is a complete Boolean algebra $P(S \times S)$, which, in addition to union, intersection, and complement, has the additional operation of composition: $R \circ R'$

- Every relation has an inverse relation:

  R^{-1} exists

The family of relations on S is a complete Boolean algebra, with two additional operations:

(1) binary operation of composition

and (2) inverse

← NB: inverse is quite different than the complement.
Observe that you have R^c, which is the complement of the relations between the sets.

$(\cup, \cap, {}^c, {}^{-1}, \circ) \Rightarrow$ complete Boolean algebra

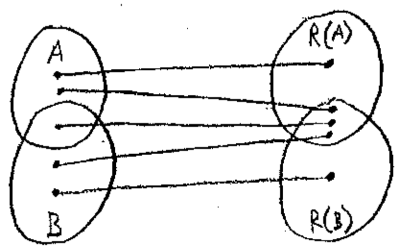
$B \subseteq S$

- With $A \subseteq S$, we define:

$$R(A) = \{ b : (a, b) \in R \}$$

Then we have:

- $$R(A \cup B) = R(A) \cup R(B)$$

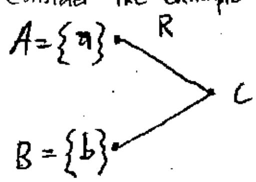


<u>Note</u> $\Rightarrow$ $R(A \cap B) \neq R(A) \cap R(B)$

and we have:

- $$R(A \cap B) \subseteq R(A) \cap R(B)$$

Consider the example:



$$R(\underbrace{A \cap B}_{\{a\} \cap \{b\} = \phi}) \subseteq \underbrace{R(A)}_{\{c\}} \cap \underbrace{R(B)}_{\{c\}}$$

$R(\phi) = \phi$

$\phi \subseteq \{c\}$

- <u>Exercise 3.1</u>

<u>Note</u> $\Rightarrow$ $R(A^c) \neq (R(A))^c$

Find a counterexample.

Exercise 3.2

Note that $(R \cup R')$ is a relation – you put all the edges together.
If you compose this with another relation R'', nothing wrong happens,

$$(R \cup R') \circ R'' = (R \circ R'') \cup (R' \circ R'')$$

And similarily:

$$(R \cap R') \circ R'' = (R \circ R'') \cap (R' \circ R'')$$

Prove these.

These are pretty much all the identities satisfied linking Boolean operations with composition.

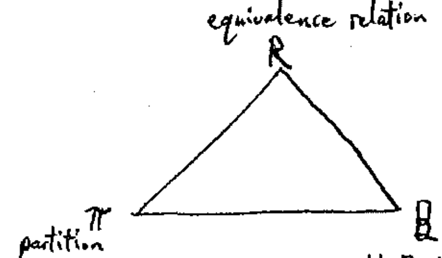
- To be honest, there is an additional operation among relations that has been studied, but it's a little hairy to discuss at this point.
  People like Lyndon + Tarski looked at this.

We continue our reasoned review.
After this, we discussed equivalence relations.

- Special relations: $R \subseteq S \times S$
  - universal relation $U_S = S \times S$ ← {every possible pair is in the universal relation}
  - identity relation $I = \{(a,a) : a \in S\}$
  - equivalence relation
    (1) $R \supseteq I$ reflexive
    (2) $R = R^{-1}$ symmetric
    (3) $R \circ R \subseteq R$ transitive

- Equivalence relations, partitions, complete Boolean subalgebras
  R π $\mathcal{B}$

  These 3 concepts are cryptomorphic
  ↑ {as I love to say.
  This is a word which will remain admirably undefined.}

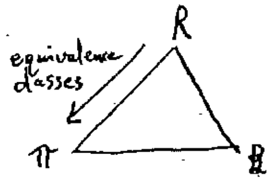


Complete Boolean subalgebra of S

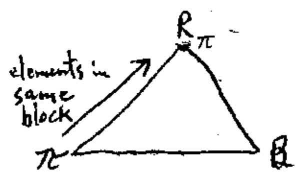
Boolean subalgebra of S is a subset of a family of sets, which is a Boolean algebra in its own right.
Complete indicates we can take arbitrary intersections and unions.

If we allowed only finite intersections + unions, all hell breaks loose.
The theory then becomes extremely weak.
The theory is easy because we allow arbitrary unions and intersections.

If you only allow countable unions and intersections, you get probability.



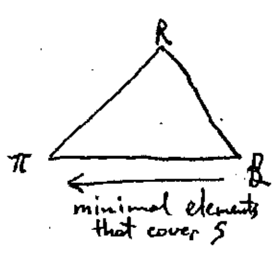
- Starting w/ an equivalence relation R, we get the partition π of equivalence classes. [2.11]



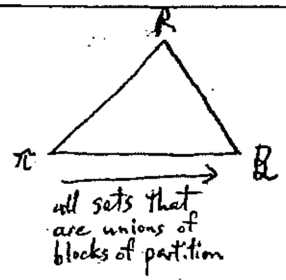
- Conversely, given partition π, we get equivalence relation R_π, where two elements are equivalent if they are in the same block of the partition.

$(a,b), (b,a) \in R_\pi$

$(a,c), (c,a), (b,c), (c,b) \notin R_\pi$

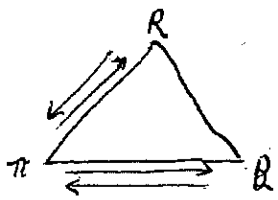


- Given a Boolean subalgebra $\mathcal{B}$, take the minimal elements of the Boolean subalgebra [p 2.13].
Any two minimal elements are disjoint (otherwise their intersection would be more minimal).
Take the disjoint minimal elements that cover set S.
These disjoint minimal elements form a partition of the set S.



Conversely, if you are given a partition, you take all sets that are unions of blocks of the partition.
That's a complete Boolean subalgebra [p 2.13-14].

So you go ring-around-the-rosie.
The three concepts are equivalent.



Basic enumeration

Take set S finite $|S| < \infty$, $|S| = n$

- How many subsets of S are there?
You've known this since the age of 5.

$$|P(S)| = 2^n$$

- How many subsets of S are there w/ k elements?
You've also known this since the cradle.

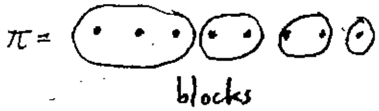
$$\left|\{A : A \subseteq S : |A| = k\}\right| = \binom{n}{k}$$
$$k \le n$$

Big deal.
I assume you already know this.
Now, let's turn the screw.
Let's do the same thing for partitions.

Here we have a finite set:

$S = \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$

And here we have the partitions:



Now we ask:

How many partitions of S are there?
How many partitions of S are there w/ k blocks?

} You better know these. Bread + butter questions.

- The number of partitions, of the set S, with k blocks

$$= S(n,k)$$

Stirling numbers of the 2^{nd} kind
(I'm very sorry. It's not my fault.
That's the way they are called.)

So, our objective is to find some formula for the Stirling numbers of the 2^{nd} kind. Guess how we're going to do that? Balls into boxes. You knew that was coming.

Let's consider the set S and the partition the set T.

$|S| = n$

$|T| = x$

{the number of elements in the set T.
x is an unusual way of denoting an integer.}

Then we take functions from S to T, where S is the balls (distinguishable) and T is the boxes (distinguishable)

The number of functions from S to T $= x^n \longleftarrow \overbrace{x \cdots x}^{n}$

{each ball can go into one of x boxes, irrespective of where other balls have gone.}

The number of $\overset{\text{1-to-1}}{\text{monomorphic}}$ functions from S to T is:

$$\overbrace{x(x-1)(x-2)\cdots(x-n+1)}^{n \text{ terms}} = (x)_n \longleftarrow \text{"}x \text{ lower factorial } n\text{"}$$

After x^n, the lower factorials are the most important polynomials.

Now, we use linear algebra.
I could compute the formula for the Stirling numbers directly (see 18.313 Super class 2 notes [SC 3/13/98, 10-14])
But I'd rather use linear algebra.

Let $R[x]$ = vector space of all polynomials $p(x)$ ← all polynomials in x w/ <u>real coefficients</u>

(all our vector spaces will have real coefficients, unless otherwise specified.)

A <u>basis</u> of $R[x]$ is $1, x, x^2, \ldots$

← set of vectors that are (a) linearly independent
(b) span the vector space

Another basis of $R[x]$ is $1, x, (x)_2, (x)_3, \ldots$

[Because you have one polynomial for each degree.
Therefore, you can express x^n as a linear combination:
$x^n = c_0 1 + c_1 x + c_2 (x)_2 + c_3 (x)_3 + \ldots + c_n (x)_n$]

Fine.
Why am I saying all this?
For the following reason.

Given $f: S \to T$, one defines the $\overbrace{\text{Kernel of } f}^{\text{Ker}(f)}$, say π_f, as the partition of S whose blocks are the sets:

$$f^{-1}(b) \ , \ b \in T$$

whenever $|f^{-1}(b)| \neq \phi$

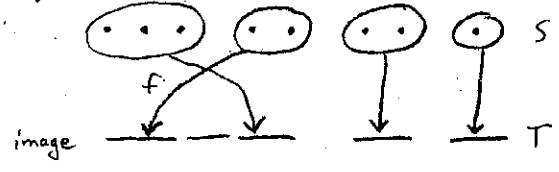
The inverse function f^{-1} of elements (i.e., blocks) of T, which are sets that are disjoint, form a linear function.

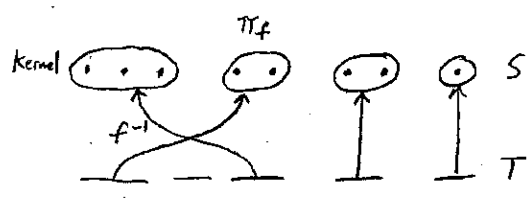
Every function has a Kernel, which is a partition.
To every function, you associate 2 things:

(1) image – which is a <u>subset</u> of T
(2) Kernel – which is a <u>partition</u> of S
} duals

Philosophically, anything you can say about subsets of T, you can turn into saying something about partitions of S.
That's the guiding principle.

Example:





OK, So what?
Now we ask:

How many functions are there w/ a given kernel?

The number of functions whose kernel is the partition π of S is:

$$(x)_{|\pi|}$$ ← x lower factorial number of blocks of π

Why? Because you treat each block as an element.
And you just put each block in a different box.

Since every function has a kernel, we obtain the following important identity:

$$x^n = \sum_{\pi \in \Pi[S]} (x)_{|\pi|}$$

total number of functions → x^n

← the number of functions whose kernel is the partition π of S.

↑ π ranges over all partitions of S

You can split the RHS sum in many ways.
In particular, you can split it by taking all partitions π that have k blocks.
The number of partitions of S that have k blocks is just the Stirling number of the 2nd kind — $S(n,k)$.

$$= \sum_{k=1}^{n} S(n,k)\,(x)_k$$

And we have our identity:

$$(*) \qquad \boxed{x^n = \sum_{k=1}^{n} S(n,k)\,(x)_k}$$

This is a purely numerical identity between polynomials.

From this identity, we need to get the formula for the Stirling numbers of the 2^{nd} kind. Here's how we do it.

Difference Operator

The difference operator is defined as:

$$\Delta p(x) = p(x+1) - p(x)$$

So, we have:

$$\Delta (x)_k = (x+1)_k - (x)_k$$

$$= \underbrace{(x+1)(x)\cdots(x-k+2)}_{(x)_{k-1}} - \underbrace{(x)(x-1)\cdots(x-k+2)}_{(x)_{k-1}}(x-k+1)$$

$$= \left((x+1) - (x-k+1)\right)(x)_{k-1}$$

$$\boxed{\Delta (x)_k = k\,(x)_{k-1}}$$

Delta acts on the lower factorials like the derivative acts on monomials.

$$\Delta : (x)_k :: D : x^k$$

$$k(x)_{k-1} :: kx^{k-1}$$

Iterating the operator Δ gives:

$$\Delta^j (x)_k = (k)_j\,(x)_{k-j}$$

We note that:

$$\Delta^j (x)_k = \begin{cases} 0 & \text{if } j > k \\ k! & \text{if } j = k \\ (k)_j\,(x)_{k-j} & \text{if } j < k \end{cases}$$

In all cases, we have:

$$\left[\Delta^j (x)_k\right]_{x=0} = k!\,\delta_{jk}$$

Kronecker Delta
Equals one only if $j=k$

if $j > k$, $\Delta^j (x)_k = 0$

if $j < k$, $\Delta^j (x)_k = (k)_j\,(x)_{k-j}$

$$= (k)_j\,\cancel{x}^{0}(x-1)_{k-j-1}$$

$$= 0$$

Now we apply Δ^j to both sides of equation (*) and set $x=0$.

$$\left[\Delta^j x^n\right]_{x=0} = \left[\Delta^j\left(\sum_{k=1}^{n} S(n,k)(x)_k\right)\right]_{x=0}$$

$$= \sum_{k=1}^{n} S(n,k)\left[\Delta^j (x)_k\right]_{x=0}$$

$$= \sum_{k=1}^{n} S(n,k)\, k!\, \delta_{jk}$$

$\delta_{jk}=0$ except when $j=k$, so only a single term survives the summation.

$$\left[\Delta^j x^n\right]_{x=0} = S(n,j)\, j!$$

And this gives us an expression for the Stirling numbers of the 2nd kind:

$$\boxed{S(n,j) = \frac{\left[\Delta^j x^n\right]_{x=0}}{j!}}$$

the number of partitions of the set S ($|S|=n$) with j blocks.

Because of this expression, the British call the Stirling numbers of the 2nd kind:

"The differences of zero"

- Now, let's look at one of the most important identities in mathematics.

Given $p(x), q(x) \in R[x]$ ← $p(x)$ and $q(x)$ are polynomials w/ real coefficients.

It is clear what we mean by $p(D)$, where $D = \frac{d}{dx}$ (the derivative). Just replace the powers of x by the powers of D.

The following identity is one of the most useful that occurs throughout algebra, linear algebra:

$$\left[p(D)q(x)\right]_{x=0} = \left[q(D)p(x)\right]_{x=0}$$

Proof:

By linearity, we only need to prove this when $p(x)$ is some power of x and $q(x)$ is some power of x.

So we check when: $p(x) = x^n$
$q(x) = x^k$

LHS:

$$\left[p(D)\,q(x)\right]_{x=0} = \left[D^n x^k\right]_{x=0}$$

if $n > k$, you differentiate the hell out of it and the result is 0.

if $n < k$, the resulting polynomial is some monomial in x, and since $x=0$, the result is 0.

if $n=k$, $D^n x^n = n!$

$$= n!\,\delta_{kn}$$

similarily:
RHS:

$$\left[q(D)\,p(x)\right]_{x=0} = \left[D^k x^n\right]_{x=0}$$

$$= k!\,\delta_{nk}$$

And we have the tautology:

$$n!\,\delta_{kn} = k!\,\delta_{nk}$$

$0 = 0$ for all $n \neq k$
$n! = n!$ for $n = k$

So it checks.

Something I forgot to tell you.
Taylor's formula.

$$\Delta = e^D - I$$

Why?

$$p(x+1) = \sum_{j=0}^{\infty} \frac{D^j}{j!}\,p(x)$$

from Taylor's formula:

$$p(u) = \sum_{j=0}^{\infty} \frac{D^j\,p(c)}{j!}(u-c)^j$$

Let $u = x+1$
$c = x$

$$p(x+1) = \sum_{j=0}^{\infty} \frac{D^j}{j!}\,p(x)$$

we can write this as:

$$p(x+1) = e^D\,p(x)$$

The difference operator gives:

$$\Delta p(x) = p(x+1) - p(x)$$
$$= e^D p(x) - p(x)$$
$$\Delta p(x) = (e^D - I)\,p(x) \implies \Delta = e^D - I$$

- We can get another expression for the Stirling numbers of the 2nd kind as follows.
Recall [p 3.10]:

$$S(n,j) = \frac{\left[\Delta^j x^n\right]_{x=0}}{j!}$$

We can express the numerator, using the fact that $\Delta = e^D - I$, as:

$$\left[\Delta^j x^n\right]_{x=0} = \Big[\underbrace{(e^D - I)^j}_{p(D)} \underbrace{x^n}_{q(x)}\Big]_{x=0}$$

We make use of the identity proved on [p 3.10-11], namely:

$$\left[p(D)q(x)\right]_{x=0} = \left[q(D)p(x)\right]_{x=0}$$

$$p(D) = (e^D - I)^j \implies p(x) = (e^x - 1)^j$$

$$q(x) = x^n \implies q(D) = D^n$$

$$= \left[D^n (e^x - 1)^j\right]_{x=0}$$

And this gives us a second formula for Stirling numbers of the 2nd kind:

$$\boxed{S(n,j) = \frac{\left[D^n (e^x - 1)^j\right]_{x=0}}{j!}}$$

- A third expression for the Stirling numbers of the 2^{nd} kind

Define the shift operator E as:

$$E\,p(x) = p(x+1)$$

Thus:

$$\Delta = E - I \longleftarrow \quad \Delta\,p(x) = p(x+1) - p(x)$$
$$= E\,p(x) - p(x)$$
$$\Delta\,p(x) = (E - I)\,p(x) \quad \Rightarrow \Delta = E - I$$

$$\Delta^j = (E - I)^j$$

We expand this by the binomial theorem

$$= \sum_{i=0}^{j} \binom{j}{i} (-1)^{j-i} E^i$$

Recalling our first expression for Stirling numbers of the 2^{nd} kind [p 3.10]:

$$S(n,j) = \frac{\left[\Delta^j x^n\right]_{x=0}}{j!}$$

Substituting above for Δ^j:

$$= \frac{1}{j!}\left[\sum_{i=0}^{j} \binom{j}{i} (-1)^{j-i} \underbrace{E^i x^n}\right]_{x=0}$$

$$E^i x^n = (x+i)^n$$

$$= \frac{1}{j!}\left[\sum_{i=0}^{j} \binom{j}{i} (-1)^{j-i} (\cancel{x}^{0}+i)^n\right]_{x=0}$$

And we obtain our third expression:

$$\boxed{S(n,j) = \frac{1}{j!} \sum_{i=0}^{j} \binom{j}{i} (-1)^{j-i}\, i^n}$$

Exercise 3.3

The formula we have just proved:

$$S(n,j) = \frac{1}{j!}\sum_{i=0}^{j}\binom{j}{i}(-1)^{j-i}\, i^n$$

reminds us of the inclusion-exclusion formula.

Prove this by the inclusion-exclusion principle
(This can be given a direct combinatorial proof)

← uses onto functions
use inclusion-exclusion to compute #ontos by taking number that exclude a specific element - take union = fcns that exclude some element and then complement. Divide #epis by (#boxes)! to get Stirling # 2nd kind.

Now we address the question of the total number of partitions.
This is more complicated.

We can make use of $S(n,k)$ to write the equation:

$$\underbrace{\text{Total \# of partitions of an } n \text{ element set}}_{B_n} = \sum_{k=1}^{n} S(n,k)$$

B_n ← we give this the name B_n.
These are called the Bell numbers.

We go back to the vector space $R[x]$.
Because this is a vector space, we can define a $\overbrace{\text{linear functional}}^{L}$ on this vector space.

You define a linear functional by telling what it does for every element of a basis.
By so doing, since the basis spans the vector space, you've implicitly defined the linear functional over the whole vector space.

Define linear functional L on $R[x]$ by setting:

$$L(\underbrace{(x)_n}) = 1 \quad , \quad n = 0, 1, 2, \ldots$$

polynomials $(x)_n$, $n = 0,1,2,\ldots$ are a basis for the vector space $R[x]$

Now watch.
This is pretty cute.
Recall [p 3.8] our formula for the total number of functions from S to T:

$$x^n = \sum_{\pi \in \Pi[S]} (x)_{|\pi|}$$

Apply L to both sides:

$$L(x^n) = L\left(\sum_{\pi\in\Pi[S]} (x)_{|\pi|}\right)$$

$$= \sum_{\pi\in\Pi[S]} L\left((x)_{|\pi|}\right)$$

When you apply the operator L to the RHS above,
every partition gives you a contribution of 1.
This is because $L((x)_n) = 1$, since $(x)_n$, $n=0,1,2,\ldots$
is a basis.
So the sum on the RHS is <u>the number of partitions.</u>
This is exactly what we are after.

$$= B_n$$

Thus, we have a formula for the Bell numbers:

$$\boxed{B_n = L(x^n)}$$

That's the formula.
It's a nice formula.
Now I know that you want something numerical.
You're not used to seeing formulas w/ linear functionals.
So next time, we'll rehash this w/o linear functionals.

John Guidi
guidi@math.mit.edu

Basic enumeration (cont'd)

We continue our study of the enumeration of the partitions of a set.
We recall that:

$$S = \text{set} \quad (\text{finite or infinite})$$

$$\Pi[S] = \text{set of all partitions of } S$$

Last time, we defined the Bell numbers as the number of elements of $\Pi[S]$:

$$\underset{\text{Bell numbers}}{B_n} = |\Pi[S]|$$

The Stirling numbers of the 2nd kind, which we studied last time, are the analogues for partitions of the binomial coefficients.
The Bell numbers are analogues for partitions of 2^n.

$$\underbrace{S(n,k)}_{\substack{\#\text{ of partitions}\\ \text{of } S \text{ w/ } k \text{ blocks}}} : \underset{\substack{\#\text{ of subsets}\\ \text{of } S \text{ w/ } k \text{ elements}}}{\binom{n}{k}} :: \underbrace{B_n}_{\substack{\#\text{ of partitions}\\ \text{of } S}} : \underset{\substack{\#\text{ of subsets}\\ \text{of } S}}{2^n}$$

- Last time we saw a formula for the Bell numbers.
You take:

$$R[x] = \text{vector space of polynomials in } x$$

Linear functional L is defined by setting:

$$L((x)_n) = 1, \quad n = 0, 1, 2, \ldots$$

A linear functional is well defined, once it is defined on the elements of a basis. As it is here.

Therefore, the linear functional is well defined on the vector space $R[x]$

From the Fundamental Identity we established last time [p 3.8]:

$$x^n = \sum_{\pi \in \Pi[S]} (x)_{|\pi|}$$

We applied the linear functional L to both sides, we observed:

$$L(x^n) = L\Big(\sum_{\pi\in\Pi[S]} (x)_{|\pi|}\Big)$$

$$= \sum_{\pi\in\Pi[S]} L\big((x)_{|\pi|}\big)$$

and since $L((x)_n) = 1$, $n=0,1,2,\ldots$ the RHS adds a count of 1 for each partition.

$$= B_n$$

And so we obtain immediately:

$$\boxed{B_n = L(x^n)}$$

[That this formula is not explicit is a predjudice.
Because you are used to seeing formulas in terms of something else.
In reality, this is as acceptable a formula as a formula given by a generating function, or any other thing.
You can work w/ it.

• But let's give in to our predjudice and give it more explicitly.
The explicit formula is:

<u>Dobinski's formula</u> – an explicit formula for the Bell numbers

$$B_{n+1} = \frac{1}{e}\left(1^n + \frac{2^n}{1!} + \frac{3^n}{2!} + \frac{4^n}{3!} + \cdots\right)$$

It's not even obvious that this is an <u>integer</u>!

<u>Proof</u>
Let's prove this.
First we notice that:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} \Longleftarrow$$

Taylor expansion:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{[k]}(c)}{k!}(x-c)^k$$

with: $f(x) = e^x$
$x = 1$
$c = 0$

$$e = \sum_{k=0}^{\infty} \frac{1}{k!}$$

- Suppose I write this expression:

$$\sum_{k=0}^{\infty} \frac{(k)_3}{k!} = \frac{\cancel{(0)_3}^{\,0}}{0!} + \frac{\cancel{(1)_3}^{\,0}}{1!} + \frac{\cancel{(2)_3}^{\,0}}{2!} + \frac{\cancel{(3)_3}^{\,1}}{\cancel{3!}} + \frac{\cancel{(4)_3}^{\,\frac{1}{1!}}}{\cancel{4!}} + \frac{\cancel{(5)_3}^{\,\frac{1}{2!}}}{\cancel{5!}} + \cdots$$

$$\frac{(k)_3}{k!} = \frac{\cancel{k}(\cancel{k-1})(\cancel{k-2})}{\cancel{k}(\cancel{k-1})(\cancel{k-2})(k-3)\cdots 1} = \frac{1}{(k-3)!}$$

$$\frac{(k+1)_3}{(k+1)!} = \frac{\cancel{(k+1)}\,\cancel{k}\,(\cancel{k-1})}{\cancel{(k+1)}\,\cancel{k}\,(\cancel{k-1})(k-2)\cdots 1} = \frac{1}{((k+1)-3)!}$$

$$= 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots$$

$$\sum_{k=0}^{\infty} \frac{(k)_3}{k!} = e$$

For clarity, I did this for lower factorial 3. But the reasoning works for any n. Therefore, we have the following:

In combinatorics, proofs are often clearer if you do them for one example and then generalize.

$$\sum_{k=0}^{\infty} \frac{(k)_n}{k!} = e$$

$$\boxed{1 = \frac{1}{e} \sum_{k=0}^{\infty} \frac{(k)_n}{k!}}$$

But what is 1?

$$1 = L\left((x)_n\right) \longleftarrow$$ Because that is my pleasure. That is how I defined the linear functional L.

Watch how this unfolds.

$$L((x)_n) = 1$$

$$L((x)_n) = \frac{1}{e}\sum_{k=0}^{\infty}\frac{(k)_n}{k!}$$

$$\sum_{n=0}^{j} a_n 1 = \sum_{n=0}^{j} a_n 1$$

$$\sum_{n=0}^{j} a_n L((x)_n) = \sum_{n=0}^{j} a_n \left(\frac{1}{e}\sum_{k=0}^{\infty}\frac{(k)_n}{k!}\right)$$

$$\sum_{n=0}^{j} L(a_n (x)_n) = \frac{1}{e}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{n=0}^{j} a_n (k)_n$$

$$L\Bigg(\underbrace{\sum_{n=0}^{j} a_n (x)_n}\Bigg) = \frac{1}{e}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{n=0}^{j} a_n (k)_n$$

Since the $(x)_n$, $n=0,1,2,\dots$ are a basis for the vector space $R[x]$, any polynomial can be writen in this form.

Namely:

$$p(x) = \sum_{n=0}^{j} a_n (x)_n$$

Note, then, that: $p(k) = \sum_{n=0}^{j} a_n (k)_n$

This gives:

$$L(p(x)) = \frac{1}{e}\sum_{k=0}^{\infty}\frac{1}{k!}\,p(k)$$

And the above is true for any polynomial $p(x)$.
We choose:

$$p(x) = x^n$$

$$L(x^n) = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}$$

And, as we've already shown, $B_n = L(x^n)$, which gives:

$$\boxed{B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}}$$

Dobinski's formula

For many years, this formula appeared to me <u>very mysterious.</u>
Then, one day, I realized it was <u>trivial.</u>
That's what happens in mathematics.
It just took a long time for me to realize that it was trivial.
Let me to you that it is trivial.
Let us <u>demythologize</u> the proof.

To do so requires that you know a little <u>probability.</u>
If you don't know any probability, take a nap.

Let X = <u>Poisson random variable</u> of <u>intensity 1</u> $\lambda = 1$

$$P(X=k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{if } k \geq 0 \\ 0 & \text{if } k < 0 \end{cases}$$

$$E(X) \doteq \sum_{k=-\infty}^{\infty} k\, P(X=k)$$

$$= \sum_{k=0}^{\infty} k \frac{1}{k!} e^{-1}$$

$$= \frac{1}{e} \underbrace{\sum_{k=1}^{\infty} \frac{1}{(k-1)!}}_{e}$$

$$E(X) = 1$$

Recall that $E(cX^n) = \sum_{k=-\infty}^{\infty} k\, P(cX^n = k)$

$$= \sum_{k=-\infty}^{\infty} k\, P\left(X = \tfrac{1}{c}\sqrt[n]{k}\right)$$

change of variables
$k \leftarrow ck^n$

$$E(cX^n) = \sum_{k=-\infty}^{\infty} ck^n\, P(X=k)$$

Now we consider the expectation of the lower factorial of this Poisson random variable:

$$E\left((X)_n\right) = \sum_{k=-\infty}^{\infty} k\, P\left((X)_n = k\right)$$

$$= \sum_{k=-\infty}^{\infty} (k)_n\, P(X=k)$$

$$= \sum_{k=0}^{\infty} (k)_n \frac{1}{k!} e^{-1}$$

$$= \underbrace{\frac{1}{e} \sum_{k=0}^{\infty} \frac{(k)_n}{k!}}_{\text{and we just showed that this is } 1 \text{ [p 4.3].}}$$

$$E\left((X)_n\right) \doteq 1$$

So the expectation of the lower factorial of the Poisson random variable w/ intensity $\lambda = 1$ is 1.

This is known in statistics as the factorial moment of this random variable.

It follows that:

$$E(X^n) = \sum_{k=-\infty}^{\infty} k\, P(X^n = k)$$

$$= \sum_{k=-\infty}^{\infty} k^n\, P(X=k)$$

$$= \sum_{k=0}^{\infty} k^n \frac{1}{k!} e^{-1}$$

$$\boxed{E(X^n) = \frac{1}{e}\sum_{k=0}^{\infty} \frac{k^n}{k!}} \longleftarrow \text{But this is } \underline{\text{Dobinski's formula.}}$$

So we have:

$$\underbrace{B_n}_{L(x^n)} = \frac{1}{e}\sum_{k=0}^{\infty} \frac{k^n}{k!} \qquad \text{and} \qquad E(X^n) = \frac{1}{e}\sum_{k=0}^{\infty} \frac{k^n}{k!}$$

$$\boxed{L(x^n) = E(X^n)}$$

[And we have that the linear functional of x^n is the same as the expectation for that random variable for the Poisson process w/ intensity $\lambda = 1$.]

This is about as simple a number as we have for the Bell numbers.

Exercise 4.1

Find the recursion formula for the Bell numbers, using linear functionals. Namely, show that:

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

Now let's consider some finer enumerations.
We've enumerated partitions by the number of blocks:

$$S(n,k) = \text{number of partitions of set with } k \text{ blocks}$$

Stirling numbers of the 2nd kind

And we've enumerated the total number of partitions:

$$B_n = \text{number of partitions of set}$$

Bell number

What other things are of interest?

Let's look at a partition of a set.

There is 1 block w/ 3 elements, 2 blocks w/ 2 elements, 4 blocks w/ 1 element.

The finer count on a partition counts how many blocks there are w/ each number of elements.

Let's make that precise:

if $\overset{\text{partition}}{\pi} \in \Pi[S]$, then the type of π is the multiset of integers:

$$\{|B| : \underset{\uparrow \text{block}}{B} \in \pi\}$$

Now, we have to say a few words about multisets.
In other words, you take the number of elements of each block.
This gives you a family of integers.

this family is not a set.
Because some of the integers may be repeated.

For example, the type of the partition is: $\{3, 2, 2, 1, 1, 1, 1\}$

this is not a set.
It's a multiset.

- A digression on multisets.

Multisets in the 19th century were called "combinations"

Hence the words permutations and combinations.

Let T = set

A multiset M is a function from T to $\mathbb{N}$. ← non-negative integers

For $t \in T$, we have $m(t)$, the multiplicity of t in the multiset M.

$\left\{ \begin{array}{l} \text{In other words, } m(t) \text{ tells you how many times element } t \text{ appears in} \\ \text{the multiset } M. \end{array} \right\}$

We say:

$$\underbrace{|M| < \infty}_{\text{multiset } M \text{ is finite}} \quad \text{when} \quad \sum_{t \in T} m(t) < \infty$$

For $t \notin T$, $m(t) = 0$.

This is the only rigorous definition I can give, other than the handwaving definition you are accustomed to.

There is an <u>algebra</u> <u>of multisets</u>.

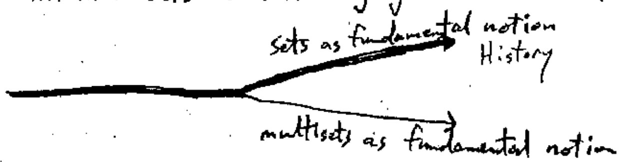
Just as we have seen for sets, where there is an algebra (i.e., Boolean algebra), there is an algebra of multisets.
But, for historical reasons, the algebra of multisets is much less developed than the algebra of sets.
I want to defer this discussion.

Let's pause for a minute and realize that this is an accident of history.
In <u>nature</u>, <u>multisets</u> occur as frequently as sets.

Statisticians talk about:

sampling w/o replacement $\rightarrow$ sets
" w/ " $\rightarrow$ multisets

Multisets are a very natural concept.
It's an accident of history that the foundations of mathematics has been developed in terms of sets, rather than multisets.
You can imagine a different evolutionary pattern, where the foundations of mathematics might have been developed using the notion of <u>multisets</u> as the <u>fundamental notion</u> and the notion of sets is something you think about later.



More about this later.

Now, back to partitions: $\{|B| : B \in \pi\}$

The <u>type</u> of a <u>partition</u> of a <u>finite set</u> is a <u>multiset</u> <u>of Integers</u>.
What kind of a <u>multiset of integers</u>?
A <u>multiset of integers</u>, whose elements add up to the number of elements of the set S.

$$\sum_{\substack{B \in \pi \\ \uparrow \\ \text{block of partition } \pi \text{ of set } S}} |B| = \underbrace{n}_{\text{number of elements in set } S.} = |S|$$

Warning:
It is unfortunate, but the word partition is used in 2 different senses.
It's not my fault.
And up until 20 years ago, people systematically confused the 2 notions.

People confused partition of a number w/ partition of a set.

But the names have stuck.

- A partition of an integer $n \in \mathbb{N}$ is a multiset of positive integers, whose sum equals n.

For example, for $n = 5$, you can list all possible partitions of 5:

5	$\{5\}$
$4+1$	$\{4,1\}$
$3+2$	$\{3,2\}$
$3+1+1$	$\{3,1,1\}$
$2+2+1$	$\{2,2,1\}$
$2+1+1+1$	$\{2,1,1,1\}$
$1+1+1+1+1$	$\{1,1,1,1,1\}$

The order of the summands does not matter, because these are multisets.
It's just convenient to arrange the summands in non-increasing order.

The theory of partitions of a number is one of the most developed branches of mathematics and the intersection of combinatorics and number theory.
There are some extremely deep results.
Some of which are due to Srinivasa Ramanujan, the great Indian mathematican.
Some of the deepest results in both combinatorics and number theory are results from the partitions of a number.

Unfortunately, there is no simple formula for the number of partitions of an integer.
There is a generating function, but I don't want to do this yet.

Note that:

- the type of $\pi \in \Pi[S]$ is a partition of the integer n

Now we come to something that sounds trivial and people take for granted for a long time until someone comes along and says:

"Hey, wait a minute. Is this really trivial?"

Then all hell breaks lose.

There are 2 notations to denote the <u>type of a partition</u> π:

$$\{|B| : B \in \pi,\ \pi \in \Pi[S]\}$$

(1) You take the multiset, whose elements are the sizes of the blocks, and <u>arrange</u> them in <u>non-increasing order.</u>

$$\lambda_1 \geq \lambda_2 \geq \ldots$$

where $\lambda_i > 0$ and $\sum_i \lambda_i = n$

(2) Look at all the sizes of the blocks,
Then count how many blocks there are w/ 1 element (r_1),
2 elements (r_2), etc.
Clearly, this gives the <u>same</u> information as notation (1).
The standard notation here is (again, don't blame me):

$$\underbrace{1^{r_1}}_{r_1 \text{ blocks w/ 1 element}} 2^{r_2} \ldots$$

So, these are 2 notations for the same concept.

The 1^{st} notation leads to a graphical representation that is extremely useful.
<u>Ferrers relation</u> of a partition of an integer n
As you recall, a relation may be defined by its <u>incidence matrix.</u>
It is the relation whose incidence matrix is represented as follows.
First we have the marginals:

$$\begin{array}{c} \lambda_1 \\ \lambda_2 \\ \vdots \\ \\ \end{array} \begin{bmatrix} \overbrace{1 \ \cdots \ 1}^{\lambda_1 \text{ 1's}} & 0 & \cdots & 0 \\ \overbrace{1 \ \cdots \ 1}^{\lambda_2 \text{ 1's}} & 0 & \cdots & 0 \\ & & 0 & \\ 0 & & \cdots & 0 \end{bmatrix}$$

Ferrers Matrix
of a relation

You get a matrix where the set of non-zero entries are contained in each other.

The matrix becomes more sparse as you go down the rows.

(It doesn't matter if you consider this an infinite matrix filled w/ 0's, or a finite matrix.)

<u>Warning:</u>
In all the books, the Ferrers relation is written w/ dots, instead of 1's + 0's.
Ex:

$$\begin{array}{|cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \\ \cdot & & & \end{array}$$

Example:
Consider the following partitions of the integer $n=7$ and their associated incidence matrices of the Ferrers relations:

$\{3,2,1,1\}$ $\{4,3\}$

$$\lambda_1=3 \begin{bmatrix} 1 & 1 & 1 & \\ 1 & 1 & & \\ 1 & & 0 & \\ 1 & & & \end{bmatrix} \qquad \lambda_1=4 \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & & \end{bmatrix}$$

- Remark

The transpose of the incidence matrix of a Ferrers relation is the Ferrers relation of another partition

this is called the dual partition

This is extremely important.

If F = Ferrers relation of a partition:

$$\underline{\lambda} = (\lambda_1, \lambda_2, \dots), \quad \lambda_1 \geq \lambda_2 \geq \dots, \quad \sum_i \lambda_i = n$$

then the transpose matrix F^* is the Ferrers relation of a partition $\underline{\lambda}^*$, called the dual partition.

- Exercise 4.2

Given $\underline{\lambda}^* = (\lambda_1^*, \lambda_2^*, \dots), \ \lambda_1^* \geq \lambda_2^* \geq \dots, \ \sum_j \lambda_j^* = n$

Express λ_j^* in terms of λ_i

This is a crisis you all must go through.
Some things I can not show you. You have to do it yourself.

- Again, there is no easy formula for the partitions of an integer.
Next time, I will show you a famous formula, due to Euler.
But it is anything but trivial.
Very hard.
Unbelievable.

But let's digress.
What do you do in math when you find a hard problem?
You try to make it easy. Right?

Since it is so hard to find the number of partitions of an integer n, let's change the problem a little bit.

Compositions of an Integer n

(by the way, there is also the notion of composition of a set, but we'll discuss that later.)

Compositions of n are partitions of integer n, where the order of the summands matter.

A linearly ordered set of positive integers, whose sum equals n.

For example: $n=3$

partitions		compositions
3	$\{3\}$	3
2 + 1	$\{2,1\}$	2 + 1
1 + 1 + 1	$\{1,1,1\}$	1 + 2
		1 + 1 + 1

these have different linear orders.

Now, we can answer the question:

How many compositions of the integer n into k summands are there?

Answer: $\binom{n-1}{k-1}$

That's easy.
See how easy things get when you linearly order them?
That's always the case.

"When things are tough, order things linearly."

Proof

$\bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet\ \bullet$ n dots

Place stoppers between the dots to delineate the ordered summands.

Ex:

$\bullet\ \bullet \,|\, \bullet\ \bullet\ \bullet \,|\, \bullet \,|\, \bullet\ \bullet$ $n=8$

$2 + 3 + 1 + 2$

With n dots, there are $n-1$ positions to place stoppers.
To get k summands, you need to use $k-1$ stoppers.
So we have a set of $n-1$ positions and a set of $k-1$ stoppers to put in these positions.

$$\binom{n-1}{k-1}$$

But, we haven't solved our problem.
Revisiting our original problem:

We know that a partition has a <u>type</u>:

$$\{|B| : B \in \pi, \pi \in \Pi[S]\} \longleftarrow \sum_{B \in \pi} |B| = n = |S|$$

The type of a partition is a partition of the integer n.

How many partitions are there of a given type?

Example: Given the type $\{3, 2, 2, 1\}$, how many partitions are there of a set $|S| = 8$ that have this type?

etc.

I want to do this with equivalence relations.

<u>Theorem</u>

The number of partitions of S, with $|S| = n$, of type $1^{r_1} 2^{r_2} \cdots$ equals:

$$\frac{n!}{\underbrace{(1!)^{r_1} r_1!}\,\underbrace{(2!)^{r_2} r_2!}\,\underbrace{(3!)^{r_3} r_3!} \cdots}$$

This is the famous formula for the number of partitions of a set with a given type.

We will prove this formula next time.
Then we'll say a few extra things about enumerative facts about partitions.
Then we'll go back to relations and finish w/ the algebra of relations.
Then we start a major chapter — namely, <u>matching theory</u>.

This is a central chapter in combinatorics.

Basic enumeration (cont'd)

What we are seeing now is to be considered extremely elementary material.
If you think this is hard, "you ain't seen nothing yet."

Last time, we stated, w/o proof, the following facts:

Given S = finite set, $|S| = n$

We are studying the family of all partitions of S.

$$\Pi[S]$$

And we have seen that:

Bell numbers B_n = how many partitions there are of set S.

Stirling numbers of the second kind $S(n,k)$ = how many partitions there are of set S with k blocks.

aka the differences of zero, if you are British.

Now, we are going to determine:

how many partitions there are of set S with a given type

Recall that the type of a partition π is the multiset:

$$\{|B| : B \in \pi\}$$

this notation for multisets imitates the notation for sets.
It should be kept in mind that some of the entries in the multiset may be multiple, as we discussed.

The type of a partition π is a partition of the integer n.
(see [p4.9-10])

as we said last time, it is unfortunate that the term partition is used in 2 completely different senses.

- The type is denoted in one of 2 ways [p 4.11] (There are no established names for these ways)

(1) $\lambda_1 \geq \lambda_2 \geq \ldots$

where $\lambda_i > 0$ and $\sum_i \lambda_i = n$

The λ_i are $\underbrace{\text{the sizes of the blocks}}_{\text{the } |B| \text{, the elements of the multiset}}$ of the partition, in non-decreasing order.

This is associated w/ a <u>Ferrers relation</u>, which I will write the British way:

$$\begin{array}{ll} \lambda_1 & \bullet\ \bullet\ \bullet\ \bullet\ \bullet \leftarrow \lambda_1 \text{ dots} \\ \lambda_2 & \bullet\ \bullet\ \bullet \leftarrow \lambda_2 \text{ dots} \end{array}$$

Dots stand for the ones of the incidence matrix of the relation.

(2) $1^{r_1} 2^{r_2} 3^{r_3} \ldots$

Where r_i is the multiplicity i (i.e., r_i is the number of elements i in the multiset).

In other words, r_1 is the number of blocks of the partition having 1 element.
r_2 " " " " " " " " " 2 " .

etc.

- <u>Exercise 5.1</u>

Derive the relationship between λ_i and r_i

- We stated last time, w/o proof, that:

The number of partitions of the set S of type $1^{r_1} 2^{r_2} 3^{r_3} \ldots$ $=$ $\underbrace{\dfrac{n!}{(1!)^{r_1} r_1! \, (2!)^{r_2} r_2! \, \ldots}}_{\text{unfortunately, these numbers don't have a name.}}$ (*)

- If you add all these numbers over all types, you get the Bell numbers:

$$\boxed{\underbrace{B_n}_{\substack{\#\text{ partitions} \\ \text{of set } S}} = \underbrace{\sum_{r_1+r_2+\cdots+r_n=n}}_{\text{sum over all types}} \frac{n!}{(1!)^{r_1} r_1!\,(2!)^{r_2} r_2!\,\cdots}}$$

This is a fantastic identity, which is impossible to derive, unless you know where it comes from.

- If you add all these numbers over all types having k blocks, you get the Stirling numbers of the second kind:

$$\boxed{\underbrace{S(n,k)}_{\substack{\#\text{ partitions of} \\ \text{set } S \text{ having} \\ k \text{ blocks}}} = \underbrace{\sum_{\substack{r_1+r_2+\cdots+r_n=n \\ \#(r_i\neq 0)=k}}}_{\substack{\text{sum over all types} \\ \text{having } k \text{ blocks}}} \frac{n!}{(1!)^{r_1} r_1!\,(2!)^{r_2} r_2!\,\cdots}}$$

- Identity (*) can be established by handwaving, but I'd rather establish it by more rigorous methods.
In order to lead up this identity, lets digress on:

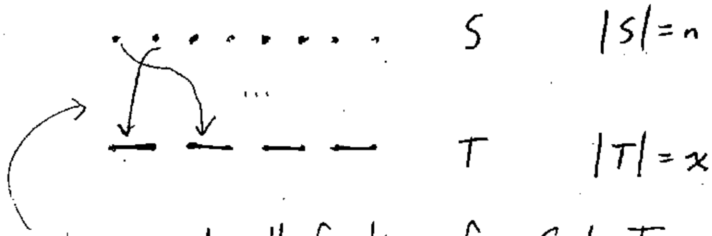
The Twelvefold Way

this term was used by Richard P. Stanley, when he took this course back in 1967.
This course has changed a lot since then, but I've decided to keep the twelvefold way.
Joel Spencer took the course in 1963 and the term "twelvefold way" is attributed to him.

The 1st time I taught this course was 1963.
The 2nd time was 1967.

(for those of you who have taken 18.313 – Probability, this will be familiar)

We have :



We consider all functions from S to T :

T^S denotes all functions from S to T

A function is, after all, a relation [p 2.4], so you can consider the function as a graph.

There are infinitely many ways of interpreting the concept of a function, depending on what business you are in.
Let's consider 3 interpretations :

(1) Distribution
(2) Occupancy
(3) Search

This is the typical situation where the same mathematical concept is given different psychological senses.

From these psychological senses, we get completely different problems.

(1) Distribution interpretation of a function

set S = set of balls
T = set of boxes

function = disposition (the way of placing) of the balls into the boxes

From the distribution point of view, one question that we can ask is that of occupation numbers :

$$\Theta_t \; : t \in T$$

Occupation Number

Θ_t is the number of balls that end up in the box labelled t

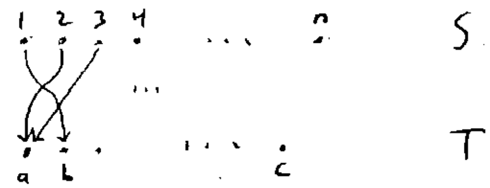
In some cases, we label elements of $S \Rightarrow 1, 2, \ldots, n$.
And elements of $T \Rightarrow 1, 2, \ldots, x$.

In some cases, we label things completely different.

(2) <u>Occupancy interpretation of a function</u>

S = viewed as a <u>linearly ordered set of places</u>

T = <u>alphabet</u>



another British custom ⇒ to label $a, b, \ldots, c$ from the 19^{th} century

in the occupancy interpretation of a function,

$1 \to b$ ⇒ letter b placed in position 1

$2 \to a$ ⇒ a 2

$3 \to a$ a 3

⋮ ⋮

function = word

Example from above:

function = baa ...

Note that this is <u>mathematically identical</u> to the distribution interpretation, but <u>psychologically, completely different</u>.

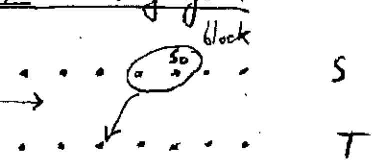
(3) <u>Search interpretation of a function</u>

This comes from <u>information theory</u> {which is extremely important nowadays, for reasons which will become clear.}

The <u>devil chooses</u> an $\underbrace{\text{element of } S}_{s_0}$ <u>w/o telling you.</u>

T = answers

function = questions



You ask a question and the devil has to give you the <u>correct answer</u>. That means, the devil <u>has to give you</u> the <u>block</u> w/in which <u>the element chosen by the devil</u> lies, when you ask <u>the appropriate question.</u>

So the whole idea of Information Theory (or the theory of search) is that you dispose of certain questions that are restricted by the problem at hand and you try to determine the element chosen by the devil effectively.

(we will discuss this later in greater detail)
(see 18.313 Probability SuperClass 4 notes [4/24/98.1-13])

- I wish I could give you 12 different interpretations of a function.
If we knock our heads together, we can come up w/ 20 different interpretations of the same concept of function.

- Let's go back to the first interpretation - that of distribution.
Then we can ask the question:

How many functions are there w/ given occupation numbers?

The number of functions w/ occupation numbers $\theta_1, \theta_2, \ldots \theta_x$

$$= \begin{cases} \dfrac{n!}{\theta_1!\,\theta_2!\,\cdots\,\theta_x!} & \text{if } \theta_1+\theta_2+\cdots+\theta_x = n \\ 0 & \text{otherwise} \end{cases}$$

{we imagine elements of T are labelled $1, 2, \ldots, x$}

[i.e., if the occupation numbers don't add up to the number of balls, there's no way.
This is called the Principle of Conservation of Balls]

⊘

- First, let's give the wrong proof.
This is a very important mistake.
Write down this mistake.

First, I say 2 functions are equivalent if they have the same occupation numbers.
Then, I consider equivalence classes of functions.

$$\left(\frac{\#\text{ equivalent functions w/ occupation numbers } \theta_1, \theta_2, \ldots, \theta_x}{\#\text{ equivalence classes}} \right) \neq \#\text{ functions w/ occupation numbers } \theta_1, \theta_2, \ldots, \theta_x$$

if you try this, you don't get an integer.
So it's wrong.

To do this correctly, you have to introduce <u>another notion</u>
And again, this proves to be the tip of an iceberg.

- <u>Disposition</u>

  Again, we can give <u>different interpretations</u> of <u>disposition</u>.
  We can consider disposition from the point of view of:

  (1) distribution

  – or –

  (2) occupancy

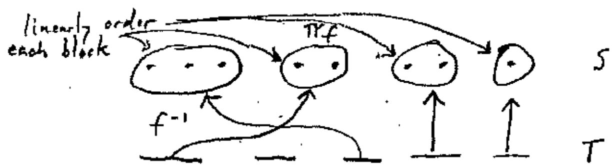
- First, a handwaving definition, from the point of view of <u>distribution</u>:

  Disposition = placement of the balls into the boxes and, <u>after</u> you place the balls into the boxes, you look at the balls in each box and you linearly order them.

  ↑ { So 2 dispositions are different if the linear order of some box or other is different, even though the occupation numbers may be the same. }

- An <u>occupancy</u> interpretation is simple to give:

  Disposition = take the <u>kernel</u> of the function, which is a partition [p 3.7], and on each block of the kernel, you put a <u>linear order</u>.



  Kernel of f, π_f, is the <u>partition</u> of S whose blocks are the sets: $f^{-1}(b)$, $b \in T$

  whenever $|f^{-1}(b)| \neq \emptyset$

- <u>Disposition</u> $\triangleq$ A <u>function</u>, together <u>with</u> a <u>linear order</u> on <u>each block</u> of its <u>kernel</u>.

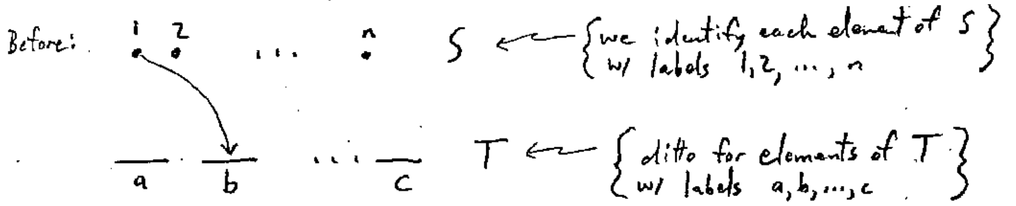
- The <u>number of dispositions</u> from S to T equals:

$$\underbrace{x(x+1)(x+2)\cdots(x+n-1)}_{n \text{ terms}}$$

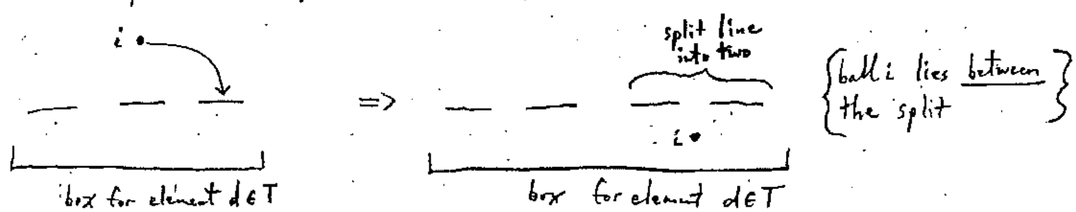
$$x^{(n)} \quad -\text{or}- \quad \langle x \rangle_n \longleftarrow \text{Read "Bracket } x\ n\text{"}$$

These are 2 <u>notations</u> that are <u>customary</u>.

<u>Proof</u> Proof by picture.
As often happens in combinatorics, there is a proof that you carry out by drawing a picture.

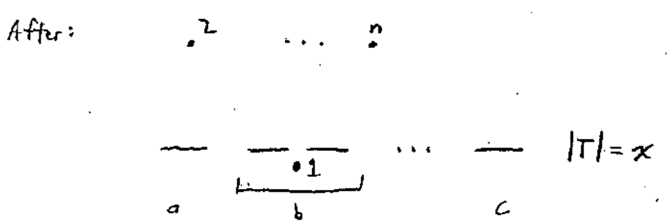


Now, we place the balls into the boxes so that the elements (i.e., balls) within each box are linearly ordered.
Whenever a ball is placed in a box, the line it is placed onto is cut into two.



Then the next ball placed in this box d can either go before or after ball i.

Thus, after placing ball 1, from the example above, we have:



The next ball placed can go into any of the $x-1$ boxes the ball was not placed in, or either of 2 places (either <u>before</u>, <u>or</u> <u>after</u>) in the same box the first ball was placed:

$$2^{\text{nd}} \text{ ball} \Rightarrow x+1$$

So, every time you place a ball, you <u>increase</u> the number of lines by 1.
And now you got it:

The number of dispositions of n balls into x boxes $= \overbrace{x(x+1)(x+2)\cdots(x+n-1)}^{n \text{ terms}}$

$= \langle x \rangle^n$

That's elementary.
Now, let's ask the question I really want to ask:

What is the <u>number of dispositions</u> of S <u>w/ given occupation numbers</u> $\theta_1, \theta_2, \ldots, \theta_x$?

(Say $\theta_i > 0$)

↑ we can assume, <u>WLOG</u>, that there are <u>no empty boxes</u>.

Answer: $n!$

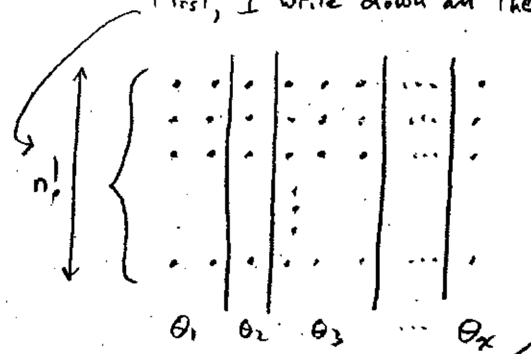
<u>No</u>, I did <u>not</u> make a <u>mistake.</u>
The number of dispositions w/ given occupation numbers is the <u>same</u> <u>irrespective</u> of the <u>assignment of the occupation numbers.</u>

This is an inherent, fundamental property.
Don't you ever forget this.
It creeps into all sorts of arguments.

So if you <u>assign the occupation numbers</u> $(\theta_1, \theta_2, \ldots, \theta_x)$, where you have n balls, and you want to count the <u>number of dispositions</u> w/ these given occupation numbers, it is <u>always the same</u>.

↑ the number of permutations of n (n = size of domain S)

<u>Proof</u> A nice combinatorial proof.
First, I write down all the <u>permutations</u> of S.



Then place the occupation numbers as stoppers.

The $n!$ permutations don't know where the occupation numbers are being placed.
Permutations don't think!

If you write all the permutations + place these stoppers, you get all the dispositions w/ these occupation numbers.
And <u>all</u> other occupation numbers.

Now, we return to counting functions w/ given occupation numbers.
We stated earlier:

- The number of functions w/ given occupation numbers $\theta_1, \theta_2, \dots, \theta_x$, where $\theta_1 + \theta_2 + \dots + \theta_x = n$, is:

$$\binom{n}{\theta_1, \theta_2, \dots, \theta_x} \leftarrow \text{multinomial coefficient} = \frac{n!}{\theta_1!\, \theta_2! \cdots \theta_x!}$$

Proof

Let's take the set of all dispositions of S into T and define an equivalence relation on dispositions.

Let d, d' be dispositions $\leftarrow$ {function from S to T (i.e., $[n] \to [x]$), Together w/ a linear order on the elements of each pre-image $d^{-1}(y)$, $y \in [x]$}

Define equivalence relation R s.t.

$$d \, R \, d'$$

when d and d' have the same occupation sets.

Example:

$d(1) = 2$ $\quad$ $d'(1) = 2$ $\quad$ but $d R d'$ since $d^{-1}(1) = d'^{-1}(1) = \{2, 3\}$

$d(2,3) = 1$ $\quad$ $d'(3,2) = 1$ $\quad$ $d^{-1}(2) = d'^{-1}(2) = \{1\}$

different linear orders

Then we have that:

An equivalence class of dispositions w/ the same occupation sets is a function, w/ occupation sets:

$$d^{-1}(1), \; d^{-1}(2), \dots, d^{-1}(x)$$

An equivalence class is a set of dispositions with the same occupation sets.
You take all the dispositions that have the same occupation sets.
What do these dispositions have in common?
They define the same function, because the order within an occupation set does not matter.

In other words:

An equivalent class corresponds to a unique function

For each possible vector of occupation sets $\left(d^{-1}(1), d^{-1}(2), \ldots, d^{-1}(x)\right)$, there is an equivalence class.

How many elements are there in an equivalence class?
Such an equivalence class has:

$$\theta_1!\ \theta_2!\ \cdots\ \theta_x! \quad \text{elements}$$

where $\theta_1 = |d^{-1}(1)|$, $\theta_2 = |d^{-1}(2)|$, $\ldots$, $\theta_x = |d^{-1}(x)|$

$\theta_1, \theta_2, \ldots, \theta_x$ are the <u>occupation numbers</u>

$\theta_1!\,\theta_2!\cdots\theta_x!$ elements, since we can have all possible permutations of elements in each set $d^{-1}(i)$.

Thus, <u>every</u> <u>equivalence class</u> is comprised of <u>a set of $\theta_1!\,\theta_2!\cdots\theta_x!$ dispositions</u>.

{unique vector of <u>occupation sets</u> $(d^{-1}(1), d^{-1}(2), \ldots, d^{-1}(x))$}

<u>Fix</u> the occupation numbers $\theta_1, \theta_2, \ldots, \theta_x$.

We've already shown that the <u>total number of dispositions</u> from S to T is [p 5.9]:

$$n!$$

And since <u>each equivalence class</u> is comprised of a <u>unique set</u> of $\theta_1!\,\theta_2!\cdots\theta_x!$ <u>dispositions</u>, we have:

$$\begin{array}{l}\text{number of } \underline{\text{equivalence classes}} \\ \text{w/ given occupation numbers} \\ \theta_1, \theta_2, \ldots, \theta_x\end{array} = \frac{n!}{\theta_1!\,\theta_2!\cdots\theta_x!} \begin{array}{l}\leftarrow \text{total \# dispositions} \\ \leftarrow \frac{\text{\# dispositions}}{\text{equivalence class}}\end{array}$$

And, since each equivalence class corresponds to a unique function:

$$\boxed{\begin{array}{l}\text{number of } \underline{\text{functions}} \\ \text{w/ given occupation numbers} \\ \theta_1, \theta_2, \ldots, \theta_x\end{array} = \frac{n!}{\theta_1!\,\theta_2!\cdots\theta_x!}}$$

- A remark about permutations of S

$n!$ = number of permutations of S ← $|S| = n$

Let w = a permutation of S ← $\{w: S \to S$, where function w is both one-to-one and onto.$\}$

ex: $w = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 4 & 2 \end{pmatrix}$

Using permutation w, we can define an equivalence relation.
We define an equivalence relation on S by setting:

$$s R_w s' \text{ iff } s w^i s'$$

↑ for some power i of the permutation w

$$w^i = \underbrace{w \circ w \circ \ldots \circ w}_{i}$$

(Remember, a permutation is a relation from S onto itself $w \subseteq S \times S$. Therefore, it can be composed w/ itself.)

So if s can be mapped to s' for some power i of permutation w, we say s is equivalence related to s'.

Equivalence classes are cycles of permutation w.

Example: $w = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$

$$\underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}}_{w} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}}_{w^2} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}}_{w^3} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}}_{w}$$

$$\text{Equivalence Class of } w = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \right\}$$

Thus, the cycles of a permutation define a partition of S, i.e., the underlying partition π_w of w.

As we discussed [p 2.11], the equivalence classes of an equivalence relation define a partition.

The blocks of the permutation π_w are the transitivity classes.

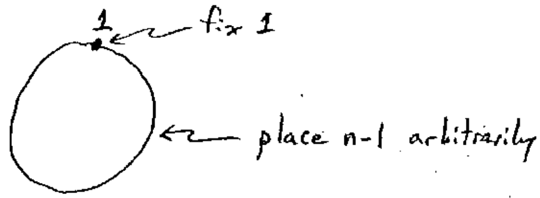
A permutation is cyclic if $\pi_w = \hat{1}$ ← (i.e., whenever the underlying partition π_w is the trivial partition $(\hat{1})$ with only one block.)

- How many cyclic permutations are there on a set w/ n elements?

$$(n-1)!$$

Because they are cyclic, you have to go round + round.
You can think of the elements disposed on the vertices of an n-gon.
So the number of different cyclic permutations is the same as the number of <u>different ways</u> of placing elements $\{1,\dots,n\}$ on the vertices of an n-gon. <u>independently of rotation</u>

So, you might as well fix the element 1,
and then place the remaining n-1 elements arbitrarily.

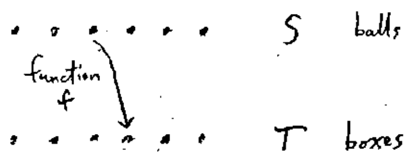


$\therefore$ $(n-1)!$ = number of different cyclic permutations on a set w/ n elements

---

- Now, we can go back to <u>The Twelvefold Way</u>

You have 2 sets S and T and a function f. You want to count the number of <u>inequivalent</u> functions



Why "Twelvefold"?

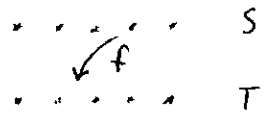
f function	S balls	T boxes
arbitrary (no restriction)	distinguishable	distinguishable
monomorphism (one-to-one)	indistinguishable	indistinguishable
epimorphism (onto)		

$$3 \times 2 \times 2 = 12$$

The "a,b,c's" of combinatorics is to learn to count using all these possibilities. We've already encountered most of them, so it's only a matter of identifying psychologically which is which.

Exercise 5.2

Verify the following portion of the Twelvefold Way table:



elements of S	elements of T	function f	number of inequivalent functions
distinguishable	distinguishable	arbitrary	x^n
		mono	$(x)_n$
		epi	$x!\,S(n,x)$

($S(n,x)$ is a Stirling number of the 2nd kind)

The Twelvefold Way (concluded)

We'll touch very briefly on this topic because it's covered in Professor Stanley's book. I want to mention that many of these topics in combinatorics are covered in Professor Stanley's book, which is very well written and readily available. And, therefore, I will not deal in this course, with any topic that is covered in Professor Stanley's book. This course is <u>disjoint</u> from Professor Stanley's book - except for definitions. There is no point in my wasting your time lecturing on stuff you can read in a well written book.

I assume you are reading this book (Enumerative Combinatorics, Cambridge University Press) on the side - for fun. Some of the things I say assume, tacitly, that you are familiar with certain things in Stanley's book.

The only thing in Stanley's book that we will cover in this course is the Twelvefold Way - largely for sentimental reasons.

- <u>The Twelvefold Way</u> is simply a list of enumerations of objects into functions:

S=balls • • • ... • $|S| = n$

T=boxes — — — ... — $|T| = x$

As usual, you have a set S of balls and a set T of boxes. Then you have functions from S to T.

\# cases:

balls	boxes	functions
{distinguishable, indistinguishable}	{distinguishable, indistinguishable}	{arbitrary, mono (1-1), epi (onto)}
2 ×	2 ×	3 = 12

the Twelvefold Way

Let's examine some of these:

i) function <u>arbitrary</u>:

balls	boxes	# inequivalent functions	
distinguishable	distinguishable	x^n	← a form of a definition of x^n
indistinguishable	distinguishable	$\left\langle {x \atop n} \right\rangle$	(Bose Einstein statistics)

↑ it's nice to have a fancy way of saying something fairly obvious.

You are putting indistinguishable balls into distinguishable boxes: what does that mean?
It means that all the data are the occupation numbers of the boxes.
Every box has a certain number of checks which correspond to the number of indistinguishable balls we put into that box. And the number of checks must add up to n.
That's what it means to put indistinguishable balls into distinguishable boxes.
So, you have the set T and you place n checks in the elements of T.

What does this mean?
You are taking a multiset out of the set T, as previously defined.
Placing n indistinguishable balls into x distinguishable boxes is just a fanciful way of saying that you are taking a multiset of size n out of a set of size x.

And everybody knows how many there are:

$$\left\langle {x \atop n} \right\rangle = \frac{\langle x \rangle_n}{n!} = \frac{x(x+1)\cdots(x+n-1)}{n!}$$

Exercise 6.1

Prove that the number of inequivalent ways of taking a multiset of size n out of a set of size $x = \left\langle {x \atop n} \right\rangle$

If you don't know this, prove it as an exercise.

ii) function mono:

balls	boxes	# inequivalent functions
distinguishable	distinguishable	$(x)_n \longleftarrow x(x-1)\cdots(x-n+1)$
indistinguishable	distinguishable	$\binom{x}{n}$ (Fermi – Dirac statistics)

That is interesting. Again, you are putting indistinguishable balls into distinguishable boxes. But every box can have at most one ball.
What does that mean?
That's just a fanciful way of saying we have x boxes and we check n of them.
That's called the binomial coefficient, for the last 3,000 years.

iii) function epi:

balls	boxes	# inequivalent functions
distinguishable	distinguishable	$S(n,x)\,x!$

This means you are placing n distinguishable balls into x distinguishable boxes and every box is occupied.
We've already discussed this (e.g., Lecture 3).

Exercise 6.2

Double check that the # inequivalent functions of placing n distinguishable balls into x distinguishable boxes, where each box is occupied (i.e., epi function) is:

$$S(n,x)\,x!$$

Exercise 6.3

indistinguishable distinguishable $\boxed{}$

For an epi function, work out the # inequivalent ways of placing n indistinguishable balls into x distinguishable boxes, where each box is occupied, is:

$$\left\langle \begin{matrix} x \\ n-x \end{matrix} \right\rangle$$

Now, let's make the boxes indistinguishable.
The only case that is interesting is epi. The other ones are trivial.

function epi:

balls	boxes	# inequivalent functions
distinguishable	indistinguishable	$S(n,x)$

This means, essentially, you are taking partitions of the balls, as the boxes are indistinguishable.
Number of partitions of a set of n elements into x blocks.

Exercise 6.4:

Double check above.

indistinguishable indistinguishable $\boxed{}$

This corresponds to partitions of a number.
Partition of integer x into n parts, as previously discussed.

That's pretty much the table. I suggest you draw a table for yourself and study it by yourself.
Now – why did I do this silly stuff?

I want to state the <u>Central Problem of Enumeration</u>

Fortunately, solved – but, unfortunately, never well written on anywhere.
Therefore, I assign it to you as a starred problem.
Every year I teach this course I tell myself I will rewrite that.
I've never done it.

It's a very interesting problem. It will take a very nice research paper to write a treatement of this problem – elegantly, of course. There are many inelegant treatements of this problem in the literature.

★ Exercise 6.5

Write up the <u>central problem of enumeration</u>, elegantly.

Let's put it, first, <u>informally</u>.

$\bullet \ \bullet \ \bullet \ \cdots \ \bullet \quad S \quad n$

$— \ — \ — \ \cdots \ — \quad T \quad x$

The balls are of different colors. Two balls of the same color are indistinguishable.
The boxes are of different shapes. Two boxes of the same shape are indistinguishable.
There are n balls and x boxes.

How many ways are there of placing the balls into the boxes? ⟸ that's the central problem of enumeration

Let me restate this <u>rigorously</u>:

Consider functions $f \in T^S$

given a partition π of S and ⟵ colors
a partition π' of T ⟵ shapes

How can we say that two functions are the same if they place balls of the same color into boxes of the same shape?
We say that in a toilet trained way.

We say that:

$f R f'$ whenever $\omega'^{j} \circ f \circ \omega^{i} = f'$
for some pair of permutations ω and ω', whose underlying partitions are π and π'

equivalent

{ω and ω' are not constrained to only cyclic permutations.}

That means if you permute the balls according to the permutation ω and then you permute the boxes according to the permutation ω', you get f'.

$f R f'$

this gives you an equivalence relation among functions

The Central Problem of Enumeration is the problem of counting the number of equivalence classes.

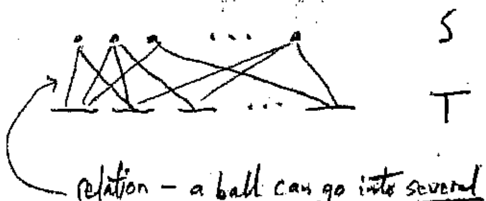
Let's jazz this up and make it a 3 starred problem:

Exercise 6.6 ∗∗∗

Develope a similar theory for relations.

This is hard. If you want to work on this problem, see me. There are papers on this and you shouldn't be working in a vacuum. I'll give you the references.

There are really 2 steps.
First of all, you are given S and T.
Then you consider relations between them.



relation – a ball can go into several boxes

It is very easy to count all relations. Trivial.
But, it is very hard to count all relations with given marginals

Count relations $R \subseteq S \times T$ with $\underbrace{\text{given marginals}}$

given # edges issuing from each ball and
edges coming into each box.
Then want to count the number of relations with these 2 numbers.

This is extremely difficult.

If you want to make this even tougher, then you make the balls partially distinguishable and the boxes partially distinguishable — like we did for a function.
You put a partition on the balls, and a partition on the boxes.
Then you define an equivalence relation, just as we defined for a function.
Except the equivalence relation is among relations.
Then you count the number of equivalence classes.

No one has ever done that. But it would be very nice if you did it.

I promised to tell you the easiest part of this problem, which is the Theorem of Gale-Ryser.

Theorem of Gale-Ryser

The necessary and sufficient conditions on marginals so that there exists a relation with these numbers.

you can't just give any numbers and expect relations to exist.
There are very subtle necessary and sufficient conditions, which were discovered rather late in the game.
You will see that. We don't have the instruments yet.

So there you are. Here's some work for you.
Don't just solve problems. I should give you only unsolved problems.
After all, this is a graduate course. What are we here for? The Mickey Mouse stuff?

That's the end of enumeration. Let's go back to pure combinatorics.
This course will oscillate between one chapter in pure combinatorics and one chapter in enumerative combinatorics.

Back to Relations

Let's consider relations of a set with itself, for simplicity.

$$R \subseteq S \times S$$

As we have seen, the set of relations is a Boolean algebra, because where there are relations there are also sets.

This Boolean algebra is endowed with an additional operation:

Composition: $R \circ R'$

Composition of relations can be visualized in many ways. Two ways:

1) analog, for relations, to composition of functions.
2) a relation is the combinatorial analog of a matrix and composition is the combinatorial analog of the product of two matrices.

This is, perhaps, more fruitful.
This will seem weird, but trust that we will gradually make it more palatable.

We'll make this last statement more precise. For the moment, just take it as it is.

** Exercise 6.7

We say that two relations commute if:

$$R \circ R' = R' \circ R$$

Find easy necessary and sufficient conditions for two relations to commute.

Probably hopeless, but I'll assign it to you anyway.
It would be nice if there were such things.

A equivalence relation R is: $R \subseteq S \times S$

reflexive $R \supseteq I$

symmetric $R = R^{-1}$

transitive $R \circ R \subseteq R$

If R is an equivalence relation, then associate to R:

$$\pi_R \in \Pi[S]$$

partition

partition of S into equivalence classes

This is one of the most primitive notions of mankind.
Balls are identified by color. Two balls with the same color are in the same equivalence class.

We will also see that partitions are the kernels of functions.

Given $f: S \to T$, the kernel of f is a partition of S

The kernel of f is defined as the equivalence relation R_f:

$$a R_f b \iff f(a) = f(b)$$

iff

The equivalence classes are the blocks.

Very nice. Now you rember ~~what~~ we said. One of the interpretations of the notion of function is in search theory, Information theory, Where f is computed as a question and T is the set of answers,
And the devil is thinking of an element of S, which you try to guess by asking the question.
The devil has to answer exactly which block (what color) the unknown element is and which block of the Kernel of f the unknown element lies in.
So, you have to ask, in general, several questions.

We look at partitions from the point of view of information theory (i.e., partitions as Kernels of functions). We are led to ask certain questions about them,

Suppose we have two questions (f, f').
They would have two Kernels (π, π').

$$\begin{array}{ccl} f & f' & \text{questions} \\ \uparrow & \uparrow & \\ \pi & \pi' & \text{Kernels} \end{array}$$

If you ask both questions, you get the meet of two partitions (the intersection of all the possible blocks). You get finer, more information.

Now, let's ask the following question about questions:

When is it that the answer to question f gives you no information whatsoever to the answer to question f'?

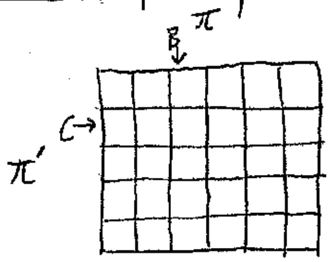
Is there a condition on the Kernels π + π' that ensures that the answer to one question bears no relevance whatsoever to the answer to the second question?
(I would not have asked this question if the answer was not positive)

We say that partitions π + π' are independent when, for every block B and C:

$$B \in \pi \ , \ C \in \pi' \ , \ B \cap C \neq \emptyset$$

This is the toilet trained way of saying that the two questions are completely independent.
Why?
Because if any two blocks meet, say the devil has chosen an element of the block B, it can be in any of the blocks of π', because every block of π' meets with block B. So you have no information whatsoever.

To visualize independent partitions like this:



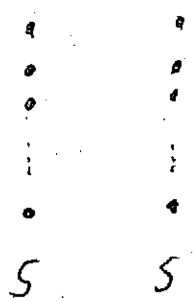
Exercise 6.8

Independent partitions can always be represented that way.
Make that precise, then prove it.

This is an extremely important concept – independent partitions.
It's made stronger in probability, where you have the concept of stochastic independence.

From the point of view of relations, how do you write the fact that they are independent?

R = equivalence relation



How do we visualize this?

Remember [2.9] that we defined the universal relation on a set B.
All possible pairs. An element of B is connected with everything else.

$$U_B = B \times B$$

An equivalence relation somehow comes from piecing together universal relations of your blocks.

we have to define what we mean by this.

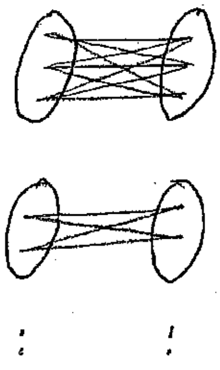
I'm sure you are wondering what I am up to.
I am up to something.

Given: Relations R_B on disjoint sets $B \in \pi$ ← partition
then define the <u>disjoint sum</u>:

$$R = \bigoplus_{B \in \pi} R_B \quad \text{is the unique relation s.t. } R\big|_B = R_B$$

restricted to B
(we haven't defined this yet)

- If R is an <u>equivalence relation</u>:



$$\underbrace{R = \bigoplus_{B \in \pi_R} U_B}$$

This is a triviality. The disjoint sum of the universal relation within each block.

- Observe that if π and π' are independent partitions then:

$$R_\pi \circ R_{\pi'} = U_S$$

- Exercise 6.9

  Prove the preceding.

Upon learning of this concept of independent relations, you are tempted to say "this is a universal concept that occurs everywhere in nature."
But nature is more sophisticated than that. It's almost the universal concept
It's not true that independent relations occur in nature.
What occurs in nature is a slight variant of independent relations, which we are now going to study in some detail.

- Universal concept - exactly when is it that two <u>equivalence relations</u> <u>commute</u>

That's the universal concept.
If you don't like it, go to church and complain.
I didn't invent the world. That's the way it is. I just tell you what is true.

We will study pairs of commuting equivalence relations.
I will sadistically withhold all examples until the end.

$$R \circ R' = R' \circ R$$

Who would ever think that this weirdo is what you find everywhere?

Notice that the concept is stated in terms of <u>equivalence relations</u>, <u>not</u> in terms of <u>partitions</u>.
It's a little more complicated to state in terms of partitions. But we're leading up to that. Stating the above in terms of the underlying partitions.

Let's first prove the following proposition:

<u>Proposition</u>

Two equivalence relations R and R' <u>commute</u> iff $\underbrace{R \circ R'}_{\text{Composition}}$ is an equivalence relation.

$R \circ R' = R' \circ R$

Proof:

1. If $R \circ R'$ is an equivalence relation:

$$\begin{aligned} R \circ R' &= (R \circ R')^{-1} && \text{symmetric property of equivalence relations} \\ &= R'^{-1} \circ R^{-1} && \text{valid for any two relations} \\ &= R' \circ R && \text{assuming } R \text{ and } R' \text{ are equivalence relations, symmetric property gives:} \end{aligned}$$

$$R'^{-1} = R'$$
$$R^{-1} = R$$

Hence, equivalence relation R commutes with equivalence relation R'.

We'll do the converse next time.

- I gave the wrong definition for Conditional Disjunction earlier [1.3].
I told you there were times that exercises were assigned in 1951. And I didn't do it.
Since then, I've had a hangup about this problem. Therefore, every time I state this problem, I make a mistake.

When I realized that the exercise I had assigned was wrong, I pulled out of my files Professor Church's original paper. And, for the first time in my life, I invented it. Sure enough, there's a mistake in it. So I ended my hangup 15 minutes ago. I should have done that a long time ago.

It's not a big error. At any rate, what I did was confuse 2 different concepts. There are two different ternary operations among sets. Both of which are used by people.

- The one I gave you is called not Conditional Disjunction – as I said, I was wrong. It's called the median. That's what I assigned to you [1.4 Exercise 1.2].

That's due to Birkhoff. I looked up Birkhoff's paper, which I happen to have because I inherited all of his papers when he died, and – sure enough – there was the median and a whole set of axioms for the median. You define a distributive lattice of sets by axioms on the median alone.

- Conditional Disjunction is a different operation. I'm not sure where Church got it from. But it was never used much. I remember who invented it. It was Post. The great American logician Post had classified all possible sets of Boolean functions, which can be used to generate Boolean algebra. An incredible tour de force. Among these is Conditional Disjunction.

- Post was a very real man. I'll explain the connotation.
All his life he taught at City College in New York. He taught like 16 hrs/week.
City College in New York, at that time, was where all the poor, brilliant students in the city of New York went. There were no scholarships to go to MIT. They had to go to City College. It was a very intense experience at that time, in the 1930's, 40's, + 50's, to be an undergraduate in City College in N.Y.

I was once invited to give a lecture at the National Academy of Sciences. It was with a group of logicians. So I mentioned the American logician Post. At the end of the lecture, 15 people came up and said "I took calculus with Post." I didn't know he was that good.

The collected papers of Post are a good sized volume. There's only one thing. Half of the volume is one paper, which he worked on for about 15 years. The paper is called "The Theory of Multi-groups." In this paper, Post tries to extend the notion of group to an n-ary operation. He did this cleverly. An n-ary operation with which you can generalize all the basic concepts of group theory. He wrote up the paper and, just before he went to press, he discovered that his n-ary operation could be reduced to a binary operation. So he put a little footnote to that effect. Because of that, it's not right, but the content of Post's solution is still correct.

- Median

$$(A, B, C) = (A \cap B) \cup (A \cap C) \cup (B \cap C)$$

an easy computation shows:

$$= (A \cup B) \cap (A \cup C) \cap (B \cup C)$$

Given median, ϕ, $\hat{I}$ (universal set), Birkhoff shows how to define $\underbrace{\cup, \cap, \text{ and the distributive law.}}_{\text{distributive lattices}}$

Birkhoff axioms for the median include:

$$(\phi, A, \hat{I}) = A$$
$$(A, B, A) = A$$
$$(A, B, C) = (C, A, B) = (B, C, A) \longleftarrow \text{cyclical permutations are equal.}$$
$$((A, B, C), D, E) = ((A, D, E), B, (C, D, E))$$

- Exercise 7.1

State correctly and prove that the median defines $\cup$, $\cap$, and the distributive law.

- Conditional Disjunction (due to Alonzo Church, from Alabama. One of the greatest logicians of all time.)

$$[A, B, C] = (B^c \cap A) \cup (B \cap C)$$

Church wrote in terms of logic, instead of in terms of sets.
For example:

if C True $\Rightarrow$ B True
if A True $\Rightarrow$ B False

Conditional Disjunction, ϕ, and $\hat{I}$ generate Boolean Algebra.

$$B^C = [\hat{1}, B, \phi]$$

$\cup$ and $\cap$ are an easy matter.

Unfortunately, he doesn't state the axioms.

* Exercise 7.2

Construct a system of axioms for conditional disjunction, analogous to what Birkhoff did for the median.

It would be nice to have such a system of axioms.
I don't know if anyone has ever done this.

<u>What's coming</u>

1. Commuting equivalence relations
2. The "pointless" point of view
3. The language of order and lattices

<u>Commuting equivalence relations (cont'd)</u>

Given: R, R' = equivalence relations.
Associate corresponding partitions π and π'.

π and π' are independent when
for every $B \in \pi$ and
for every $C \in \pi'$
we have $B \cap C \neq \phi$

To visualize this concept, it's convenient to take the following special case.

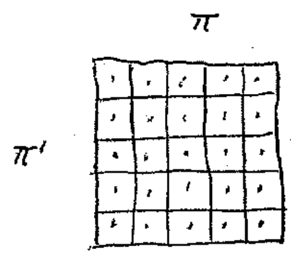
Example:

Say $|B \cap C| = 1$ for every $B \in \pi$ and every $C \in \pi'$

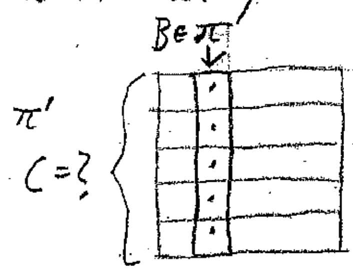
Thus, every element of S belongs to exactly one pair of blocks $B \in \pi$ and $C \in \pi'$

Hence, we can code S as $\{(B,C) : B \in \pi, C \in \pi'\}$

The set of these pairs is isomorphic to S.
This means that every element of S has 2 coordinates.
You have the π coordinate and the π' coordinate.



Asking a π question and asking a π' question are independent questions.
The answer to the first question gives you no information whatsoever as to what block of the second partition the element chosen by the devil lies in.



- If R and R' are independent, then $R \circ R' = U_S = R' \circ R$

  Two independent equivalence relations commute.
  Trivial.
  We were then embarking on finding a structure theorem for commuting equivalence relations, in general.

  It is tempting to say this is the only example where equivalence relations commute (i.e., when the relations are independent), but it's not true.

Let's finish the Proposition we started last time [6.11]:

- Proposition:

Two equivalence relations R and R' commute iff $R \circ R'$ is an equivalence relation.

We proved one direction last time.
Now for the second half of the proof.

Proof:

2. If equivalence relations R and R' commute:

$$NTS \quad R \circ R' = R' \circ R \text{ is symmetric, reflexive, transitive.}$$

Reflexive - trivial

Symmetric

$$(R \circ R')^{-1} \stackrel{?}{=} R \circ R'$$

$$R'^{-1} \circ R^{-1} =$$

{and since R, R' are equivalence relations, $R = R^{-1}$, $R' = R'^{-1}$}

$$R' \circ R =$$

{since it is given that $R \circ R' = R' \circ R$}

$$R \circ R' = R \circ R'$$

Transitive

$$(R \circ R'^{-1}) \circ (R \circ R'^{-1}) \stackrel{?}{\subseteq} R \circ R'$$

{Composition is an associative operation - a fact I should have observed before.}

$$R \circ \underbrace{R'^{-1} \circ R}_{} \circ R'^{-1} \subseteq$$

{given that $R \circ R'^{-1} = R'^{-1} \circ R$}

$$\underbrace{R \circ R}_{R} \circ \underbrace{R'^{-1} \circ R'^{-1}}_{R'^{-1}} \subseteq$$

$$R \circ R'^{-1} \subseteq R \circ R'^{-1}$$

QED

R^{-1} should be R' in Transitivity proof

Now we are almost ready to classify a structure theorem for pairs of commuting equivalence relations.
Let's see if you can guess it.

If you have two independent equivalence relations, they commute.
Trivial.

Last time we discussed the disjoint sum of equivalence relations.
Let's review.
What I'm going to say is that the disjoint sum of independent equivalence relations will also give you pairs of commuting equivalence relations.

If R_B and R_B' are independent equivalence relations on the set B, then: $\oplus$ is the disjoint sum

$$\bigoplus_{B \in \pi} R_B = R \quad \text{and} \quad \bigoplus_{B \in \pi} R_B' = R' \quad \text{commute}$$

Why?
You have disjoint blocks:

Each disjoint block has 2 independent equivalent relations in there. They don't interfer with each other.
They commute.
Trivial.

So an easy way of constructing pairs of commuting equivalence relations is to take pairs of disjoint sums of independent equivalence relations.

That's very easy.
The surprising thing is that the converse of this is true.
Which brings us to Mme. Dubreil's Theorem.

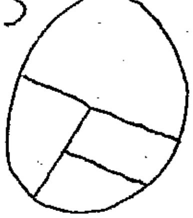
Mme. Dubreil's Theorem

Two equivalence relations R and R' commute iff they are disjoint sums of independent equivalence relations.

This is a very famous result, which unfortunately has not found its way into very many books.
She tried to develope the foundations of mathematics based on the theory of relations. Unfortunately it didn't work. Nothing wrong with it. Sorry.
In the meantime, she got this nice theorem.

We have 2 commuting equivalence relations on a set S.

S



Then we partition S in such a way that we restrict the pair of equivalence relations in each of these blocks.
Then, such a restriction on each block is a pair of independent equivalence relations on that block.

I know what you're thinking. You're thinking – "how weird." But, as soon as you see the example, which, as I've said, I am sadistically withholding, you'll see it's not weird at all.

To repeat: Any 2 equivalence relations commute iff there is a partition of S into blocks in such a way that if you restrict them to each block, then they become independent.

i.e., join or rather meet

Proof:
We have just seen, of course, that one part of the Theorem is immediate.
Now, suppose we have 2 commuting equivalence relations:

$$R \circ R' = R' \circ R$$

By the preceding Proposition [7.5], we know that $R \circ R'$ is an equivalence relation.
Observe that:

$$\left.\begin{aligned} R \circ (R \circ R') &= (R \circ R) \circ R' \\ &= R \circ R' \end{aligned}\right\} \qquad \begin{aligned} R \circ (R \circ R') &= \underset{\text{associative}}{R \circ (R' \circ R)} \\ &= \underbrace{(R \circ R)}_{\text{reflexive}} \circ R' \\ &= R \circ R' \end{aligned}$$

Hence, each block of π_R is contained in a block of $\pi_{R \circ R'}$.

By symmetry, each block of $\pi_{R'}$ is contained in a block of $\pi_{R \circ R'}$.

So we can restrict to the blocks of $R \circ R'$.
Therefore, we can assume, without loss of generality, that there is <u>only one block.</u>
So we consider only 1 block at a time.
But, if there is only 1 block, they are independent.

$$\left.\begin{array}{l} \text{Say } R \circ R' = U_S \\ \text{Then } R \text{ and } R' \text{ are commuting equivalence relations,} \\ \qquad \text{their composition is } U_S \end{array}\right\} \text{a moments meditation shows that they are independent.}$$

So, in general, you take the blocks of $R \circ R'$ and restrict parts of R and R' to the blocks and apply this observation. ~~Until~~ you get one of them. unclear!

Now you say "will you at last give us an example?"
<u>Glaring Example</u>

V = vector space.
Pick your favorite vector space. In this course, we take only vector spaces of the real numbers. But, if you wish, you can take a vector space over any field.

W = subspace of [redacted] V

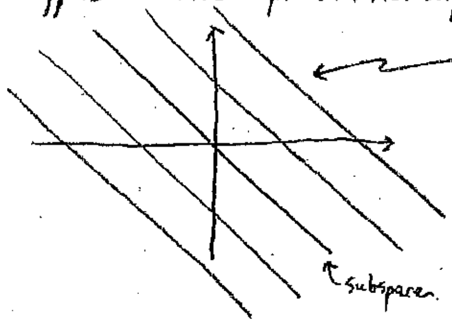
Given a <u>subspace</u> of a <u>vector space</u>, can <u>define</u> an <u>equivalence relation</u>.

Define an equivalence relation R_W, as follows, on the set V.

$x, y \in V$

Say $x R_W y \underset{\text{iff}}{\Longleftrightarrow} x - y \in W$

What do the equivalence classes look like?
Suppose we have a plane. And suppose our subspace is a line.



Now, let's suppose we have another subspace W'.

If W' is also a subspace of V, then the equivalence relations R_W and $R_{W'}$ commute.

Too bad this stuff isn't in any book. Once you know it, you can think differently.

Isn't that nice.
Any vector space, take the set of all subspaces, any two of them will give you a pair of commuting equivalence relations.
What more do you want? There you have it. They're all over the place.

Proof:

Let $W'' = \text{span}(W, W')$ ← subspace spanned by elements of W and elements of W'

$$= \{w + w' : w \in W, w' \in W'\}$$

It turns out that:

$$\underbrace{R_W \circ R_{W'}}_{\text{composition of equivalence relations}} = \underbrace{R_{W''}}_{\text{equivalence relation of span}}$$

Suppose that:

$$x R_W \circ R_{W'} y \quad \text{for } x, y \in V$$

Want to show that:

$$x - y \in W''$$

So, let's unscramble what $x R_W \circ R_{W'} y$ is like.
It means, by definition of composition of relations:

There is a $z \in V$ s.t. $x - z \in W$ and $z - y \in W'$

This gives:

$$\left.\begin{array}{l} x - z = w \in W \\ z - y = w' \in W' \end{array}\right\} \text{ if you add these two, you get: } x - y = w + w' \in W''$$

And, vica versa ~~if~~ you ~~take~~ any element of W/W', you can get $x-y$ by the process. And ~~therefore~~, the conclusion holds.

So, it is interesting that the set of all subspaces of a vector space is a fountain in that any two of them defines commuting equivalence relations. That's very important. It was very late to be recognized.

Conversely suppose $x-y \in W'' = W + W'$ so $x-y = w + w'$, then $W \ni x-(w'+y) = w$ and $(x-w)-y = w' \in W'$ where we may take $z = x-w = y+w'$ so $x R_W z$ and $z R_{W'} y$ and thus $x R_W \circ R_{W'} y$. //

You are not expected to do any two or three star problems.
But you are expected to do one one star problem and 1/3 of the unstarred problems.
If you do a two star or three star problem, you are excused from any more duties in the course. Some two star problems and three star problems I will assign are very interesting and challenging.

I shall put my jacket back on. I can't lecture without my jacket.
For the kind of tuition you pay, it is very professional to wear a jacket. All the time. Most of the time.

JNG: Content before form.
GCR: Form matters too. Form gives backbone to content.
When you have no content, you fall back on form.
I hope we have some content today.

- Commuting equivalence relations (conclusion) {Then we'll start on the "pointless" point of view.}

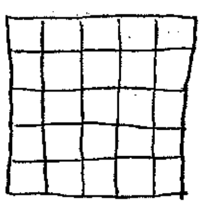
Last time, we proved Mme. Dubreil's theorem, which I now summarize by picture.

First of all we begin to systematically confuse the notions of <u>equivalence relations and partitions</u>. We interchange these, as they are <u>cryptomorphic</u>.

Two partitions are independent if the blocks of one are one way and the blocks of the other are the other way:

π

π'



The blocks of each partition can be used as coordinates for the intersection, assuming that the intersection has one block.

Two equivalence relations commute iff the underlying set S can be written as the disjoint sum of several blocks such that if you restrict the two relations to any one of these blocks, you get two independent relations.

The only way to get two commuting equivalence relations is to take a disjoint sum of independent equivalence relations.

- ** Exercise 8.1

Find a search theoretic meaning for two commuting equivalence relations.

I would really appreciate it if someone worked this out.
Like all important problems, it is not clearly stated.

The problem is very simple – almost infantile.
If we have 2 independent equivalence relations, then there is an obvious search theoretic meaning; which we have discussed [6.8]. You are asking 2 independent questions.

There has to be a search theoretic meaning to 2 commuting equivalence relations.
But people in Course 6 don't know about commuting equivalence relations and, therefore, they haven't worried about it.
There has to be a meaning to this in terms of information theory.
Once you discover that, people will prove all sorts of things.

This problem is made all the more interesting because, as we began to see last time, pairs of commuting equivalence relations are a dime a dozen.
Last time we saw the classical example [7.8–10], which I will remind you of.

Glaring example of ~~[illegible]~~ commuting equivalence relations

V = vector space

W, W' = subspaces of V

Given vectors $x, y \in V$, can define the relation:

$$x R_W y \iff x - y \in W$$

It is immediate that this equivalence relation has an obvious geometric meaning.
You take parallel subspaces and those are the equivalence classes.
That's what parallel means, rigorously.

We verified last time:

$$R_W \circ R_{W'} = R_{\mathrm{span}(W, W')}$$

But this doesn't depend on the order of W and W'.
If you don't see that, I can't explain it.

$$= R_{W'} \circ R_W$$

Therefore the equivalence relations commute.

Any 2 subspaces define 2 commuting equivalence relations.
This took a long time to sink in.

- Another Example:

This time the example is hard.
If you don't know the underlying math, take a nap.

G = group , H and H' are normal subgroups

imitating the preceeding example, we are going to define 2 equivalence relations.

For $x, y \in G$, set:

$$x R_H y \iff x y^{-1} \in H$$

Then:

$$R_H \circ R_{H'} = R_{H'} \circ R_H$$

This is what normal subgroups are all about.
So the family of all normal subgroups of a group are such that any 2 provide commuting equivalence relations.

If you work through the definitions, you find that you have an equivalence relation because H is a normal subgroup.

- Exercise 8.2

Prove the preceeding.

- Another Example:

Again, if you don't know the math, take a nap.

A = ring , I and I' are ideals

For $x, y \in A$, set:

$$x R_I y \iff x - y \in I$$

Then:

$$R_I \circ R_{I'} = R_{I'} \circ R_I$$

The ideals of a ring give you commuting equivalence relations.
So you begin to see that commuting equivalence relations have a deep relation with the coset structure of an algebraic system. This was brought out by the Russian mathematician Mulsev in a famous discovery of what went on in a general algebraic system that made this commutativity work.

That's the end of this chapter.
Now we begin the next chapter:

The "pointless" point of view

Let me motivate this w/ a few words on probability.
If you don't know probability, take another nap.
Say we have a function:

$$f : S \to T$$

This function has a kernel:

$$\pi_f = \text{Kernel} \quad (\text{This is a partition of } S)$$

In probability $\underline{S}$ becomes a <u>sample space</u> and
function $\underline{f}$ is called a <u>random variable</u>.

In a <u>discrete</u> sample space (a finite sample space, for example),
every random variable has a kernel. So you can visualize it
and ask information theoretic questions about it.

In the <u>continuous</u> case, for example you have a normally distributed random variable,
you don't have an obvious partition for the kernel. Yet, you'd like to talk about the
kernel of a random variable, even in this case. There is a partition, but the blocks
all have probability 0.

Therefore, you want to extend the notion of partition so that every random variable would
have a kernel in the extended notion.
What strategy should we follow in performing such an extension?
By the way, this is only of several extensions which are possible — and not only for
probability, but topology, algebraic geometry, what not.
In order to perform such extensions, we have to rephrase them in a way, which is
called <u>"pointless."</u>

↑ the word "pointless" is from von Neumann

Let's take the case of a partition.
We saw that there are <u>3 cryptomorphic concepts</u> going on here.

equivalence relation

partition — Complete Boolean subalgebra of sets

You have a set S. You get the complete Boolean subalgebra by taking the atoms (the minimal elements) you get a partition. And the the Boolean algebra is obtained by taking the unions of these minimal sets.

So every complete Boolean subalgebra determines a partition, and a partition determines an equivalence relation.

Now, we could have defined the notion of partition by starting with the notion of complete Boolean subalgebra. That would have been the "pointless" definition of a partition. Because the pointless definition uses only the definitions of the Boolean algebra and nothing about the points.

So the program of the pointless point of view is to redefine several basic concepts. In combinatorics, we don't use the underlying points of a finite set, with the idea of eventually generalizing it to one of the various branches of mathematics.

Let's see how this idea works in a number of cases.
For example, relations.

Relation - "pointlessly"

$$R \subseteq S \times T$$

Can we invent a pointless notion equivalent to this?
Take a subset A of S and define $R(A)$ as:

$$A \subseteq S, \quad R(A) = \{ b \in T : (a,b) \in R \text{ for some } a \in A \}$$

a well known concept of functions

And we have seen that:

$$R(A \cup B) = R(A) \cup R(B)$$

This gives us the lead to the pointless definition.
Let's take the Boolean algebra of subsets $P(S)$ and $P(T)$
Now, we take a mapping φ:

$$\varphi : P(S) \to P(T)$$

We say that φ is a hemimorphism whenever:

$$\varphi(A \cup B) = \varphi(A) \cup \varphi(B)$$

Or, more generally, since we are dealing with a complete Boolean algebras:

$$\varphi\left(\bigcup_i A_i\right) = \bigcup_i \varphi(A_i)$$

Claim: Every hemimorphism defines a relation.
It's almost trivial.

$R \subseteq S \times T$
Take any $a \in S$, $\varphi(a) \subseteq T$.
Then all pairs (a, b) for $b \in \varphi(a)$ shall belong to R.

$$R(A) = \bigcup_{a \in A} R(a) = \bigcup_{a \in A} \varphi(a) = \varphi(A)$$

This is easy because we can take arbitrary unions + intersections.

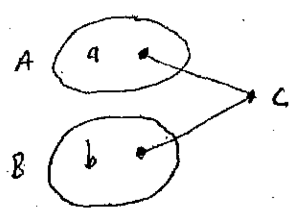
Therefore, given a hemimorphism, it trivially defines a relation.
And this relation implements the hemimorphism.

Now we come to the interesting example.
Example
Given the function $f: S \to T$

of course $f(A \cup B) = f(A) \cup f(B)$

But, in general, $f(A \cap B) \neq f(A) \cap f(B)$

Proof by picture:



$$f(A \cap B) \neq f(A) \cap f(B)$$
ϕ $\qquad$ c

However, for the inverse function, we do have:

$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$ true of all relations

$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$

We also have that:

$f^{-1}(\phi) = \phi$

$f^{-1}(T) = S$ where the function is everywhere defined.

And this gives us the lead to <u>defining</u> a <u>function</u> <u>pointlessly</u>.
How do we define it?
Here's how we define it:

Suppose Ψ is a <u>homomorphism</u> of the Boolean algebra of subsets $P(T)$ into the Boolean algebra of subsets $P(S)$ when:

$$\Psi(A \cup B) = \Psi(A) \cup \Psi(B)$$

$$\Psi(A \cap B) = \Psi(A) \cap \Psi(B)$$

$$\Psi(A^c) = \Psi(A)^c$$

$$\Psi(\phi) = \phi$$

<u>Complete</u> means:

$$\Psi\left(\bigcup_i A_i\right) = \bigcup_i \Psi(A_i)$$

$$\Psi\left(\bigcap_i A_i\right) = \bigcap_i \Psi(A_i)$$

Now we have the following simple, but important theorem.
Whenever you have a homomorphism of the Boolean algebra $P(T)$ into the Boolean algebra $P(S)$, there is <u>always</u> a <u>function</u> that <u>implements</u> the <u>homomorphism</u> in this way.

The equations indicate that the inverse of the function is always a homomorphism of Boolean algebras. So there is an inverse relation here between homomorphisms and Boolean algebra functions going the other way.
That's very important. Let's write that down as a theorem.

<u>Theorem</u>

Boolean algebras

Ψ is a complete homomorphism of $P(T)$ to $P(S)$
iff $\Psi = f^{-1}$ for some function $f: S \to T$

That is the <u>pointless version</u> of a <u>function.</u>

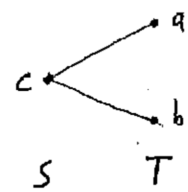
We will generalize the notion of function by taking homomorphisms.
What will we do?
We'll drop the word "complete". We'll just take homomorphisms.
And we'll get continuous functions, random variables, etc.
Do you get the idea?

This is the kind of theorem that, once stated, is almost trivial.
I can give you a proof by gestures.

Proof (by gesture)

Ψ is a hemimorphism

Therefore Ψ is implemented by a relation from T to S



But this relation can not have this diagram
Because otherwise, the intersection can not work.
You can't have 2 elements in T to the same element in S.
Otherwise, you get:

$$\underbrace{\Psi(A \cap B)}_{\emptyset} \neq \underbrace{\Psi(A) \cap \Psi(B)}_{c}$$

Therefore, you can't have this.
That means everything in T goes only to one place.
That's called a function in my book.
All you have to check is that the function is everywhere defined.
And that comes from the fact that it preserves complements.

Exercise 8.3

Write down this proof in all detail.

So, here we have 2 examples of the "pointless" rendering of concepts:

relation corresponds to a hemimorphism

function has an inverse correspondence to a homomorphism of Boolean algebras.

Now you say – "Sure, that's easy. What about something more complicated?
For example, 2 independent equivalence relations."

- 2 independent equivalence relations – "pointlessly"

We have independent partitions π and $\pi' \in \Pi[S]$

Then B_π and $B_{\pi'}$ = the corresponding Boolean subalgebras

What properties of the Boolean subalgebras B_π and $B_{\pi'}$ are equivalent to the partitions being independent?

Easy:

Independence is equivalent to the following "pointless" property of B_π and $B_{\pi'}$:

For every $A \in B_\pi$, $B \in B_{\pi'}$ s.t. $A \neq \emptyset$, $B \neq \emptyset$,

we have $A \cap B \neq \emptyset$

This means that it's not only true for the blocks, but also any union of blocks. And you can convince yourself of the equivalence.

- Exercise 8.4

Prove the preceeding property.

Now we come to the tough one.
This was an open problem that was solved by Catherine Yan in 1995 in her PhD Thesis.

- 2 partitions which correspond with commuting equivalence relations – "pointlessly"

Let π and π' be commuting partitions.

Again, we have the corresponding Boolean subalgebras B_π, $B_{\pi'}$

Theorem Yan (1995)

π and π' commute iff B_π and $B_{\pi'}$ satisfy the following condition:

whenever $A \in B_\pi$, $B \in B_{\pi'}$ s.t. $A \cap B = \emptyset$,

there exists $C \in (B_\pi \cap B_{\pi'})$ s.t. $A \subseteq C$ and $B \subseteq C^c$

- *Exercise 8.5

Prove Yan's Theorem when the underlying set is finite.

I think this should be out any day. If you crib from the paper, it should take a couple of weeks.

Now you are thinking – Why on earth should we worry about giving pointless definitions,
Why? What is that good for?
Let's see what that's good for.
We're going to have an excursion into probability.
The point of "pointless" definitions. One point – there are many points.
There are many ways of justifying it. If you're a topologist, you justify it one way.
If you're an algebraic geometricist, you justify it another way.
We'll use probability.

Measure

If sets S, T finite, $|S| \in \mathbb{R}$, $|T| \in \mathbb{R}$;
one axiomatically characterizes the number of elements as:

$$|S \cup T| + |S \cap T| = |S| + |T|$$

If you don't see that, I can't help you. I can, but I won't.

Furthermore:

$$|\phi| = 0$$

Mathematicians have abstracted the notion of measure from this property of number of elements.
If you want the continuous analog of the number of elements, you introduce the concept of a measure.

Let $\mathcal{L}$ be a family of subsets of a set S (not complete) closed under $\cup$, $\cap$, and containing ϕ.

Note that complements do not play a role.
This is not a Boolean algebra.

This is also called Distributive lattice of sets

A measure μ on $\mathcal{L}$ is a function from $\mathcal{L}$ to $\mathbb{R}$ (positive or negative) s.t.
for $A, B \in \mathcal{L}$ we have:

1. $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$
2. $\mu(\phi) = 0$

Exercise 8.6
Show by an example that the 2nd property doesn't follow from the 1st.

Exercise 8.7

Show that every measure satisfies the inclusion-exclusion formula, Namely, for $A_i \in \mathcal{L}$,

$$\mu(A_1 \cup \dots \cup A_n) = \sum_{i=1}^{n} \mu(A_i) - \sum_{i<j} \mu(A_i \cap A_j) + \sum_{i<j<k} \mu(A_i \cap A_j \cap A_k) - \dots$$

How do measures connect up w/ the "pointless" point of view?
Time's almost up.
We'll continue this next time.

The point of the "pointless" point of view

This is largely cultural.
We'll discuss how the stuff we've been doing relates to other branches of mathematics.
We said, last time, we have:

set S

$\mathcal{L} \subseteq P(S)$

← family of subsets

$\mathcal{L}$ is closed under finite unions and intersections.

Such a family is called a distributive lattice of subsets

The point of distributive lattice of subsets is that they are used to define measures.
A measure is, in general, not defined on all the subsets of a set, because if the set is empty, there are too many.
So you take a suitable family — a distributive lattice of subsets — and that is what you use to define a measure.

$\phi \in \mathcal{L}$

A measure on $\mathcal{L}$ is a function $\mu : \mathcal{L} \to \mathbb{R}$ satisfying:

1. $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$
2. $\mu(\phi) = 0$

Sometimes measure is known as a valuation, especially by Geometers.
If you look at my book on Geometric Probability, measures are sometimes called valuations, following custom from geometry.

We will have occasion to study some remarkable measures that arise in combinatorics.
The most famous of all measures is NOT the number of — as you think — but the measure which is one of the fundamental concepts of mathematics, which is called the Euler characteristic. We'll study this in great detail.

I assigned you, last time, as an exercise, the fact that every measure satisfies the inclusion-exclusion formula.

Since we're finishing this 1st chapter of our course, why don't you start doing the exercises maybe due next Monday.
You have to the problems. Otherwise, you don't learn.
$\frac{1}{3}$ of the exercises and 1 starred exercise in the term.
And, of course, I might examine your notes.

If, in addition, $\phi \in \mathcal{L}$ and $\mu(A)$ lies between 0 and 1, for every $A \in \mathcal{L}$, then μ is called a probability.

And you can extend it to complements:

And one can define, consistently:

$$\mu(A^c) = 1 - \mu(A)$$

Exercise 9.1

Prove the preceeding.

Therefore, we might as well assume that $\mathcal{L}$ is a Boolean subalgebra (not necessarily complete).

closed under finite unions and intersections.

Example – measure

S = any infinite set

There's a famous Boolean subalgebra of any infinite set.
Of course, there's the Boolean algebra of all subsets.
But, there's another one:

We say that $A \subseteq S$ is cofinite when A^c is finite.

The family of all finite and cofinite sets of S is a Boolean algebra $\mathcal{L}_{fin}$

finite

This should be obvious to you.
The $\cap$ of two cofinite sets is cofinite.
The $\cup$ of a finite and a cofinite set is a cofinite set.
The $\cup$ of two finite sets is finite.
The complement of a finite set is cofinite.

On $\mathcal{L}_{fin}$ we define a measure μ as follows:

set $\mu(A) = 0$ if A is finite
$\mu(A) = 1$ if A is cofinite

This measure is very important in logic. Logicians use it all the time.

You can see that this measure has some pathological properties.

Example:

$$S = \mathbb{N}$$

$$\mu(\mathbb{N}) = 1$$, since $\mathbb{N}$ is cofinite

But:

$$\mu\left(\bigcup_{i=0}^{\infty} i\right) \neq \sum_{i=0}^{\infty} \mu(i)$$

Measure of the union is not the sum of the measures, even though the sets are disjoint.
So, it's not countably additive.
This measure won't do for the purposes of probability.

More generally, even though

$$\mu(A_1 \cup A_2 \cup \ldots \cup A_n) = \mu(A_1) + \mu(A_2) + \ldots + \mu(A_n)$$

whenever the A_i are disjoint, for any measure —

In other words, any measure is "finitely additive," as they say.
In fact, this is not true if you take infinite sets.
In fact, it is <u>never</u> true if you allow more than countable sets.

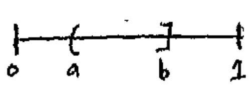
— it is <u>seldom true</u> that:

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_i \mu(A_i)$$

for A_i disjoint and I infinite.

This does not make any sense. That's why we have the "pointless" point of view.

<u>Example</u>



Take the interval $[0,1]$. Define a measure on $[0,1]$ to be the length of the interval:

$$\mu([a,b]) = b - a$$

There is a theorem of measure theory, which I don't want to state or prove, which says that this measure extends to lots of other sets.

However, note that:

$$\mu((a,b]) = \mu\Big(\underbrace{\bigcup_{a<p\le b} p}_{\substack{\text{the interval is the}\\ \text{union of all the points}\\ \text{between } a \text{ and } b}}\Big) \not= \sum_{a<p\le b} \mu(p)$$

$$\underbrace{b-a \neq 0}$$

We get the classical contradiction.
Take the whole interval and we would get $1=0$. I call that contradiction.

Therefore, the equality above can not be true.
However, it's partly true.

The equality is true when we allow only countable unions of the interval.
That's where probability generalizes combinatorics.

We say that $\mathcal{L}$ is a Boolean σ-algebra of sets when $\mathcal{L}$ is a Boolean algebra and:

whenever $A_1, A_2, \ldots \in \mathcal{L}$ disjoint, we have:

$$A_1 \cup A_2 \cup \ldots \in \mathcal{L}$$

i.e., the union of a countable number of disjoint elements of $\mathcal{L}$ belongs to $\mathcal{L}$.

If so, then a measure μ is countably additive when

$$\mu(A_1 \cup A_2 \cup \ldots) = \mu(A_1) + \mu(A_2) + \ldots$$

For example, the measure defining the lengths extends to the Boolean σ-algebra of subsets generated by intervals.
That's a non-trivial result.
This gives you ordinary probability on the interval $[0,1]$.

In particular, the triple $S, \mathcal{L}, \mu$ where:

S = set

$\mathcal{L}$ = Boolean σ-subalgebra of subsets

μ = probability (i.e. a countably additive measure taking values between 0 and 1, including the extremes)

is called a <u>sample space.</u>

For example, the interval $[0,1]$, together with the Boolean σ-algebra generated by all intervals, is a sample space.
The Boolean σ-algebra generated by the interval $[0,1]$ is called, as you know, the Boolean σ-algebra of <u>Borel sets.</u>

Now, the point of the "pointless" point of view is that, in general, in probability, you want to <u>generalize</u> the idea of asking a question and getting an answer.
Just like we did for search theory.

But just having a partition of a sample space isn't enough.
Even in the simplest cases, because random variables, as some of you know, can be continuous.
There is a <u>substitute</u> for <u>partition.</u>
And that is sub Boolean σ-subalgebra of the Boolean σ-algebra.

The analog of a partition is a sub Boolean σ-algebra.
Sub Boolean σ-algebras are generalizations of partitions. (in some sense, sub Boolean σ-algebras are more important than partitions.)

That's the point!
You don't take an arbitrary complete Boolean algebra. Because a complete Boolean subalgebra would determine a partition, as we've seen in one of the earlier theorems we've proved here.
You take a Boolean σ-subalgebra. That will determine a partition - however, because of the "pointless" point of view you think about it as if it determined a partition. If you have any partitions, you rewrite it pointlessly and generalize it to Boolean σ-algebras.
Thereby obtaining the <u>probabilistic analog.</u>

In general, if you have a Boolean σ-algebra and a measure on it, it's very hard not to make it countably additive.
↑ you have to go out of your way

Now, something marvellous happens for Boolean σ-algebras.
If you take a set. Let's say the set is finite.
Then you take partitions. Then the transposition of partions is complicated.
Because we have to classify partitions according to their types.
Two partitions would be equivalent, in the certain sense we have defined, if they have the same type - same number of blocks of 1 element, 2 element, etc.

For Boolean σ-algebra, a marvellous thing happens, which we'll prove later.
They are all isomorphic. So you don't have to worry about type and all that.
This is the famous theorem of von Neumann.
Assuming you don't have any pieces of minimal measure (they're non-atomic), then they're all isomorphic.
It's a marvellous theorem for which there is no simple proof.

Any 2 non-atomic Boolean σ-subalgebras of a Boolean subalgebra are isomorphic.

Intuitively, it should be obvious. You cut into 2, cut into 2, etc.
Then you piece together again.

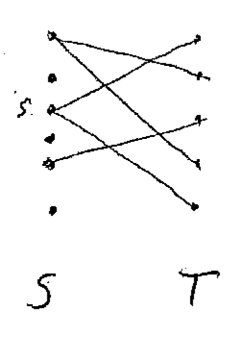
We'll talk about it later.

This is one point of the "pointless" point of view.
I wanted to get this far to show you how the "pointless" point of view gets to apply.

You get a partition and you rewrite in a Boolean σ-subalgebra, which has no obvious relation to a partition. Nonetheless, by transfering from the language of partitions to the language of Boolean σ-algebras, you are able to go to the probabilistic case.
If you are careful, you can extend the notions of dependent, independent, and commuting partitions to Boolean σ-algebras.

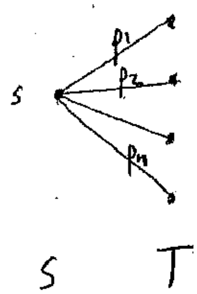
There is more we could say about the "pointless" point of view.
Let me conclude w/ another example of a probabilistic generalization of a notion we have already studied.
This is purely cultural.

We've been studying relations, which you can visualize as:



The probabilistic analog of a relation is a Markov chain.

Take point s. Take all the edges issuing from it. And to these edges, assign probabilities that add up to 1.
Intuitively, that means that s goes w/ one point w/ probability p_1, w/ another point w/ probability p_2, etc.



Similarly, with all points.
That gives you a Markov chain w/ transition probabilities.

So from relation, you go to Markov chain by putting probabilities on the edges, which add up to 1. (You can even put probabilities that don't add up to 1, because you can include sinks).
This is just an example of how you go from combinatorial to probabilistic.

The following is culture. You don't have to know. You can take a nap.
Historically, how did the notion of relation arise?

This is a very interesting lesson in mathematical history.
It shows something that happens again and again in mathematics.
The notions of mathematics arise, first, in their most complicated form.
Then they gradually get refined.

The notion of relation first arose in its most complicated form, which is this (if you don't know this, I won't explain it):

You take 2 algebraic varieties. You take the product of these, in the sense of algebraic geometry. Then you take a sub variety of the product.
That's what algebraic geometers call a correspondence.
This is a relation. It has this algebraic structure.

That's how they arose in the 19th century.
They couldn't think of a relation as purely a subset. They had to think of them in terms of equations. Things were not defined for them unless you gave them an equation.

There are some very deep theorems, like the Riemann-Brock theorem, which are about correspondences, which are the algebraic analog of the notion of relation.

I want to fill in a couple of odds + ends on the theory of relations before we leave and start in on the language of order, which is the next in this course.

A couple of things. I'd feel guilty if I didn't tell you this.
It's a very fundamental fact about a relation.

- Relation on a set to itself

$$R \subseteq S \times S$$

To this relation we have seen that we can associate an incidence matrix,
That's a good way of visualizing a relation. There are other good ways (an oriented graph, for example).

There is something kinky about the incidence matrix.
If you take the composition:

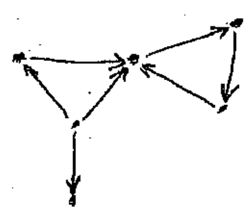
$$R \circ R$$

↑ that doesn't correspond to the product of the incidence matrices.
If you take the product of the incidence matrices, you're likely to get a matrix whose entries are no longer just 0's or 1's

That's cute. We'd like to do something about this kinkiness.
I'll tell you one now and one later when we do matroids.
One way is to define a new kind of incidence matrix (it's very natural):

- <u>Edge-vertex incidence matrix</u>

For this, we visualize the relation $R \subseteq S \times S$ as a graph:



Assume that R has <u>no</u> loops:
for all $a \in S$, $(a,a) \notin R$

Then, with this graph, we associate a matrix, as follows.
It is a matrix of 0, 1, and -1.

The edge-vertex incidence matrix of R:

$$\begin{array}{cc} & \text{edges} \\ \text{vertices} & \begin{bmatrix} \;\; a_{ij} \;\; \\ \boxed{} \end{bmatrix} \end{array} = M$$

0 = vertex doesn't belong to edge
-1 = vertex at <u>end</u> of arrow
$+1$ = vertex at <u>beginning</u> of arrow

There is no point in making up this matrix if it didn't have some remarkable property. And it does. It has the extraordinary property that it is <u>totally unimodular.</u>

We say that matrix $M = (a_{ij})$, $\begin{array}{l} 1 \le i \le k \text{ rows} \\ 1 \le j \le n \text{ columns} \end{array}$ is <u>totally unimodular</u> when every minor $= +1, -1,$ or 0.

Now you say – what a weirdo notion.
These matrices – there aren't lots of them.
And I say – no, no, no. They're all over the place. In fact, they are very popular.
They are a great subject of research right now.
Because they just discovered a couple of years ago that the <u>theory of the representation</u> of the <u>infinite symmetric group</u> is <u>intimately related</u> to the <u>theory of totally unimodular matrices.</u>
So they are very much in the news.

Now, you say, give me an example of a totally unimodular matrix.

<u>Theorem</u> (David Gale at UC Berkeley)

The edge-vertex incidence matrix of a graph (i.e., relation w/o loops) is <u>totally unimodular.</u>

<u>Exercise</u> 9.2

Prove the above.

This is a very difficult proof, actually, which is the result of successive simplifications. The original proof was massive.

So here we have an enormous class of totally unimodular matrices. They are all over the place.

- The question as to when a unimodular matrix is the edge-vertex incidence matrix of a graph has been solved, by one of the most outstanding graph theorists of all times, Tutte, in hair raising detail. This condition is very deep. One of the deepest theorems of combinatorics. Not fully understood to this day. You follow it line by line, but you really don't see why it should be true.

We have 3 minutes
I'll give you some problems.

- * Exercise 9.3 (Mme. Dubreil)

$R, R' \subseteq S \times S$ are sesquicommuting when:

$$R \circ R' \circ R = R' \circ R \circ R'$$

Find a structure theory for sesquicommuting.

In other words, what do they look like?

- Exercise 9.4 (Rignet)

We say that R is a Ferrer's relation when S is finite and can be ordered so that:

$$R(a_1) \supseteq R(a_2) \supseteq \ldots$$

The incidence matrix has lots of 1's in the first blocks, a subset of those 1's in the second, etc.

It turns out these relations can be characterized algebraically. Combinatorially.

Prove that R is a Ferrer's relation iff:

$$R \circ R^{c-1} \circ R \subseteq R$$

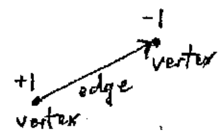
Very elegant.

John Guidi
guidi@math.mit.edu

Last Words before Order

• From last time, we saw that given a relation $R \subseteq S \times S$ w/o loops (i.e., an oriented graph), we can associate an edge–vertex incidence matrix M.

$$M = (a_{ij}), \quad i \in S, \; j \in R$$

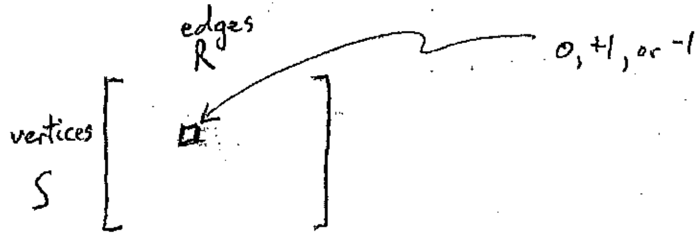


set $a_{ij} = 0$ if $i \notin j$

$a_{ij} = 1$ if $(i,a) = j$ for some $a \in S$
(i is at beginning of edge j)

$a_{ij} = -1$ if $(a,i) = j$ for some $a \in S$
(i is at end of edge j)

In this way, you obtain a matrix:



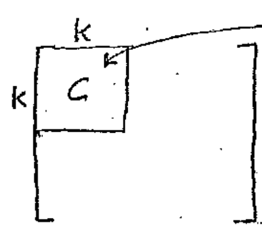
Of course, when writing down the matrix, you can linearly order the vertices and edges.

Theorem

The matrix M is totally unimodular.

That means that every minor of the matrix is equal to +1, -1, or 0.
Then you might ask what are totally unimodular matrices good for. We touched on this last time. But certainly this is a remarkable property.

Proof:



Take a minor
You may, w/o loss of generality take the first $k \times k$ submatrix.
A minor is always a square submatrix.

Need to show that the determinant of this submatrix is +1, -1, or 0.

So there are 3 cases.

Let me tell you a story.
John von Neumann – everytime he listened to a lecture and the lecturer said "And now there are 3 cases" he got up and left. He couldn't stand it.
Now I tell you "And now there are 3 cases."

Observe that since columns (edges) correspond to elements of R, each column contains all 0's except exactly one $+1$ and one -1.

Every column is an edge and every edge has a beginning and an end.
Remember – no self loops.

So, when we take this minor, there are 3 cases.

Case 1: Every column has <u>exactly two</u> non-zero entries

necessarily one $+1$, the other -1. And when summed, these cancel.

Hence, the sum of the rows, for each column vector, equals 0.

Hence $\det C = 0$.

Since the sum of the rows is the zero vector, they are <u>linearly dependent</u> and the determinant is 0.

Case 2: One column has all 0 entries.

Trivially, $\det C = 0$

Case 3: At least one column has exactly one non-zero entry.

$$C = \begin{bmatrix} \pm 1 & \cdots \\ 0 & \\ 0 & C' \\ \vdots & (\text{stuff}) \\ 0 & \end{bmatrix}$$

We can assume, wlog that it's the first column and row (if you don't see that, I can't explain it).

We compute $\det C$ by expanding by the first column.

$$\det C = (\pm 1) \cdot \underbrace{\det C'} + 0$$

all the other cofactors are 0.

C' is a smaller minor ($k-1 \times k-1$)
So we compute this by induction, until we get a 1×1 minor.
Then we back up.

$$\det C = \pm 1$$

What are totally unimodular matrices good for?

There are people who make their living on totally unimodular matrices.
But that's not a really honest answer.

Example:

M is a totally unimodular matrix

Say it's square and say $\det M \neq 0$

Then, consider the system of linear equations:

$$M\underline{x} = \underline{b}$$

since $\det M \neq 0$, $\underline{b}$ has a unique solution.

What happens when M is totally unimodular?

When M is totally unimodular, whenever $\underline{b}$ has integer entries, then the solution $\underline{x}$ has integer entries.

Why? I'll do this by hand
How do you solve?
By Cramer's Rule.
Cramer's Rule tells you that the solution $\underline{x}$ is obtained by taking M, replacing one of the columns by $\underline{b}$ and dividing it by the determinant.
When you expand the minors, the minors are all ± 1, so you get linear combinations of the entries of $\underline{b}$ and the determinant is ± 1, so you get an integer.

Ultimately, this is the reason people study totally unimodular matrices.
The field is called integer programming.

Exercise 10.1

The above example also extends to the case where $\det M = 0$.
Do this as an exercise.

In that case, you don't have a unique solution.
You can have a space of solutions.

• *Exercise 10.2*

Theorem (Ryser at Cal Tech)

Given a relation $R \subseteq S \times S$ $\quad |S| < \infty$ (i.e., S is finite)

relation R has some marginals

(sum of the rows and the columns of the incidence matrix [2.7])

A lot of work has gone into studying the set of all relations w/ given marginals.
Studying the structure of the set.
In particular, the question of how many there are is largely open.
But a number of results have been obtained.
Here is an elegant and useful result about the set of relations which are marginals.

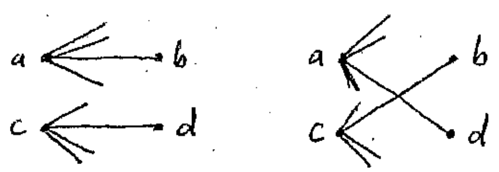
A switch of R is a relation R' obtained from R as follows:

Pick a pair $(a,b) \in R$ $\quad$ where $(a,d) \notin R$
$\qquad\qquad (c,d) \in R \qquad\qquad\quad (c,b) \notin R$

then:

$$R' = \left(R - \underbrace{(a,b) - (c,d)}_{\text{take away these edges}}\right) \underbrace{\cup\ (a,d) \cup (c,b)}_{\text{add these edges}} \quad \Bigg\}\ \text{switch}$$

Let's visualize:



$R \qquad\qquad R'$ switch

everything else (all other edges) remain the same.

From the point of view of incidence matrices:

$$\begin{array}{cc} & \begin{array}{cc} b & d \end{array} \\ \begin{array}{c} a \\ c \end{array} & \begin{pmatrix} \cdots 1 & 0 \cdots \\ 0 & 1 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \begin{array}{cc} b & d \end{array} \\ \begin{array}{c} a \\ c \end{array} & \begin{pmatrix} \cdots 0 & 1 \cdots \\ 1 & 0 \end{pmatrix} \end{array}$$

$R \qquad\qquad R'$ switch

If R' is a switch of R, then R and R' have the same marginals.
The sum of the rows and the sum of the columns remain the same. The marginals are unchanged.

The theorem is:

If R and R'' have the same marginals, then R'' may be obtained from R by a series of switches.

This result has found lots of applications.
The set of relations w/ given marginals is connected through switches.
There are many proofs of this theorem.
I don't know a snappy proof.
Please find a snappy, elegant proof. Not just any old proof.
There must be something really central.

This is just one result about marginals.
There are oodles of them.

The Language of Order

A partial order on a set P is a relation $R \subseteq P \times P$ with the following properties:

1. $R \supseteq I$ (reflexive) R contains the Identity relation
2. $R \cap R^{-1} = I$ (anti-symmetric)
3. $R \circ R \subseteq R$ (transitive)

An ordered relation is usually written w/ a different notation.
Instead of aRb, one writes $a \leq b$ "less than or equal" (interpreted differently according to which partial order is being considered)

In terms of $\leq$, the 3 properties of a partial order become:

1. $a \leq a$ (reflexive)
2. if $a \leq b$ and $b \leq a$ then $a = b$ (anti-symmetric)
3. if $a \leq b$ and $b \leq c$ then $a \leq c$ (transitive)

Now we have to go through all the language of partially ordered sets, so we can speak.

- <u>Covered relation</u>

For $a, b \in P$,
henceforth, when we see P, we mean a set endowed with a partial order. Automatically. Strictly, we should write:

$$(P, \leq)$$

we say $a \prec b$ (a is <u>covered by</u> b)
when:
$a < b$ (i.e., $a \leq b$ and $a \neq b$)
and $\underbrace{\text{if } a \leq c \leq b \text{ then either } c = a \text{ or } c = b}$

this is a polite way of saying that there is <u>nothing between</u> a and b.

We write $a \preceq b$ to mean $\underbrace{a \prec b}_{a \text{ is covered by } b}$ or $a = b$

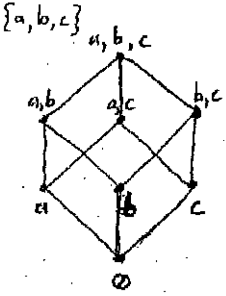
- If P is finite, the graph (necessarily oriented) of the covering relation is the Hasse diagram of P.
This graph is usually not written as an oriented graph, but from the top, down.

- $\left\{\begin{array}{l}\text{Partially ordered sets}\\ \text{Ordered sets}\\ \text{posets}\end{array}\right\}$ Synonyms

- <u>Example</u> – Hasse diagrams

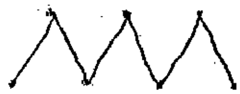
We give an example of a poset (partially ordered set) in terms of its Hasse diagram:

$P(\{a, b, c\})$ ← w/ apologies for the confusion, here P is the <u>Boolean algebra of subsets of the elements $\{a, b, c\}$.</u>
Powerset of $\{a, b, c\}$



- **Poset example** – fence Hasse diagram

Another example fence:



- **Example** – Relation

$R \subseteq S \times T$ defines a partial order on $P = S \cup T$ by setting:

$$a \leq b \text{ whenever } a \in S,\ b \in T, \text{ and } (a,b) \in R$$

Thus, every relation can be viewed as a partially ordered set.

$R \subseteq S \times T$

$a \in S$

$b \in T$

You have $a \leq b$ wherever there is an edge connecting a and b in the "balls into boxes" diagram.

- **Antichain**

Set with a trivial partial order

$$\underbrace{a \leq a}_{}$$

a is related to a only.
If a and b are different they are unrelated.
This satisfies the 3 conditions.

- **Chain** (or linearly ordered set)

A chain is a poset P where:

for all $a, b \in P$, we have $a \leq b$ or $b \leq a$

For example, $\mathbb{R}$ is a chain.
The set of real numbers is ordered.

A <u>finite chain</u> has a Hasse diagram:

- If $Q \subseteq (P, \leq)$ then Q inherits a partial order from P.
  Subset Q may be viewed as a partially ordered set in its own right.
  Because you can restrict the partial order to Q, generally, and satisfy the 3 conditions.

- A <u>maximal element</u> of $(P, \leq)$ is an element $x \in P$ s.t.
  if $y \geq x$ then $y = x$

- A <u>minimal element</u> of $(P, \leq)$ is an element $x \in P$ s.t.
  if $y \leq x$ then $y = x$

- The <u>dual</u> of $(P, \leq)$ is the set $(P^*, \leq)$ where:

  $$a \leq b \text{ in } P^* \text{ iff } b \leq a \text{ in } P$$

  Informally, you get P^* by turning P upside down.
  If P is finite, you literally turn the Hasse diagram upside down to get P^*.
  Of course, $P^{**} = P$. That's trivial.

"one cap"

- A <u>maximum</u> element (if any) is the <u>unique</u> maximal element, written as $\hat{1}$.
  A <u>minimum</u> element (if any) is the <u>unique</u> minimal element, written as $\hat{0}$.

Example: $\mathbb{R}$ has <u>no unique minimal</u> or <u>maximal</u> element.
So this poset has no minimum and no maximum.

Example: The Boolean algebra of subsets of 3 element sets [10.6] has:
the null set $\emptyset$ is the minimum element
the set $\{a, b, c\}$ is the maximum element

Example:  This Hasse diagram has no maximum element and no minimum element.
There is no <u>unique</u> maximal element and no <u>unique</u> minimal element.

John Guidi
guidi@math.mit.edu

The Language of Order (Cont'd)

We saw last time the definition of a partially ordered set.
All of the following are the same notion:

$$P = \text{partially ordered set} = \text{poset} = \text{ordered set}$$

These all refer to a set P and an ordered relation $\leq$.
We tend not to explicitly write the ordered relation and assume it, implicitly.

$$(P, \leq)$$

We have seen that if P is finite, we can associate w/ P a graph (namely a relation) which visualizes the covering relation in the partial order.

We have begun to list the various terms that are used in connection with partial orders:

- minimal element
- maximal element
- minimum element — the unique $\hat{0}$
- maximum element — the unique $\hat{1}$

For the poset represented by this Hasse diagram, there is no $\hat{0}$ or $\hat{1}$.

- antichain — trivial order
- chain — linear order

  a poset where any 2 elements are comparable. So you can visualize it as a linear chain, even though there may be continuous chains.
  Even a linear order is extremely complex. For example, you have transfinite ordering.

- the subset of a partially ordered set inherits the partial order.

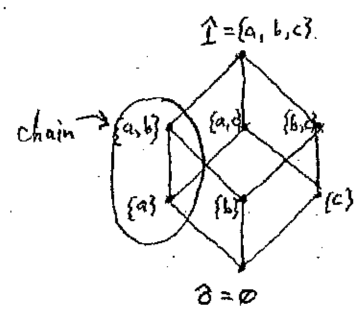
  $Q \subseteq P$

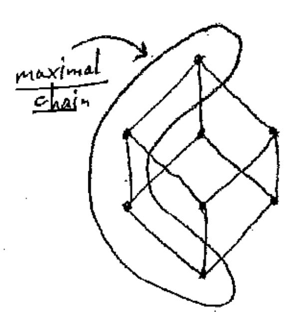
  We are particularly interested in subsets of partially ordered sets that are chains or antichains.

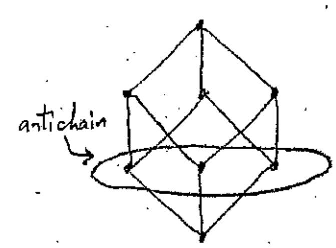
Let's next define:

maximal chain of P = flag = complete chain = saturated chain

For example, if you take the Boolean algebra of subsets of 3 elements, whose Hasse diagram we have seen [10.6], we have:







- **Rank**

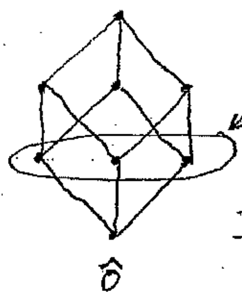
Suppose P has a $\hat{0}$.
We say P is ranked if, for all $x \in P$, all maximal chains from $\hat{0}$ to x have the same size (usually called length), say $r(x)+1$.

(in which case $r(x)$ is said to be the rank, so that: $r(\hat{0}) = 0$)

- **Atom:**

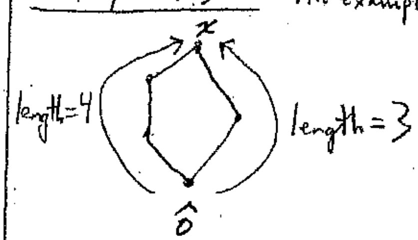
$x \in P$ is an atom when $\underbrace{x \succ \hat{0}}$

Atoms have rank 1. Always. x covers the minimum (i.e., unique minimal) element



It is easy to see that this partially ordered set is ranked.

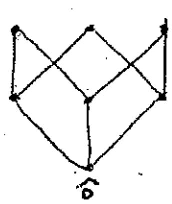
- **Example - N_5** An example of a partially ordered set that is not ranked.



Maximal chains from $\hat{0}$ to x have different lengths.

N_5 is not ranked.

- **Example - ranked poset without $\hat{1}$**

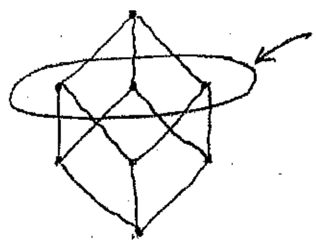


Ranked, as for all $x \in P$, all maximal chains from $\hat{0}$ to x have the same length.
Note that there is no $\hat{1}$.

Coatom

$x \in P$ is a coatom when $\underbrace{x \prec \hat{1}}_{x \text{ is covered by the maximum (i.e., \underline{unique} maximal) element}}$

$\hat{1}$

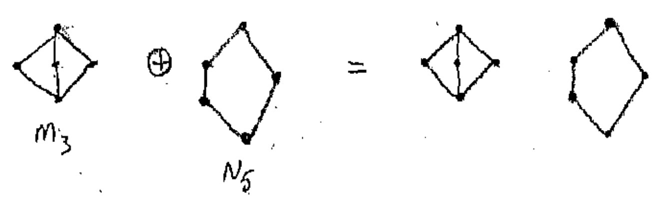


There are two basic operations on partially ordered sets:

- disjoint sums
- products

Disjoint Sum

$P \oplus Q$ = the disjoint sum of the sets P and Q, considered to be disjoint, and the partial order is the original partial order in each set.

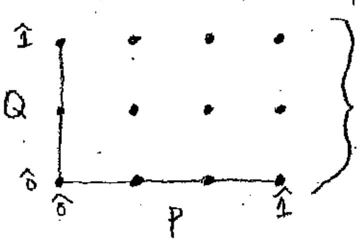


[redacted] Product

$$P \times Q = \{ (x,y) : x \in P,\ y \in Q \}$$

and $(x,y) \geq (x',y')$ iff $x \geq x'$ and $y \geq y'$

For example, let's take the product of 2 chains:



The product $P \times Q$ is simply the partially ordered set represented by the rectangle.

Warning: Disjoint sum and product are not the only two operations on partially ordered sets.
There are many, many others.
And they have never been completely classified.
For example, you can take the lexicographic product. You can put one poset on top of the other. You can stitch them together in various ways.
There are infinitely many ways of combining partially ordered sets with one another.

- **sup**

If $x, y \in P$, we say that $\sup(x,y)$ exists if there is an element $z \in P$ s.t.:

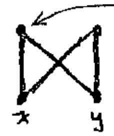
$z \geq x$ and $z \geq y$ and, furthermore,

every element u s.t. $u \geq x$ and $u \geq y$ must have $u \geq z$

In other words, if there is one element z above both x and y and, furthermore, anything else above both x and y is also above z, then $z = \sup(x,y)$.

[Mathematicians are so silly.
When they give a definition, very often what they should give is something that does not satisfy the definition.
Many times that is the way to understand the definition.
They should give you something that shows what the definition is meant to guard you against.
Nobody has learned this lesson.]

Examples of partially ordered sets where sup doesn't exist:



both candidates
Therefore, no $\sup(x,y)$

No $\sup(x,y)$

Example of poset where sup exists:

Given our friend the Boolean algebra of subsets of 3 elements [10.6], given any 2 subsets, if you take their union, that is their sup.

- **inf** (dually, one defines inf, if it exists)

If $x, y \in P$, we say that $\inf(x,y)$ exists if there is an element $a \in P$ s.t.:
$a \leq x$ and $a \leq y$ and, furthermore,
every element u s.t. $u \leq x$ and $u \leq y$ has $u \leq a$

Lattice

A partially ordered set L where $\sup(x,y)$ and $\inf(x,y)$ <u>always</u> exist for any 2 elements x and y is called a <u>lattice</u>.

A lattice is a poset w/ sups and infs all over the place.
For example, the following poset is <u>NOT</u> a lattice.

No $\sup(x,y)$

There was a great discovery in the 19^{th} century, by the German mathematician Dedekind, that the notion of a lattice can be axiomized algebraically.
You can see an equivalent definition of a lattice using an algebrazation of the two notions of sup and inf.
This was a tremendous step forward. Not uncontroversial, because Kronecker, who was a friend of Dedekind, until Dedekind published his first paper on lattices, said: "you've become so abstract, you're going crazy."

Dedekind algebraic axiomization of lattice

Let L be a set endowed with two operations, everywhere defined, $\vee$ (= join) and $\wedge$ (= meet) satisfying:

$x \vee x = x$	$x \wedge x = x$
$x \vee y = y \vee x$	$x \wedge y = y \wedge x$
$x \vee (y \vee z) = (x \vee y) \vee z$	$x \wedge (y \wedge z) = (x \wedge y) \wedge z$
Absorption Law: $x \vee (y \wedge x) = x$	$x \wedge (y \vee x) = x$

there is a complete duality between $\vee$ and $\wedge$

Theorem:

Any set endowed with join and meet satisfying the above 8 properties is a lattice.

Theorem:

If we set $x \leq y$ to mean $x \wedge y = x$, we obtain a partially ordered set where:

$$\sup(x,y) = x \vee y \quad \text{and}$$
$$\inf(x,y) = x \wedge y$$

In other words, if you define exactly these operations ($\wedge$ and $\vee$), then it turns out that these operations automatically define a partial order.
And this partial order is a lattice, where sup is the operation of join and inf is the operation of meet.

That's Dedekind's contribution.

Proof:

1. First, we need to show that $x \le y$, when defined as $x \wedge y = x$ is a partial order.

   In other words, we need to show that the reflexive, anti-symmetric and transitive properties hold [10.5].

   a) reflexive

   $$x \le x \iff x \wedge x = x$$

   By the definition of $\le$ and then observing that this is precisely one of the properties of the meet operation, as defined [11.5]. ✓

   b) anti-symmetric

   $$\begin{matrix} x \le y \text{ and} \\ y \le x \end{matrix} \Rightarrow x = y \iff x \wedge y = x \text{ and } y \wedge x = y$$

   The commutative property for meet states that $x \wedge y = y \wedge x$.
   Therefore:

   $$x = y \checkmark$$

   c) transitive

   $$\begin{matrix} x \le y \text{ and} \\ y \le z \end{matrix} \Rightarrow x \le z \iff x \wedge y = x \text{ and } y \wedge z = y \Rightarrow x \wedge z = x$$

   $$x \wedge z = x$$

   given $x \wedge y = x$ $\longrightarrow$ $(x \wedge y) \wedge z = x$

   By the associative property of meet:

   $$x \wedge (y \wedge z) = x$$

   Given that $y \wedge z = y$:

   $$x \wedge y = x$$

   And since it is given that $x \wedge y = x$:

   $$x = x \checkmark$$

That's the easy part.

2. Now we need to show that $\vee$ and $\wedge$ are sup and inf in this partially ordered set.
We have a partial ordering, but we don't yet know that $\vee$ is sup and $\wedge$ is inf.

Lemma

$$x \wedge y = x \quad \text{iff} \quad x \vee y = y$$

Proof:

$$x \vee y = (x \wedge y) \vee y$$ given that $x \wedge y = x$

Rewrite using commutative law:

$$= y \vee (x \wedge y)$$

what haven't we used yet?
The absorption law [11.5].

$$y \vee (x \wedge y) = y$$

$$= y \checkmark$$

Since $\vee$ and $\wedge$ are self dual, this proof goes the other way around, as well.
So you have this Lemma.
Now we show that we have, everywhere, sup and inf.

Proof that $x \vee y = \sup(x, y)$:

observe $x \vee y \geq x$ ← why?
Because, from the definition of $\leq$, we have:

$$x \leq x \vee y \iff \underbrace{x \wedge (x \vee y) = x}$$

And this is exactly the absorption law.

Similarily $x \vee y \geq y$

That's not yet enough to show that $x \vee y = \sup(x, y)$.
You have to show that if there is an element that is greater than both x and y, it's also greater than $x \vee y$.

Suppose $z \geq x$ and $z \geq y$.

From the Lemma, we have:

$$z \vee x = z \quad \text{and} \quad z \vee y = z$$

But
$$\left.\begin{aligned} z \vee (x \vee y) &= (z \vee x) \vee y \\ &= z \vee y \\ &= z \end{aligned}\right\}$$
This gives:
$$z \vee (x \vee y) = z$$
Hence:
$$(x \vee y) \leq z$$

[redacted]

Thus $z \geq x \vee y$.

Q.E.D.

For once, I gave a complete proof.
This is a famous argument.
This holds dually for $\wedge$ and inf.

At this point, we can make a big list of partially ordered sets that are lattices.

- A lattice L is said to be <u>complete</u> when, for every subset $A \subseteq L$, $\sup(A)$ and $\inf(A)$ exist.

  This is the analogue of what we call a Boolean algebra of sets which have arbitrary unions and intersections.

- A lattice is <u>distributive</u> when it satisfies the identity:

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$$

Exercise 11.1

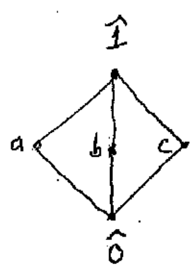
Show that the preceeding identity is equivalent to the <u>dual identity</u>:

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$$

[i.e., interchange join and meet.
$\vee \to \wedge$
$\wedge \to \vee$

Not every lattice is distributive.
Now I'll be a good boy and immediately give you an example of a lattice that is not distributive.

<u>Example</u> – Lattice M_3 is non distributive



M_3

$$a \wedge (b \vee c) \overset{?}{\cancel{=}} (a \wedge b) \vee (a \wedge c)$$
$$a \wedge \hat{1} \qquad \hat{0} \vee \hat{0}$$
$$a \neq \hat{0}$$

Therefore, this lattice is not distributive

It is <u>very rare</u> for a <u>lattice</u> to be <u>distributive</u>.

- If P and Q are posets, $f: P \to Q$ is <u>order preserving</u> when:

$$x \le y \text{ in } P \underset{\text{implies}}{\Longrightarrow} f(x) \le f(y)$$

In particular,

- f is an <u>isomorphism</u> of P onto Q when:

  f is an order preserving monomorphism onto Q and
  f^{-1} is also order preserving and
  $f \circ f^{-1} = I$ and $f^{-1} \circ f = I$ (I = identity)

Now we look at some famous <u>examples</u> of posets and lattices:

- $P(S)$, the Boolean algebra of subsets, is a lattice for any set S, where

$$\vee = \cup$$
$$\wedge = \cap$$

In fact, this is a <u>complete lattice.</u>

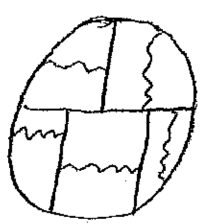
- $B[S]$ = the family of all Boolean subalgebras of S

  setting $B \le B'$ when $B \subseteq B'$

  This is a partially ordered set. I haven't verified that this is a lattice yet.

- $\Pi[S]$ = all partitions of S

  setting $\pi \le \pi'$ when the partition π is a refinement of the partition π'
  (i.e., every block $B \in \pi$ is contained in some block C of π')



The picture is like this.
Here's your set S and the partition π' = —— cuts.

To get π the refinement of π', you cut the blocks C of π' up. Cuts = 〰.

The blocks of π are each contained in a block of π'.

In particular, $\hat{0}$ is the partition where every element is a block to itself.
$\hat{1}$ is the partition with only one block.

Next time, we'll show that $B[S]$ and $\Pi[S]$ are <u>isomorphic.</u>

Some of this material can be found in my book (Gian-Carlo Rota on Combinatorics) pp. 516-560.

- Maximum element

A maximum element $\hat{1}$ of a poset P is the element $\hat{1}$ s.t.

$$\hat{1} \geq x, \text{ for all } x \in P,$$

if any.

We have been discussing the notion of partially ordered sets $(P, \leq)$.
Before we go any further, I feel it is my duty to inform you that there is another notion, which is in a sort of no man's land, which we have to briefly discuss.
The notion of a quasi-ordered set. And it does come up. You have to know it exists.

- A quasi-ordered or pre-ordered set Q is the set with a relation $R \subseteq Q \times Q$ s.t.

1. $R \supseteq I$ reflexive
2. $R \circ R \subseteq R$ transitive

But not anti-symmetric.

What happens if you don't have anti-symmetry?
Fortunately, there's a structure theorem for quasi-ordered sets.
It allows us to deduce the study of quasi-ordered sets from the study of partially ordered sets.
Namely:

If Q is a quasi-ordered set, let R' be defined as follows:

$a, b \in Q$

$aR'b$ whenever aRb and bRa

It follows that the relation R' is an equivalence relation on Q.

because it's reflexive, symmetric, and transitive.

Let $\overline{Q}$ be the set of equivalence classes. (or blocks of the partition)

For $\alpha, \beta \in \overline{Q}$, set $\alpha \leq \beta$ whenever:

aRb for some $a \in \alpha$, $b \in \beta$

This is well defined and defines a partial order on $\overline{Q}$.

Thus, every quasi-order splits into an equivalence relation and a partial order.

Example

Let R = any relation $R \subseteq S \times S$

The transitive closure of R is the relation:

$$R_{trans} = I \cup R \cup R \circ R \cup R \circ R \circ R \cup \dots$$

You can verify that R_{trans} is a quasi-order.

These constructions, quasi-orders, are extremely frequent in mathematics.

- You remember that a lattice is a partially ordered set with a sup and an inf. As we showed last time [11.5-7], if we set $x \leq y$ to mean $x \wedge y = x$, the sup is $\vee$ (join) and the inf is $\wedge$ (meet).

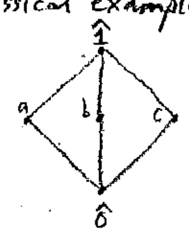
It is a remarkable fact that these joins and meets can be viewed as abstract algebraic operations, completely defined by identities, as we've seen. Any algebraic system that can be identified by identities automatically enjoys a number of properties, which maybe we'll talk about, if time permits. It is very important, when given an algebraic system, to see if you can redefine it using only operations and identities, because that allows you to apply general theorems of universal algebra.

It just so happened that Dedekind discovered that sup and inf can be defined with algebraic identities. As you recall, the identities are that they are idempotent, commutative, associative, and they satisfy the absorption law. From this, we can recover the partial order of the set.

- We say a lattice is distributive when:

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \quad \text{and, dually,}$$
$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

We saw that not every lattice is distributive. A classical example is M_3:

M_3 

If a lattice turns out to be distributive, we're very lucky. There are some really dissimilar lattices that turn out to be distributive, as you will see.

- What are some examples of distributive lattices? $P(S)$
  Well, one example we have seen, Namely, the Boolean algebra of subsets of a set, where join is $\cup$ and meet is $\cap$.

  What about another example:

- The order ideal of a partially ordered set P is a subset I of P s.t.

  if $x \in I$ and $y \leq x$ then $y \in I$

  this is also known as a descending set.

- The dual notion of an order ideal is a filter ← this is also known as an ascending set.

- Intuitively, if you visualize the Hasse diagram of P, then the order ideal consists of taking the total anti-chain and taking everything underneath.

- If I and I' are order ideals, then so are:

  $$I \cap I' \quad \text{and} \quad I \cup I'$$

- Theorem

  The family $I(P)$ of all order ideals of a partially ordered set P is a distributive lattice.

  This is how you get distributive lattices galore.
  You just take a partially ordered set and the set of all its order ideals.
  You get a distributive lattice.

  These are not Boolean algebras, because the complement is a filter.
  So these distributive lattices are not Boolean algebras.
  They are not closed under complement.

  If I = order ideal, then its complement I^c is a filter.

  We will see shortly that if P is a finite partially ordered set, then the converse of this theorem is true, That's the famous theorem of Birkhoff.

- Theorem – Birkhoff

  Every finite distributive lattice is isomorphic to the lattice of order ideals of some partially ordered set.

  For infinite distributive lattices, that's not true. That is part of the chapter of profinite Combinatorics.

So, here we have one prime example of a lattice.
We take the family of order ideals of a partially ordered set.

• Remark

To every partially ordered set, you can associate a topological space.
The order ideals of P define the closed sets of a topology.

In this way, to every poset, you can associate a topological space.
All the pathologies of algebraic topology can already be found by examples of this kind of topology. Homotopy theory, etc. You can always get it from this topology. Even finites.
These topologies include a wide variety of topological spaces.

Partitions and Boolean Subalgebras

$B[S]$ denotes the family of all Boolean subalgebras of S

$$B_1, B_2 \in B[S], \text{ set } B_1 \leq B_2 \text{ when } B_1 \subseteq B_2$$

$\Pi[S]$ denotes the family of all partitions of a set S

$\pi, \pi' \in \Pi[S]$, set $\pi \leq \pi'$ when every block of π is contained in some block of π'

We want to show that $B[S]$ and $\Pi[S]$ are lattices.
And we want to get a clear idea what sup and inf look like.

$$\text{Set } B_1 \wedge B_2 = \underbrace{B_1 \cap B_2}_{\text{the intersection of two Boolean subalgebras is a \underline{Boolean subalgebra}}}$$

$$\text{Set } \pi \wedge \pi' = \{B \cap C : B \cap C \neq \emptyset,\ B \in \pi,\ C \in \pi'\}$$
the partition whose blocks are so defined.

Earlier in this course, we showed that to every Boolean subalgebra, there corresponds a partition. And to every partition, there corresponds a Boolean subalgebra. [2.13]
Now, let's exploit this.

Given $\underset{\text{partition}}{\pi} \rightarrow Bool(\pi)$ = Boolean subalgebra whose atoms are the blocks of π

$\underset{\text{Boolean subalgebra}}{B_1} \rightarrow Part(B_1)$ = set of atoms of B_1

We can do this, of course, because we can take arbitrary unrestricted unions and intersections.
And we have shown that this correspondence is a <u>bijection</u>.
This bijection is order preserving.

This gives an order preserving bijection of $\Pi[S]$ onto $B[S]^*$

[We use the $*$ to indicate the <u>reversal</u> here. The <u>finer</u> the partition is, <u>the bigger</u> the resulting Boolean subalgebra that is generated.]

So we have that the <u>partially ordered set of all partitions</u> is <u>isomorphic</u> to the <u>partially ordered set of all Boolean subalgebras</u>.

Observe that it is easy to define the <u>meet</u> for Boolean subalgebras and partitions, as we have just done.
But, under the order inverting isomorphism, the meet becomes a join.
Therefore, you define the join by exploiting this isomorphism.

Hence, we define:

$$\pi \vee \pi' = Part\big(Bool(\pi) \cap Bool(\pi')\big)$$

$$B_1 \vee B_2 = Bool\big(Part(B_1) \wedge Part(B_2)\big)$$

Thus, both $B[S]$ and $\Pi[S]$ are <u>lattices</u>.
Of course, they are complete lattices, because you can take arbitrary unions and intersections.

$\Pi[S]$ and $B[S]$ are <u>complete lattices</u>
They are <u>not</u> distributive lattices.

The lattice $\Pi[S]$ is, to my mind, the most interesting lattice there is.
You can find everything in it.
The lattice of partitions of the 4 element set is, already interesting.
Let's get a feel for it.

<u>Example</u> - the lattice $\Pi[S]$

Let $S = \{a, b, c, d\}$

You can't write the Hasse diagram. It would take the rest of the period.
But let's see what the Hasse diagram looks like - roughly.

$\hat{1} = \{a,b,c,d\}$ rank=3

$\{a,b\}\{c,d\}$ $\{a,b,c\}\{d\}$ rank=2 ⟵ has two types of partitions. One of the form 2+2, the other 3+1.

$\{a,b\}\{c\}\{d\}$ rank=1 (atoms)

$\hat{0} = \{a\}\{b\}\{c\}\{d\}$ rank=0

$\Pi[\{a,b,c,d\}]$

rank = #elements in the set - # blocks in partition

if S <u>finite</u> then

$$r(\pi) = |S| - |\pi|$$

Now you say - that's good. That works for <u>partitions of a set</u>.
What about <u>partitions of a number</u>?
Let's see what we can do.

Something funny happens.
There are 2 partial orders on the partition of a number.
A good one and a bad one.
First, let's talk about the bad one, as that's the first one that will occur to us.

<u>Bad</u> – Partitions of a number

Given $n \in \mathbb{N}$

$P(n)$ = partitions of n
In other words, a multiset of integers whose sum is n [4.16]

If $\alpha, \beta \in P(n)$, say $\alpha \prec \beta$ when β is obtained from α by replacing two "summands" of α by their sum.

{ β covers α if you can take 2 elements of α, add them, and then get another multiset, which is β }

The <u>transitive closure</u> of this covering relation is a partial order $\leq$ on $P(n)$, called <u>refinement</u>.

However, this is <u>NOT</u> a lattice.

This partially ordered set $(P(n), \leq)$ is what people use when they want to find a bad partially ordered set.
In other words, if they have a property, and they want to find some partially ordered set where the property doesn't hold, chances are that it doesn't hold in $(P(n), \leq)$.
So it is often used for counterexamples.
That's what it is mostly used for.
The simplest questions are not answered by this partially ordered set.
It's weirdo.

<u>Good</u> – Partitions of a number

Another partial order on the set $P(n)$, which is really good is the <u>dominance</u> order.
Given $n \in \mathbb{N}$

$P(n)$ = partitions of n

$\lambda \in P(n)$

arrange the entries of the multiset λ in non-increasing order

$$\lambda = (\lambda_1 \geq \lambda_2 \geq \ldots), \quad \lambda_i \in \mathbb{N}, \quad \sum_i \lambda_i = n$$

We say that $\lambda \geq \lambda'$ in the <u>dominance order</u> when:

$$\lambda_1 \geq \lambda_1'$$
$$\lambda_1 + \lambda_2 \geq \lambda_1' + \lambda_2'$$
$$\lambda_1 + \lambda_2 + \lambda_3 \geq \lambda_1' + \lambda_2' + \lambda_3'$$
$$\vdots$$
etc.

This defines a partial order.
This partial order arose first in statistics. There are a tremendous number of applications.
And, strangely enough, it was first defined in the continuous case.
If you take a function on $[0,1]$, you can define a non increasing rearrangement of that function.
A function is $\leq$ to another in the partial order if the definite integral of one from 0 to x is $\leq$ the definite integral of the other from 0 to x, for every x.

How do we <u>visualize</u> the dominance order?
One way to visualize this partial order is to visualize the covering relation.
To visualize the covering relation, we can associate a Ferrers matrix [4.11~12] with λ and λ'.

- Ferrers Matrix:

$\lambda = (5,1,1)$ $\qquad\qquad$ $\lambda' = (3,2,1)$

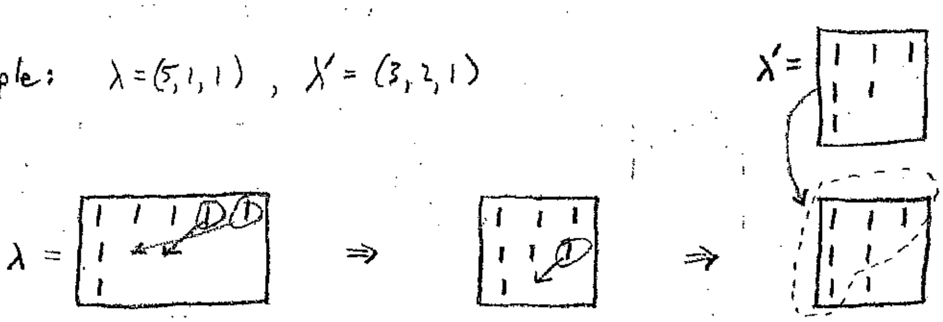
$$\begin{array}{c} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{array}\left[\begin{array}{ccccc} 1&1&1&1&1\\ 1& & O & & \\ 1 & & & & \end{array}\right] \qquad \begin{array}{c} \lambda_1' \\ \lambda_2' \\ \lambda_3' \end{array}\left[\begin{array}{ccc} 1&1&1\\ 1&1& \\ 1 & & O \end{array}\right]$$

The covering relation $\lambda_i \underset{}{\overset{\text{covers}}{\geq}} \lambda_i'$ is equivalent to saying that you can move "1" entries down in such a way that:

(1) "1" entries, in the covering Ferrers matrix, can be moved down in such a way that the resulting matrix is a Ferrers matrix (maintains the Ferrers relationship).

(2) this resulting Ferrers matrix contains the covered Ferrers matrix.

Example: $\lambda = (5,1,1)$, $\lambda' = (3,2,1)$



$\lambda \geq \lambda'$
↑ in the dominance order

Let me mention 2 funny facts about the dominance order.

1. The dominance order is a linear order up to $n=5$. So it was missed early in the game. People would test things up to $n=5$ and say, — "oh, it's a linear order." Funny things happen when $n=6$, since the dominance order is no longer a linear order.

2. There is an ortho complement in the dominance order.

If P is a partially ordered set, an ortho complementation is a map

$$x \to x^{\perp}, \quad x \in P$$

s.t.

1. if $x \leq y$ then $x^{\perp} \geq y^{\perp}$
2. $x^{\perp\perp} = x$

You have that the dominance order is ortho complement.

Setting $\lambda^{\perp}$ = the partition whose Ferrers matrix is the transpose of the Ferrers matrix of λ, we obtain an ortho complement.

It is very rare for a partially ordered set to have a complement, as we will see.

** Exercise 12.1

Open problem.
Give a structural characterization of the dominance order.

Give an order theoretical characterization of the dominance order, in terms of the properties of its orthocomplementation. In other words, the dominance order is the only order with the following properties. There is every reason to believe there is such a characterization, but no one's got it yet.

The dominance order is a lattice. I leave it to you to prove this.
But it is not a distributive lattice.

Theorem

The dominance order is a lattice

What's another example of an ortho complemented partially ordered set?
Boolean algebra.

Given a Boolean algebra, you take the complement of the set — that's an ortho complement,
Whereas, with the lattice of partitions, there is no ortho complement.

The Wonderful World of Order (cont'd)

The last partially ordered set we discussed last time was the dominance order.

- Dominance order

$P(n)$ = family of all partitions of the positive integer n

Take $\lambda \in P(n)$ and members of the multiset in non-increasing order

$$\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots), \quad \sum_i \lambda_i = n, \quad \lambda_i \geq 0 \text{ integers}$$

The dominance order is defined $\lambda \geq \lambda'$ whenever:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_i \geq \lambda_1' + \lambda_2' + \cdots + \lambda_i', \quad \text{for } 1 \leq i \leq n$$

This is the right (i.e., good) kind of order for partitions of a number.

As we say, this partially ordered set has an ortho complement.

- An ortho complement in a partially ordered set P is a map

$$x \to x^{\perp}$$

s.t.

1. $x \leq y \Rightarrow x^{\perp} \geq y^{\perp}$ (ortho complement is order inverting)
2. $x^{\perp\perp} = x$

In the dominance order, $\lambda^{\perp}$ is the partition whose Ferrers matrix is the transpose of the Ferrers matrix of λ.

I stated as a two star problem to give a structural characterization of the dominance order in terms of the ortho complement and some properties that remain to be discovered.
I did not prove that the dominance order is a lattice, but you can verify that to your heart's content.

- Another example of ortho complement is, of course, $P(S)$ the Boolean algebra of all subsets of a set is ortho complemented.
We set $A^{\perp} = A^{c}$ to get the ortho complement.

Ortho complemented partially ordered sets are quite rare.

Exercise 13.1

Given P = partially ordered set

$\mathcal{I}(P)$ = lattice of order ideals (distributive lattice)

Show that $\mathcal{I}(P)$ is ortho complemented iff P is an anti chain.

In which case the lattice of order ideals $\mathcal{I}(P)$ is actually a Boolean algebra.

* Exercise 13.2

Theorem of Gale-Ryser

Suppose we have a relation: $R \subseteq S \times T$, finite, $|S| = n$, $|T| = k$
And this relation has marginals.
We stated earlier that there are necessary and sufficient conditions for two given sequences of numbers to be the sequences of marginals of some relation.
Now we can answer the question.

Given sequences of positive integers:

$$r_1 \geq r_2 \geq \dots \geq r_n \quad \text{and} \quad s_1 \geq s_2 \geq \dots \geq s_n$$

when does there exist a relation $R \subseteq S \times T$ whose marginals are $\underline{r}$ and $\underline{s}$?

This is a very important question.
The answer is the Theorem of Gale-Ryser.

The answer is:

$$\text{iff} \quad s \leq r^{\perp} \text{ in the dominance order}$$

Very elegant. Observe that this relation is symmetric. If you $\perp$ both sides, you get:

$$s^{\perp} \geq r$$

Later on we'll see that this theorem comes out cheapo as a consequence of matching theory of matroids.
For now, I want you to do it by rolling up your sleeves.

The idea is this. You pack up, in the Ferrers matrix, all the 1's together. And you start shifting them to the right. And you shift them to the right in the correct position to get the right marginals.

This is a very important theorem.

Now we continue with our list of famous partially ordered sets, followed by a list of "red hot" partially ordered sets.

Our next examples have to do w/ vector spaces.
Here, we have 2 kinds of examples:

1. convexity
2. projective space

Some of you have not been introduced to these notions, so we have to review.

Convexity in $\mathbb{R}^n$ (a seeming digression)

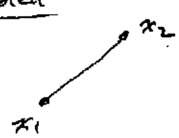
Convexity is a very important chapter in combinatorics.
We could easily spend the rest of the term on convexity alone.
And you will see this limited discussion of convexity in my book "Introduction to Geometric Probability." Half of this material is in my book.

First, we define a convex linear combination of vectors or points

Given vectors or points $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$, a convex combination is a vector or point of the form:

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \quad , \quad \lambda_i > 0 \quad , \quad \sum_i \lambda_i = 1$$

Idea



Given the points x_1 and x_2, the set of all convex combinations of x_1 and x_2 span the segment joining x_1 and x_2.

The closure of the set of all convex linear combinations of a set $A \subseteq \mathbb{R}^n$ is called the convex closure of A.

It is the smallest convex set containing the set A.

For example, if you have the set A:

the convex closure includes all this

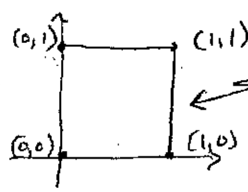
- A convex polyhedron is the convex closure of a finite set of points.

Example : $(\delta_{i1}, \delta_{i2}, \ldots, \delta_{in}) = e_i \qquad i = 1, 2, \ldots, n$

Kronecker δ

The convex closure gives an $n-1$ simplex
In 2 dimensions, a segment
3 " , triangle
4 " , tetrahedron
etc.

Example : Take all points $(a_1, a_2, \ldots, a_n)$ where $a_i = \{0 \text{ or } 1$
Take the convex closure. What do you get?
In $\mathbb{R}^2$:



In $\mathbb{R}^n$, you get the n-cube

- It is a fact, which we may prove later, that in dimensions 5 or greater, there are only 3 regular polyhedra:

1. n-simplex
2. n-cube
3. dual of the n-cube ← the dual of the n-cube is obtained by placing the points in the middle of each face.

in $\mathbb{R}^2$ you again get an n-cube
in $\mathbb{R}^3$ the n-cube dual is an octahedron.

So there are 3 regular solids.
They are the analog of the tetrahedron, the cube, and the octohedron.
These are the only ones that exist in dimension 5 or greater
It's a very basic fact of life. And there are certain consequences.
In each dimension, how many are there?

2 D — ∞ many (any regular n-gon)
3 D — 5 platonic solids (and only 5) ← cube, tetrahedron, octohedron, icosohedron, dodecahedron.
4 D — low and behold, there are 6

Let's take the n-simplex and look at it and the lattice of its faces.

- Lattice of faces of n-simplex

For the case $n=3$, we have a tetrahedron.

The faces are the vertices, the sides, the 2D faces, and the 3D face.

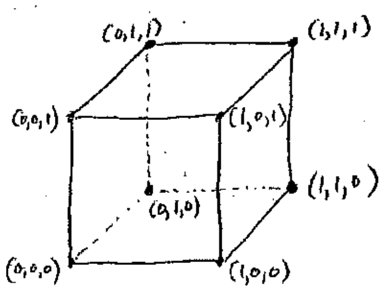
view from top

Any idea what that looks like when you draw the Hasse diagram?
The Boolean algebra of subsets of a 4 element set.
If you take any subset of vertices, that subset spans a face.

The lattice of faces of the n-simplex is isomorphic to the Boolean algebra of subsets of an n-set.

This is a good way of visualizing a Boolean algebra.
You can visualize the complement to a set by flipping a face across.

- Lattice of faces of n-cube



Faces are described as follows:
fix certain number of coordinates to be either 0 or 1.
Let all other coordinates vary entirely, to include both 0 and 1. We'll use x for this purpose.

A face is uniquely determined by a sequence of 0's, 1's, and x's.
assigned all possible combinations of 0's and 1's.

Examples:

$(1,1,1)$ = vertex
$(x,1,0)$ = side
$(x,0,x)$ = face
(x,x,x) = all faces

- Compare this with the <u>simplex</u>, where the faces are sequences of 0's and 1's, but <u>no</u> x's. This is because the faces correspond to a subset of an n-set.
You put a 1 where each element is in the set and a 0 where each element is not in the set.
For example, the tetrahedron face $(0,1,1,0)$.

For the n-simplex, we represent faces with a sequence of 0's and 1's.
For the n-cube, we represent faces with a sequence of 0's, 1's, and x's.

Now, having this so defined, I can define an <u>infinite dimensional cube</u>.

A <u>cubical lattice</u> is the family of all <u>signed "subsets"</u> of a set S.

to every element of S, you assign 0, 1, or x.

Now we define the order by secretly using the face numbers of the cube.
If we look at the faces of the cube, the <u>more x's</u> you have, the <u>bigger the face</u>, because of the greater the number of combinations.

A, B = signed subsets of S

We partition A and B into 3 blocks, which are the set of all elements of A and B signed 0, 1, and x, respectively:

$$A = (A_0, A_1, A_x)$$

$$B = (B_0, B_1, B_x)$$

We say that $A \le B$ when:

$$B_x \supseteq A_x$$

$$B_0 \subseteq A_0$$

$$B_1 \subseteq A_1$$

You can verify, from the preceeding reasoning, that if S is a <u>finite</u> set, you obtain a lattice, which is a lattice of faces of a cube.
Now we can define cubical lattices for any n.

You can find in my book "Gian-Carlo Rota on Combinatorics" pp. 561-563 a structural characterization of the lattice of faces of the cube.
Besides the join $(\vee)$ and meet $(\wedge)$, there is also an analogue for <u>complement</u> for the lattice of faces of the cube. Which means <u>flipping a face, across a face</u>. In this paper, we have characterized all these flippings - that's called <u>diagonal maps</u>, which are the cubical analog of a combinatorial set.

*Exercise 13.3

Rewrite pages 561-563, with all details.

*** Exercise 13.4

Now, let's think philosophically.
If you take the dual of the cube, namely, the other regular solids that exist in n dimensions, the lattice of faces will be the dual of the lattice turned upside down.
Therefore, from the point of view of lattices in n dimensions, there are only 2:

1) lattice of faces of the n-simplex
2) lattice of faces of the n-cube
(the lattice of faces of the n-hyperoctohedron is dual to the lattice of faces of the n-cube)

This leads us to the following speculation.
Boolean algebra is what you use to do ordinary logic.
There must be another kind of logic that goes with the lattice of faces of the cube.
That is sketched towards the end of the paper, but never fully developed.

Develope cubical logic.

intuitively, this is the logic of 1 = yes,
0 = no,
x = not yet known.

But this has never been formalized.

Furthermore, in combinatorics, we prove theorems about sets.
So prove the analogues in cubical lattices,
Write some cubical lattice analogue of a theorem of sets.
If you do it, you'll be in the best of families.
You can do it. It's a good exercise.
Anything you can think of has a cubical lattice analogue.

} this is a nice, cheapo way of writing papers.

I want to show that there are two other lattices associated with a convex set.

The family of all convex closed sets in $\mathbb{R}^n$ is a lattice where:

$$A \wedge B = A \cap B$$

$$A \vee B = \text{convex closure of } A \cup B$$

← unfortunately, the union of 2 convex sets is not convex.
You need to take the convex closure.

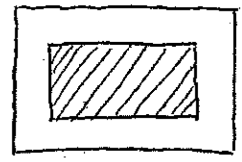
This is not a very nice lattice.
This has also been characterized structurally. But we are not very interested in it.

Lattice of polyconvex sets – a more interesting lattice

A polyconvex set is a finite union of convex closed sets.

A polyhedron is a finite union of convex polyhedra.

↑ also known as a polytope.



Polyconvex sets are a distributive lattice
Polyhedra are a distributive lattice (Polyhedra are a sublattice of this distributive lattice of polyconvex sets.)

These facts will have enormous consequences, as we will see.

These are the most famous distributive lattices that are not finite.

Order (Cont'd)

Last time, we began listing examples of famous partially ordered sets and lattices that have to do with vector spaces.
Last time we focused on convexity. Today, projective space.

Projective Space

This is one of the most important examples of a lattice.
This is part of what everybody should know about mathematics.
It is too bad this doesn't even get taught – except in algebraic geometry courses.

Projective space, in a strictly theoretical sense, is the study of one lattice

$$V = \text{vector space of dimension } n < \infty$$

$$L(V) = \text{poset of linear subspaces of } V$$

(all the subspaces pass through the origin, by definition)

$L(V)$ is a lattice where, for $W, W' \in L(V)$ — W and W' are linear subspaces

$$W \wedge W' = W \cap W'$$ ← As you know, the intersection of 2 linear spaces is a linear space.

$$W \vee W' = \text{span}(W, W') = \underbrace{\{x+y : x \in W, y \in W'\}}_{\text{set of all vectors } x+y}$$

After Boolean algebra, this is the most important lattice there is. Let's discuss some of the properties of this lattice.
The study of this lattice is called projective geometry.
Why it's called geometry, we'll see in a minute.
First of all:

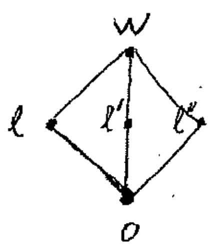
$L(V)$ is not distributive.

If W is a plane, ℓ, ℓ', ℓ'' are lines in W in general position, — meaning no 2 lines are really the same.
then:

$$\ell \vee \ell' = W, \quad \ell \vee \ell'' = W, \quad \ell' \vee \ell'' = W$$

$$\ell \wedge \ell' = 0 \text{ subspace}, \quad \ell \wedge \ell'' = 0, \quad \ell' \wedge \ell'' = 0$$

Therefore, the configuration gives the following Hasse diagram:



This configuration prevents anything from being distributive.
Therefore, $L(V)$ is not a distributive lattice.

How close does L(V) come to being a distributive lattice?

Modular Law

For $x, y, z \in L(V)$, the distributive law holds if 2 of $\{x, y, z\}$ are comparable.

The atoms of $\underbrace{L(V)}$ are the straight lines.

L(V) is a ranked partially ordered set

The rank is the same as the dimension:

$$r(W) = \dim(W)$$ ← the rank of a straight line = 1

Here's another property that the lattice of subspaces of a vector space shares with the Boolean algebra. Namely:

$$\dim(W \vee W') + \dim(W \wedge W') = \dim(W) + \dim(W')$$

How do you prove this?
You take a basis of the intersection. This is a partially linearly independent set.
This set can be completed even to a basis of W onto a basis of W'.
You add up and you get the sum of two dimensions.

Warning - When we see the above identity, it is tempting to say:
"Oh - then we can apply inclusion-exclusion."
But, the answer is: No, No, No.

Inclusion-exclusion is not valid here.

I leave it to you to find a counterexample.

The analog of inclusion-exclusion is a whole theory, which is called the Schubert Calculus.

Let's look at another property of Boolean algebra and see what the analog is in the lattice of linear subspaces:

Every element of L(V) is the sup of a set of atoms.

That's like saying you take a basis.
W is a subspace. You take a basis of W. Every element of a basis spans a line.
The sup of this line is in W.

This is kind of like Boolean algebra. But notice, however, that in a Boolean algebra, every element is the sup of a unique set of atoms.
In L(V) an element can be the sup of atoms in many ways.

Let's talk about complements now.
This is most striking.

In Boolean algebra, every element has a complement.

We defined an ortho complement for a lattice.
I should have defined the concept of a complement first.
Let's digress and define the notion of a complement in a general lattice.

In lattice L with $\hat{0}$ and $\hat{1}$, we say that an element y is a complement [redacted] of an element x when:

$$x \vee y = \hat{1} \text{ and}$$

$$x \wedge y = \hat{0}$$

In general, an element of a lattice need not have a complement.
For example, take the line. That's a lattice. But it doesn't have a complement.
It's a rare event for an element of a lattice to have a complement.
In a Boolean algebra, every set has a complement and a unique one.

What happens in the lattice $L(V)$?
(this is an extraordinary finding — one of the deepest theorems of combinatorics)

In $L(V)$, every element has a complement.

Why?
Take a subspace W. Take a basis of W. The basis of W can be completed to a basis of the whole space.
Take the elements of the basis of the whole space that are not in W.
They span another subspace W'.
And together:

$$W \vee W' = \hat{1} \quad \text{and} \quad W \wedge W' = \hat{0}$$

Because they are linearly independent.

So, every element of $L(V)$ has a complement — but NOT a unique one.
Here's how I describe this property.

The set of all complements of an element $W \in L(V)$ is an antichain.
Why?

Because if W has dim k, then the complement has dim $n-k$.
So any two complements would have dim $n-k$! Therefore, they can not be comparable.

The non-trivial fact is that the <u>converse</u> is true.

[If you have a lattice, which is atomic (in other words, every element is the sup of atoms) and which has the property that every element has a set of complements which is an antichain,

then it is the lattice of all subspaces of some vector space, over some field, not necessarily a linear subspace.]

This is a <u>very deep</u> theorem.
In fact, it has a 150 year history. And, to this day, it takes about 30 pages to prove.
There is no really simple proof.
So I'll just state it for you.

<u>Theorem</u> – von Staudt-von Neumann (Must be dimension 3, at least. This does not apply to the plane)

Conversely, a lattice L with finite chains having a chain of length ≥ 4 which is atomic (i.e., every element is a sup of atoms) and with the property that the set of complements of any element $x \in L$ is an antichain is isomorphic to the lattice of all subspaces of a vector space over a field.

You pull out a whole field from this.
It is astonishing.

(L is also assumed to be irreducible – see [15.2])

Some ancedotal history of the proof of this Theorem.
This was first discovered in the 19th century by the famous German geometer von Staudt in order to construct the Algebra of Froes. A very complicated algebra. He didn't have the concept of a lattice.

Then it was forgotten. In the 1930's, von Neumann rediscovered it from scatch. Not knowing of von Staudt's work. When someone told him of this, he had a fit. Literally. 3 years for nothing.
However, von Neumann went immediately one up on von Staudt.
Because he generalized this to infinite dimensional behavior.
And he constructed vector spaces that take continuous values from $[0,1]$.
So he built up continuous geometry, motivated by quantum mechanics.

Then Emil Artin, the father of Michael Artin here, gave one of the simplest possible proofs. A very elegant, short proof.
Unfortunately, the proof was just mimeographed and distributed to graduate students at U. of Notre Dame.
So it's very hard to get hold of it.

Since then, people have simplified the proof by all sorts of methods. Mark Haiman, who wrote his thesis here in 1984, constructed a very simple proof, assuming that the lattice L is a lattice of commuting equivalence relations. We will see that this assumption is not outrageous.

Now let's look at some additional properties of lattices.
Why did I bring up lattices of commuting equivalence relations?

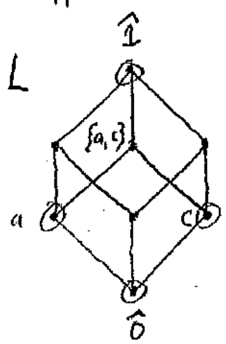
The lattice $L(V)$ is isomorphic to a sublattice of the lattice of all partitions of the set V. And these partitions correspond to commuting equivalence relations.

The most important thing in this statement is the notion of sublattice.

Sublattices

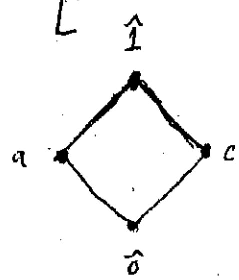
Suppose we take the Boolean algebra of subsets of 3 element sets,

L



Now I take the circled elements $\longrightarrow$

L'



Together, these elements form a partially ordered set, whose Hasse diagram is as shown. This poset doesn't know that the other elements existed. Indeed, this poset is a lattice. $\sup(\cdot,\cdot) = \hat{1}$, $\inf(\cdot,\cdot) = \hat{0}$.

$\sup(a,c) = \{a,c\} \neq \hat{1}$
$\inf(a,c) = \hat{0}$

$\sup(a,c) = \hat{1}$
$\inf(a,c) = \hat{0}$

L' is NOT a sublattice of L, because sup and inf are not the same for all corresponding elements in the two lattices.

To say a subset of a partially ordered set is a sublattice is a very strong statement.

A sublattice L' of a lattice L is a subset of L, which is a lattice, s.t.

$$\left.\begin{aligned} \sup_{L'}(x,y) &= x \vee_L y \\ \inf_{L'}(x,y) &= x \wedge_L y \end{aligned}\right\}$$

In other words, the sup and the inf coincide.

It is very easy to embed any partially ordered set in another.
But it's not so easy to embed a lattice in another as a sublattice.

It is, therefore, remarkable that the lattice $L(V)$ is isomorphic to a sublattice of the lattice of all partitions of the set V.

Theorem:

the equivalence relation defined by the subspace

The map $W \to R_W$ is an isomorphism of the lattice $L(V)$ into the lattice $\underbrace{\Pi[V]}_{\text{partitions of } V \text{, viewed as a set.}}$.

Proof: (much of this we have seen before [7.8–10])

Recall that for $\underbrace{x, y}_{\text{vectors}} \in V$, $x R_W y$ iff $x - y \in W$ ← that's how we defined the equivalence relation R_W.

We have shown that:

$$R_W \circ R_{W'} = R_{W \vee W'}$$ ← we wrote $R_{span(W, W')}$ before

$$R_W \cap R_{W'} = R_{W \wedge W'}$$ ← this is trivial to show

It remains to be shown that $R_W \circ R_{W'}$ is the sup.
I haven't shown this yet. I forgot.
A small digression.

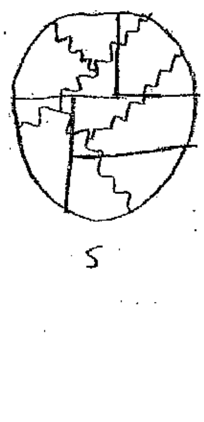
What do $\vee$ (join) and $\wedge$ (meet) in $\Pi[S]$ look like?

How did we define them?
We defined them indirectly by the isomorphism between the lattice of partitions and the dual of the lattice of Boolean subalgebras.

The intersection of two Boolean algebras is a Boolean algebra – that corresponds to the join of two partitions.
And the intersection of the blocks, pairwise, of two partitions will give you the meet.
We get the join by using Boolean algebra.
We got the meet directly from partitions.

Now we want to see how to construct the join in terms of partitions alone.
What is the idea?

You have a set with two partitions:



S

meet – to intersect all the blocks, you take all the blocklets and discard the empty blocks.

join – the roughest partition that contains them both as refinements.

How do we define that?

$\pi, \pi' \in \Pi[S]$

A reminder that:

$$\pi \wedge \pi' = \{B \cap C : B \in \pi,\ C \in \pi',\ B \cap C \neq \emptyset\}$$

How do we define $\pi \vee \pi'$?

We take R_π, $R_{\pi'}$
Then we define an equivalence relation R' as follows:

For $x, y \in S$, set $x R' y$ whenever $x R'' y$ for some R'' of the form:

$$R'' = R_\pi \circ R_{\pi'} \circ R_\pi \circ R_{\pi'} \circ \dots \circ R_\pi$$
$$\text{or} \quad R'' = R_{\pi'} \circ R_\pi \circ R_{\pi'} \circ R_\pi \circ \dots \circ R_{\pi'}$$

These equivalence relations don't commute, in general.
However, you compose the relations <u>any number of times</u> in this manner and you say $x R' y$.

the number of compositions can not be restricted, in general.

Exercise 14.1

It turns out that $R' = R_{\pi \vee \pi'}$. Show this.

What I've really done is rephrase the stuff of Boolean algebra.
The join is obtained by iterating the occurrances of R_π and $R_{\pi'}$.
You can not have two R_π consecutively, for example, because $R_\pi \circ R_\pi = R_\pi$, since this is an equivalence relation.

In particular, if R_π and $R_{\pi'}$ commute then $R_{\pi \vee \pi'} = R_\pi \circ R_{\pi'}$

Say $R'' = R_\pi \circ R_{\pi'} \circ R_\pi \circ R_{\pi'} \circ \dots \circ R_\pi$
iteratively commute and reduce $R_\pi \circ R_\pi = R_\pi$ and $R_{\pi'} \circ R_{\pi'} = R_{\pi'}$.
And, after you simplify, you have:

$$R'' = R_\pi \circ R_{\pi'}$$

That's exactly what we are doing in the theorem.
We know that these equivalence relations commute.
We have verified that in detail before.

Therefore, the composition of the two equivalence relations is the join of the partitions.
That completes the proof of the theorem.

This is an extremely remarkable fact.
You have a sublattice of the lattice of partitions of the set where any two partitions commute.
And that's given by a vector space. Any vector space gives you a sublattice of partitions where any two of them commute.
If this is not extraordinary, I don't know what is.
In fact, we give it a name: I don't like this term.

A linear lattice (or type I lattice) is a sublattice of $\Pi[S]$, the lattice of partitions of a set, in which any two partitions commute.

$L(V)$ is a linear lattice.

This is a fundamental result.
Are there any other linear lattices besides $L(V)$?
Yes. They're all over the place.

The lattice of all normal subgroups of a group - that's a linear lattice.
Because I told you that every normal subgroup defines an equivalence relation.
And any two normal subgroups define commuting equivalence relations.
So the lattice of normal subgroups of a group is a linear lattice.

The lattice of all ideals of a ring - that's a linear lattice.

The lattice of all submodules of a module - that's a linear lattice.

They're all over the place.
So they ought to have interesting properties.
And we will see that they do.

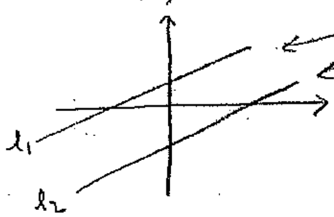
Now you say "that's all fine and dandy, We talk about $L(V)$ and call that projective geometry. But, where's the geometry?"

It's not easy to visualize $L(V)$. And you visualize it in terms of $\vee$ (joins) and $\wedge$ (meets).
In fact, von Neumann used to say you couldn't do this. The idea was that you would visualize $\vee$ and $\wedge$ in $L(V)$ just as you would view $\cup$ and $\cap$ of sets by Venn diagrams. But that's pretty tough. And you can't do it by Venn diagrams because the lattice is not distributive.

So, what are we going to do?
We introduce <u>projective space</u>.

Kepler was the first to introduce projective space. (Kepler, the famous astronomer)
This was the greatest discovery he ever made.

Let's take the plane, by way of motivation.
I want to take <u>points and lines within the plane</u> and I <u>want</u> the set of all points and lines to form a <u>lattice</u>.
That is <u>not</u> possible. Because if I take two parallel lines, then their intersection is the empty set, not a point.



l_1 l_2

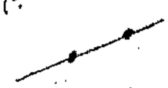
$l_1 \parallel l_2$
$l_1 \cap l_2 = \phi$

So that's not right.
But, any two lines which are not parallel intersect at a point.
So Kepler had this great idea. One of the greatest ideas he had.

He introduced points at infinity.
But, what's a point at infinity?

A <u>point at infinity</u> is an <u>equivalence class</u> of <u>parallel lines</u>.

This is the first occurrance of the notion of an equivalence relation in mathematics.
Kepler said this is an "ideal class."
There are equivalence relations among lines. Many.
And we say that two lines are equivalent if they are parallel.
So you include this class of points at infinity, which are equivalence classes of parallel lines.
Lo and behold, this has the same property as a point.
Why?
Two points determine a unique line. Fine.



l_1 defined by <u>2 points</u>

Now a <u>point</u> and a <u>point at infinity</u>, which is an equivalence class of parallel lines, and this is still a <u>unique line</u>.
A line with a given parallelism.
Now you have a lattice.
This was tremendous.

l_1 defined by <u>point</u> + parallelism (i.e., <u>point at infinity</u>)

You can extend this to n dimensions.
You can do it, but you get into a mess defining all the equivalence relations.
Instead of points at infinity, you have <u>lines at infinity</u>, <u>planes at infinity</u>, etc.
What a mess.
This was done by the geometers of the 19th century, with great care. There are reams of papers – a flurry.
Until some day someone came along and said – "Look, we can do it very easily. In a completely different way."
And I will tell you next time.

Projective Space (cont'd)

This is an idea used in algebraic geometry.
And you also see it in combinatorics. Recall that we are studying:

$V =$ finite dimensional vector space

$L(V) =$ lattice of subspaces of V
(linear subspaces through the origin, naturally)

$L(V)$ has the following properties:

1. every $W \in L(V)$ is the join of atoms
   $\hat{0}$ is the 0 subspace
   an atom is a line

2. $r(W) = \dim(W)$
   ↑ rank

3. $\dim(W \vee W') + \dim(W \wedge W') = \dim(W) + \dim(W')$

   This identity is also satisfied by measures of sets, with unions and intersections. However, measures of sets satisfy higher order identities, which are called inclusion-exclusion identities, whereas this metric only satisfies the second order – <u>NOT</u> the third order.

4. If W' is a complement of W (i.e., $W \vee W' = V$, $W \wedge W' = 0$) then:

$$\dim(W) + \dim(W') = \dim(V)$$

   Thus, the set of all complements of W is a non-empty anti-chain.

   Any two complements have the same dimension, therefore they can't be complements of each other.

   Last time I stated the <u>converse</u> of this was true. Let me state it more precisely this time. (I left out an assumption in the von Staudt – von Neumann theorem last time).

$L =$ lattice where all chains have length $\leq n+1 < \infty$
So, the maximum length of a chain is $n+1$

Then it is a fact, which I won't do because it is too dull, that L can be written:

$L = L_1 \times L_2 \times \cdots \times L_K$

L is factored into products, in the sense of partially ordered sets,
where the L_i are irreducible,
L_i can not be factored

Any lattice can be factored into irreducible factors.

Theorem of von Staudt - von Neumann ([14.4] restated)

Von Neumann - was a professor at the Institute for Advanced Studies.
Later a member of the first Atomic Energy Commission.
He died in 1950.

von Staudt - one of the founders of projective geometry

Conversely, a lattice L where all chains have length $\leq n+1 < \infty$, is irreducible, has chain of length ≥ 4, and has the property that for every $x \in L$, the set of complements is a non-empty anti-chain

then L is isomorphic to the lattice of subspaces of a vector space over a field.

This is an extraordinary result.
Because, all you are given is the lattice structure and this one little fact,
And you pull out the whole field - addition, multiplication, division, 0, 1, everything.
Everything gets pulled out of the lattice.
No wonder it takes 30 pages to prove it. You have to construct the whole theorem in terms of the latest operations. Later on I'll give you the secret of how this theorem works.
In other words, I'll show you how to construct $+$ and $\times$ in terms of $\vee$ and $\wedge$.
There's a secret - discovered by von Staudt and rediscovered by von Neumann.

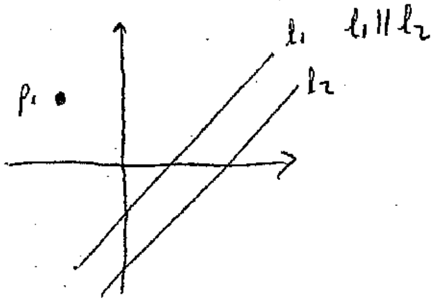
This is one of the most magnificent results ever obtained in combinatorics.
You characterize vector spaces of a field in terms of lattices.
In principal, you should be able to do everything you do with vector spaces using only the lattice operations.

Namely all of linear algebra, geometry should be encryted in the lattice.
Doing geometry using only the lattice operations is what is known as synthetic geometry.

So, the lattice of subspaces is a wonderful lattice.
The Theorem of von Staudt – von Neumann tells us that we should be able to do geometry only using $\vee$ (joins) and $\wedge$ (meets).
But we don't have a good way of visualizing that geometry.
The way of visualizing that geometry is projective space.

Last time we saw how to define a projective plane.
Let's review.

Projective plane



You want to make the points and lines, not necessarily through the origin, into a lattice.
You want to do synthetic geometry. Build triangles, stuff like that.

But, the problem is that two parallel lines have meet of the empty set. So the dimension axiom is not valid, because parallel lines intersect at the empty set.
So, although the set of points and lines is a lattice, it's a badly behaved lattice.

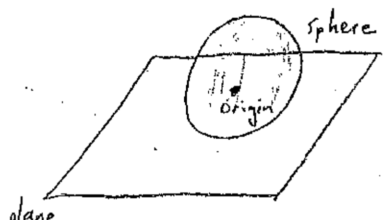
You are obliged, as we started last time, to add points at infinity so that the dimension condition is still true.
That is done by saying:

Equivalence classes of parallel lines are called points

Because I say so.
These are visualized as points at infinity.

By adding these points at infinity, then the dimension axiom is true.
Two lines always meet at a point.

Then I said this can be debunked.
Sure it can be debunked. Watch this.



We're in 3 dimensional space.
The lattice of subspaces of 3 dimensional space consists of all spaces through the origin, identified in the drawing
You can visualize that that is a subspace by intersecting it with a sphere.
In that way, a line becomes a set of 2 opposite points.

You can visualize the lattice of subspaces by identifying the appropriate parts of the sphere. That's what topologists do.

But there's a better one:
I take a line and I intersect it with a plane. I identify a line with this intersection.
Then we see that a line that is parallel to the plane will correspond at the point of infinity.
At the point of infinity, there will be exactly one line, corresponding to the <u>great circle</u>, parallel to the plane.

In this way, we have an interpretation of the projective plane.
It is simply the lattice of all subspaces of a 3 dimensional vector space.
Where we say that a line is a point.
And this works for any number of dimensions.
You take an $n+1$ dimensional sphere and project everything. And in that way we get n dimensional projective space.
We can visualize $L(V)$ as points, lines, and planes.
Points will be assigned rank of 0, because you lower the dimension by 1.

So that's what a projective space is. It's the central projection of an $n+1$ dimensional sphere onto a hyperplane. The hyperplane is the projective space.
Every point is given by $n+1$ coordinates – the coordinates of a line.

A point in projective space of dimension n is an equivalence class of $n+1$ tuples of numbers

$$(x_0, x_1, \ldots, x_n)\, R\, (x_0', x_1', \ldots, x_n')$$

whenever there is a number $\lambda \neq 0$ for which:

$$x_i' = \lambda x_i \quad , \; i = 0, 1, \ldots n$$

$\leftarrow$ in other words, the vectors are <u>proportional</u>.

A single point in n space is given by infinitely many coordinates.
The points with $x_0 = 0$ are at infinity. There is always a hyperplane of one dimension lower.
If you want ordinary cartesian coordinates, you take those with $x_0 = 1$ $\leftarrow$ i.e., points not at infinity.

This is a great advantage – to use an extra coordinate which may be 0.

For example, let's check that 2 lines always meet.

$$\left.\begin{aligned} 3x + 2y &= 5 \\ 3x + 2y &= 4 \end{aligned}\right\}$$

these are, in ordinary coordinates, parallel lines.

Now, let's use projective coordinates.

Set $x = \frac{x_1}{x_0}$, $y = \frac{x_2}{x_0}$, where x_0 = point at infinity

Then the lines become:

$$\left.\begin{aligned} 3x_1 + 2x_2 &= 5x_0 \\ 3x_1 + 2x_2 &= 4x_0 \end{aligned}\right\}$$

Now you see that these 2 equations do have a common non trivial solution. With $x_0 = 0$, at infinity. Where the lines meet.
And that's how it works.

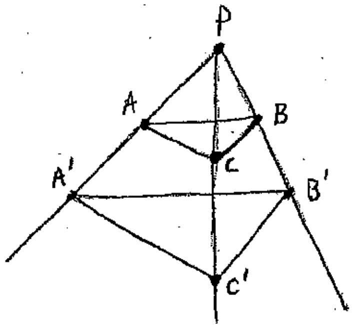
In projective geometry, you don't even have to tell whether a point is at infinity or not. It works at the point of infinity like any point.
If you are working w/ lattices, you don't even know which one is the point of infinity.

This point of infinity originated from the idea of perspective, which originated in the Renaissance.
The first painter ever to use this concept was the Italian painter Paolo Cellor
Then it was developed by Leonardo da Vinci.
Then Desargues and Kepler, in the foundations of projective geometry.

Let's now state the Fundamental Theorem of Projective Geometry:

- Desargues' Theorem

In $\mathbb{R}^3$, we have 3 lines and two triangles.



If the lines $A \vee A'$, $B \vee B'$, $C \vee C'$ pass through one point P then the points

$$(A \vee B) \wedge (A' \vee B')$$
$$(A \vee C) \wedge (A' \vee C')$$
$$(B \vee C) \wedge (B' \vee C')$$

lie on one line.

Zen proof:

Take the plane spanned by ΔABC and the plane spanned by $\Delta A'B'C'$.
By the dimension axiom, two planes meet in one line.
Therefore these two planes (i.e., the ones spanned by ΔABC and $\Delta A'B'C'$) meet in one straight line.
But, the intersection of $A \vee B$ and $A' \vee B'$ lies on both planes
Therefore, it must lie on that line.
So do the others.
And that's the end of the proof.

This proof I have given you uses only the dimension axiom.
It's an intuitive proof. Once you get it, you can't forget it.

Now let's give an analytic proof, using coordinates.
I really rub it in.

Analytic Proof of Desargues' Theorem

We are in 3 dimensional space, so the point A has 4 coordinates.
These 4 coordinates are determined to be linear multiplicative compounds, as they say, which is an equivalence class.
The same point is given by an infinity of multiples, which are proportional.

So, the point P is a linear combination of A and A'.
But we can work the constant of the linear combination of A and A' into the coordinates of the rest,
So we can say:

$$P = A + A' = B + B' = C + C'$$

Now we want to find the coordinates of the intersection $A \vee B$ and $A' \vee B'$.
How are we going to do that?
Easy.
There is one, and only one, point which is a linear combination of $A \vee B$ and a linear combination of $A' \vee B'$. That's the intersection:

$$\left.\begin{aligned} A - B &= B' - A' \\ \text{Similarily}\quad C - A &= A' - C' \\ B - C &= C' - B' \end{aligned}\right\} \text{ these are the 3 points of intersection.}$$

We want to show that these 3 points of intersection lie on one line.
That means there is one linear combination of them, which is 0.
Add the above up. You get $0 = 0$.
That proves the theorem.

** Exercise 15.1

Give similar analytic proofs of Haiman's generalization of Desargues' Theorem.
↑ 1984 Thesis

- Recall, from last time [14.6], we stated the theorem:

  There is an isomorphism of the lattice $L(V)$ into a sublattice of the lattice $\underbrace{\Pi[V]}_{\text{partitions of } V, \text{ viewed as a set.}}$ given by $W \to R_W$

  I stressed that joins in the lattice of subspaces correspond to joins in the lattice of partitions.

  $$R_W \circ R_{W'} = R_{W'} \circ R_W$$

  The sublattice of the lattice of partitions, which is the image $L(V)$, under this isomorphism, is a <u>linear lattice</u> – a lattice of commuting equivalence relations.

  $L(V)$ is a lattice of commuting equivalence relations.

- Suppose you have two partitions $\pi, \pi' \in \Pi[S]$.
  How is $\pi \vee \pi'$ defined?

  $s, t \in S$

  Say that [redacted] $s R_{\pi \vee \pi'} t$ whenever there is a sequence $s_1, \dots, s_n$ and a sequence $t_1, \dots, t_k$ where:

  $$\left.\begin{array}{l} s R_\pi s_1, \overset{\text{and}}{} s_1 R_{\pi'} s_2,\ s_2 R_\pi s_3,\ \dots,\ s_n R_\pi t \\ \text{and } s R_{\pi'} t_1,\ t_1 R_\pi t_2,\ t_2 R_{\pi'} t_3,\ \dots,\ t_k R_{\pi'} t \end{array}\right\} \text{ and, dually, when all } \begin{array}{l} R_\pi \to R_{\pi'} \\ R_{\pi'} \to R_\pi \end{array}$$

  You're taking the intersection of two Boolean algebras.
  You need to take the smallest blocks that contain blocks of π and blocks of π'.
  You have to go around π and π'.

  In particular, if π and π' commute, then $s R_{\pi \vee \pi'} t$ iff there exists s_1 and t_1 s.t.

  $$\left.\begin{array}{l} s R_\pi s_1,\ s_1 R_{\pi'} t \\ \text{and } s R_{\pi'} t_1,\ t_1 R_\pi t \end{array}\right\} \text{ in } \underline{\text{one}} \text{ step}$$

  Pictorially:

  $$\begin{tikzcd} t_1 \arrow[r, "t_1 R_\pi t"] & t \\ s \arrow[u, "s R_{\pi'} t_1"] \arrow[r, "s R_\pi s_1"'] & s_1 \arrow[u, "s_1 R_{\pi'} t"'] \end{tikzcd}$$

- ** Exercise 15.2

  Help me finish my paper on this.

So, the lattices of subspaces of projective geometry can be also visualized by the language of commuting equivalence relations.

Theorem of B. Jónsson

We've seen that a lattice of projective geometry is isomorphic to a lattice of commuting equivalence relations.
And it's proved that, for this lattice, Desargues' Theorem holds.

Desargues' Theorem holds in every linear lattice.

So Desargues' Theorem has nothing to do with geometry.
It's a purely combinatorial fact. It's about equivalence relations.
Remember that linear lattices are a dime a dozen.
This is an extraordinary discovery.

This is the deepest theorem that we will prove, so far.
The proof will take us half an hour. I want to give you the whole proof.

As a matter of fact it has been discovered very recently that just about every theorem of projective geometry also holds in linear lattices.
That puts projective geometry in a very difficult situation. Where's the geometry?
It's all purely combinatorial.

Last time, I announced the fact that Desargues' Theorem, which looks so much like a theorem of geometry, has, in reality, nothing to do with geometry.
The analogue of Desargues' Theorem holds in every lattice of commuting equivalence relations, also known as linear lattices.
Our job now is to state this fact correctly and prove it.
And that will be the end of this chapter.
This raises the following question.
What do you want me to do next? I've made a list of 10 topics which can come next. We can have a show of hands as to which ones are the favorite topics. Again, I've made it an example not to deal with any topics that I explain in my book, because if you can read about it in this book, what's the point of the lecture?
Everybody votes, as many times as you care to.
(Our goal is to get to 3 topics)

1st vote	2nd vote	Possible topics to be treated
8	6	1. More lattice theory
13	19	2. Matroid theory and matching theory
0	—	3. Basic results on convexity (including linear and integer programming)
9	8	4. Theory of species (a fashionable contemporary theory)
12	10	5. The Umbral Calculus
10	13	6. Möbius functions ← (This would be a real challenge for me, as I'd have to do things not in Stanley's book. Completely differently.)
6	—	7. The profinite point of view
12	12	8. Geometric probability
2	—	9. Greene's Theorem
8	7	10. Homology of posets

the agenda for the rest of the term:

1. Matroid theory and matching theory
2. Geometric probability
3. Möbius functions
4. Umbral Calculus ← (I'll keep the Umbral Calculus for last. If I don't get to it, I'll do it next term. The course is called Multilinear Algebra.)

Jónsson's generalization of Desargues' Theorem

Let L = linear lattice ← that means it's a lattice of commuting equivalence relations. and it's a sublattice of the lattice of partitions $\Pi[S]$. Sublattice means the joins and meets are the same as the joins and meets of partitions.

If partitions $\pi, \sigma \in L$,

$$R_\pi \circ R_\sigma = R_\sigma \circ R_\pi$$

By definition of a linear lattice.

We proved, a long time ago, that two equivalent relations commute iff their composition is an equivalence relation [6.11].
And we showed that this equivalence relation is the join [14.7]

$$R_{\pi\vee\sigma} = R_\pi \circ R_\sigma = R_\sigma \circ R_\pi$$

When two equivalence relations commute, their join is simply their composition.
This is what makes linear lattices tick.

I want to write out exactly what this means, in combinatorial terms.

Let $s, s' \in S$

$$s R_{\pi\vee\sigma} s' \iff s R_\pi \circ R_\sigma s' = s R_\sigma \circ R_\pi s'$$

The following 2 conditions have to be satisfied:

These 2 conditions together are equivalent to the equivalence relations commuting.

1. $s R_\pi \circ R_\sigma s'$:

   There is a $t \in S$ s.t.

   $s R_\pi t$ and $t R_\sigma s'$

2. $s R_\sigma \circ R_\pi s'$:

   There is a $u \in S$ s.t.

   $s R_\sigma u$ and $u R_\pi s'$

So $s R_{\pi\vee\sigma} s'$ iff conditions 1 and 2 above hold.
The whole art of working with linear lattices is being able to exploit this.

Let's warm up to Jónsson's Theorem by first establishing the modular identity of linear lattices.

The modular identity

$$\text{If } \alpha \geq \gamma \text{ the } \alpha \wedge (\beta \vee \gamma) = (\alpha \wedge \beta) \vee \gamma$$

A lattice that satisfies this statement (the modular identity) is said to be modular.
It just so happens that, in real life, all linear lattices known to man are modular.
The only modular lattice known to man that is not linear is the free modular lattice.

Theorem

The modular identity holds in all linear lattices.

Proof: (every equality is 2 inequalities. We prove = by first proving $\geq$ and then $\leq$)

1. Actually, the inequality:

$$\alpha \wedge (\beta \vee \gamma) \geq (\alpha \wedge \beta) \vee \gamma, \text{ for } \alpha \geq \gamma$$

holds for all lattices.

$$\left.\begin{array}{r}\text{By definition, we know } \alpha \geq \alpha \wedge \beta \\ \text{we are given } \alpha \geq \gamma\end{array}\right\} \alpha \geq (\alpha \wedge \beta) \vee \gamma$$

By definition of sup ($\vee$), α has to be $\geq$ than the sup (lowest upper bound) of these two.

$$\left.\begin{array}{l}\text{A moment's thought shows that } (\beta \overset{\text{sup}}{\vee} \bullet) \geq (\beta \overset{\text{inf}}{\wedge} \bullet)\text{:} \\ \text{Thus, we know that } \beta \vee \gamma \geq \alpha \wedge \beta \\ \text{By definition, we know } \beta \vee \gamma \geq \gamma\end{array}\right\} \beta \vee \gamma \geq (\alpha \wedge \beta) \vee \gamma$$

For the same reason as above (i.e., definition of sup).

This gives $\alpha \geq (\alpha \wedge \beta) \vee \gamma$
$\beta \vee \gamma \geq (\alpha \wedge \beta) \vee \gamma$

By definition of inf ($\wedge$):

$$\alpha \wedge (\beta \vee \gamma) \geq (\alpha \wedge \beta) \vee \gamma$$

We've just shown that this inequality is true for all lattices, using only definitions of sup and inf.

One lecture that I skipped was the general theory of inequality in lattices. Since you didn't want yet more lattice theory, then we skip it. You'll never know.

2. So now we have to prove:

$$\alpha \wedge (\beta \vee \gamma) \leq (\alpha \wedge \beta) \vee \gamma, \text{ for } \alpha \geq \gamma$$

Now we have to roll up our sleeves.
Now you see how it works. The real McCoy.
I've built up the whole term to this point. To get you to understand this part.
To get you through B. Jónsson's Theorem, which is a maze of reasoning.

$s, s' \in$ linear Lattice

If $sR_{\alpha \wedge (\beta \vee \gamma)} s'$, then we want to prove that $sR_{(\alpha \wedge \beta) \vee \gamma} s'$ is the same or a bigger relation.

$$sR_{\alpha \wedge (\beta \vee \gamma)} s' = sR_{\alpha} s', \underbrace{sR_{\beta \vee \gamma} s'}$$ (by definition of meet of 2 relations)

Recall the conditions for $R_{\beta \vee \gamma}$ where we have commuting equivalence relations [16.2]. Now, we use it. By definition of join, which is composition if they commute:

$$sR_{\beta \vee \gamma} s' = \boxed{sR_{\beta} u}, \boxed{uR_{\gamma} s'} \text{ for some } u$$
$$sR_{\gamma} t, \; tR_{\beta} s' \text{ for some } t$$

Since $\alpha \geq \gamma$ (given):

$$\boxed{uR_{\gamma} s'} \leq uR_{\alpha} s'$$

$$\leq \underbrace{sR_{\alpha} s', uR_{\alpha} s'}, \boxed{sR_{\beta} u}, \boxed{uR_{\gamma} s'}$$

by the transitive law:

$$\underbrace{sR_{\alpha} u, \boxed{sR_{\beta} u}}, \boxed{uR_{\gamma} s'}$$

by definition of meet:

$$\underbrace{sR_{\alpha \wedge \beta} u, \boxed{uR_{\gamma} s'}}$$

by definition of join:

$$sR_{\alpha \wedge (\beta \vee \gamma)} s' \leq sR_{(\alpha \wedge \beta) \vee \gamma} s'$$

$$\alpha \wedge (\beta \vee \gamma) \leq (\alpha \wedge \beta) \vee \gamma$$

That gives us two inequalities that, combined, give the equality we sought to prove.

part 1 : $\alpha \wedge (\beta \vee \gamma) \geq (\alpha \wedge \beta) \vee \gamma$

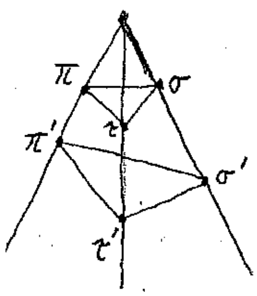
part 2 : $\alpha \wedge (\beta \vee \gamma) \leq (\alpha \wedge \beta) \vee \gamma$

$\alpha \wedge (\beta \vee \gamma) = (\alpha \wedge \beta) \vee \gamma$ for all linear lattices

Q.E.D.

That proves the theorem.
You ain't seen nothing yet. Let's prove Desargues' Theorem.
Remind me to discuss the Modular Law philosophically next time. What it's about. Where did it come from.

Desargues' Theorem for linear lattices



We have the assumption that these 3 lines meet at one point.
How do we say that lattice theoretically?
Easy.
That means the meet of two lines is contained in the 3rd line.

This is Desargues' Theorem:

Assuming $(\pi \vee \pi') \wedge (\sigma \vee \sigma') \leq (\tau \vee \tau')$,

we prove:

$$(\pi \vee \sigma) \wedge (\pi' \vee \sigma') \leq \big((\pi \vee \tau) \wedge (\pi' \vee \tau')\big) \vee \big((\sigma \vee \tau) \wedge (\sigma' \vee \tau')\big)$$

Let's start with the LHS of what we want to prove. We'll do this in columnar form. Suppose:

$$s\, R_{(\pi \vee \sigma) \wedge (\pi' \vee \sigma')}\, s'$$

unwinding the meet:

$$s\, R_{\pi \vee \sigma}\, s' \qquad s\, R_{\pi' \vee \sigma'}\, s'$$

unwinding the join:

$$(*)\ s\, R_{\pi}\, u, \quad (+)\ u\, R_{\sigma}\, s' \qquad (**)\ s\, R_{\pi'}\, v, \quad (++)\ v\, R_{\sigma'}\, s'$$

by transitive law:

$$u\, R_{\pi \vee \pi'}\, v \qquad u\, R_{\sigma \vee \sigma'}\, v$$

by definition of meet:

$$u\, R_{(\pi \vee \pi') \wedge (\sigma \vee \sigma')}\, v$$

↑ this is exactly the LHS of the assumptions
End of Act 1.

Let's next start with the RHS of the assumptions:

$$u R_{\tau \vee \tau'} v$$

unwinding join:

$$u R_{\tau} w \qquad w R_{\tau'} v$$

from previous page $\longrightarrow$ $(*)\ s R_{\pi} u \quad (+)\ u R_{\sigma} s' \qquad (**)\ s R_{\pi'} v \quad (++)\ v R_{\sigma'} s'$

by definition of join:

$$s R_{\pi \vee \tau} w \qquad w R_{\tau \vee \sigma} s' \qquad s R_{\pi' \vee \tau'} w \qquad s' R_{\sigma' \vee \tau'} w$$

by definition of meet:

$$s R_{(\pi \vee \tau) \wedge (\pi' \vee \tau')} w \qquad w R_{(\tau \vee \sigma) \wedge (\tau' \vee \sigma')} s'$$

by definition of join:

$$s R_{((\pi \vee \tau) \wedge (\pi' \vee \tau')) \vee ((\tau \vee \sigma) \wedge (\tau' \vee \sigma'))} s'$$

↑ this is exactly the RHS of what we needed to prove.

Q.E.D.

So now you see what a non trivial theorem looks like.

- There are two other major theorems that I'll state next time.
  Bricard's Theorem, which is a statement about points and lines, like Desargues.
  And the Theorem of Pappus, which goes back to Greek times.

  For a long time, it was felt that Bricard's Theorem could not be dealt with in a linear lattice.
  I just got the paper last week from Catherine Yan where she generalizes it to a linear lattice.

  On the other hand, Pappus' Theorem can not be generalized to a linear lattice.
  That was discovered centuries ago. I'll tell you next time why. It was a great discovery by Hilbert.

- Back to the Modular Law:

$$\alpha \geq \gamma \implies \alpha \wedge (\beta \vee \gamma) = (\alpha \wedge \beta) \vee \gamma$$

We proved this in a linear lattice. So, by implication, this is true in $L(V)$.
However, $L(V)$ is a lattice of subspaces and these are joins and meets of subspaces.
So there should be a simple linear algebra way of seeing this.
That's what we'll do next.
Prove this in $L(V)$ using elementary linear algebra.

A digression: Remember the notion of an ortho complement [12.9].

Ortho complement in L is the map $x \to x^{\perp}$ where:

1. $x \leq y \Rightarrow x^{\perp} \geq y^{\perp}$
2. $x^{\perp\perp} = x$

$x^{\perp}$ is a complement of x

When is there an ortho complement in $L(V)$?
Answer – when you have the notion of perpendicularity.
And when do you have the notion of perpendicularity in a vector space?
When you have a bilinear form that gives you the dot product.

The conditions above hold if you have a dot product

$$x \cdot y = \underbrace{(x, y)}_{\text{also written as this}} = \sum_i x_i y_i \quad \text{is given,}$$

in which case

$$W^{\perp} = \{ y : x \cdot y = 0 \text{ for all } x \in W \}$$

subspace perpendicular to W

Whenever you have the dot product defined, you have the ortho complement.

Exercise 16.1

The exercise is the converse of this.

Theorem of Kakutani – Mackey

If $W \to W^{\perp}$ is an orthocomplement in $L(V)$
then there exists an inner product (x, y) in V for which

$$W^{\perp} = \{ y : (x, y) = 0 \text{ for all } x \in W \}$$

dot product $x \cdot y$

In other words, if you have an ortho complement, it forces you to find an inner product in the vector spaces

I would appreciate an elementary proof of this fact. The only proof in the literature is a complicated one. It would be interesting to get a self contained proof.

Again, it turns out that most of the theorems of projective geometry hold in linear lattices. There's a deep mystery there. Why do these theorems hold only for commuting equivalence relations?
Very weird.

Kultur

Before we start on matching theory, let's do some cultural topics.
In the German newspapers, you have sections national news, then international news, then sports, and then you have Kultur. But there is no equivalent word in English. There's one in Spanish, French, and Italian. Culture does not mean Kultur.
The origin of the word is Spanish. It was first introduced by Luis Vives.

Last time we saw that Desargues' Theorem is valid in all linear lattices.
Today, I want to show you how the Modular Law can be done by elementary linear algebra, as promised.

Modular Law (via linear algebra)

$$a \geq c \implies a \wedge (x \vee c) = (a \wedge x) \vee c \quad \text{for all } x \in L$$

Let's take $L = L(V) \leftarrow$ lattice of subspaces of a vector space

Let's verify the Modular Law by elementary linear algebra.
There's probably some easier way than what I'm about to do.
If there is, raise your hand.

a, x, c are subspaces of a vector space

So we take basis of a, x, and c and we reason with bases.
The basic fact is that when you have a subspace and a subsubspace, then any basis of a sub subspace can be completed to the basis of a subspace.
That's all you've got.
So anything you can squeeze out of this is true.

You can squeeze anything you can say about 2 subspaces, but nothing about 3 subspaces (unless one is contained in the other). That's your problem.

	basis of vectors in V	
$x \wedge c$	$\{v_1, \dots, v_i\}$	Let's say this is the basis for $x \wedge c$
$x \wedge a$	$\{v_1, \dots, v_i, \underbrace{w_1, \dots, w_j}\}$	since $a \geq c$, a basis of $x \wedge a$ will have this extra stuff, compared with that of $x \wedge c$.
c	$\{v_1, \dots, v_i, c_1, \dots, c_k\}$	A basis of $x \wedge c$ can be completed to a basis of c.
a	$\{v_1, \dots, v_i, w_1, \dots, w_j, a_1, \dots, a_l, c_1, \dots, c_k\}$	completing $x \wedge a$ and noting $a \geq c$
x	$\{v_1, \dots, v_i, w_1, \dots, w_j, x_1, \dots, x_m\}$	
$x \vee c$	$\{v_1, \dots, v_i, w_1, \dots, w_j, x_1, \dots, x_m, c_1, \dots, c_k\}$	

Then we have:

$$\left.\begin{array}{ll} (x \vee c) \wedge a & \{v_1, \dots, v_i, w_1, \dots, w_j, c_1, \dots, c_k\} \\ (x \wedge a) \vee c & \{v_1, \dots, v_i, w_1, \dots, w_j, c_1, \dots, c_k\} \end{array}\right\} \quad a \wedge (x \vee c) = (a \wedge x) \vee c$$

So the Modular Law comes out in linear algebra.

Pappus' Theorem

If I have to reconstruct Pappus' Theorem, what do I do?
I do some secret computations, which I'll then erase, and then tell you the result.
Right

I'll tell you a story. There was a time in algebra when everything had to be done without a basis. Prof. Chevalier was brought up using matrices. Every once in a while, he'd get lost and he'd go to the corner and do the computation using matrices.
"Basis free" algebra.

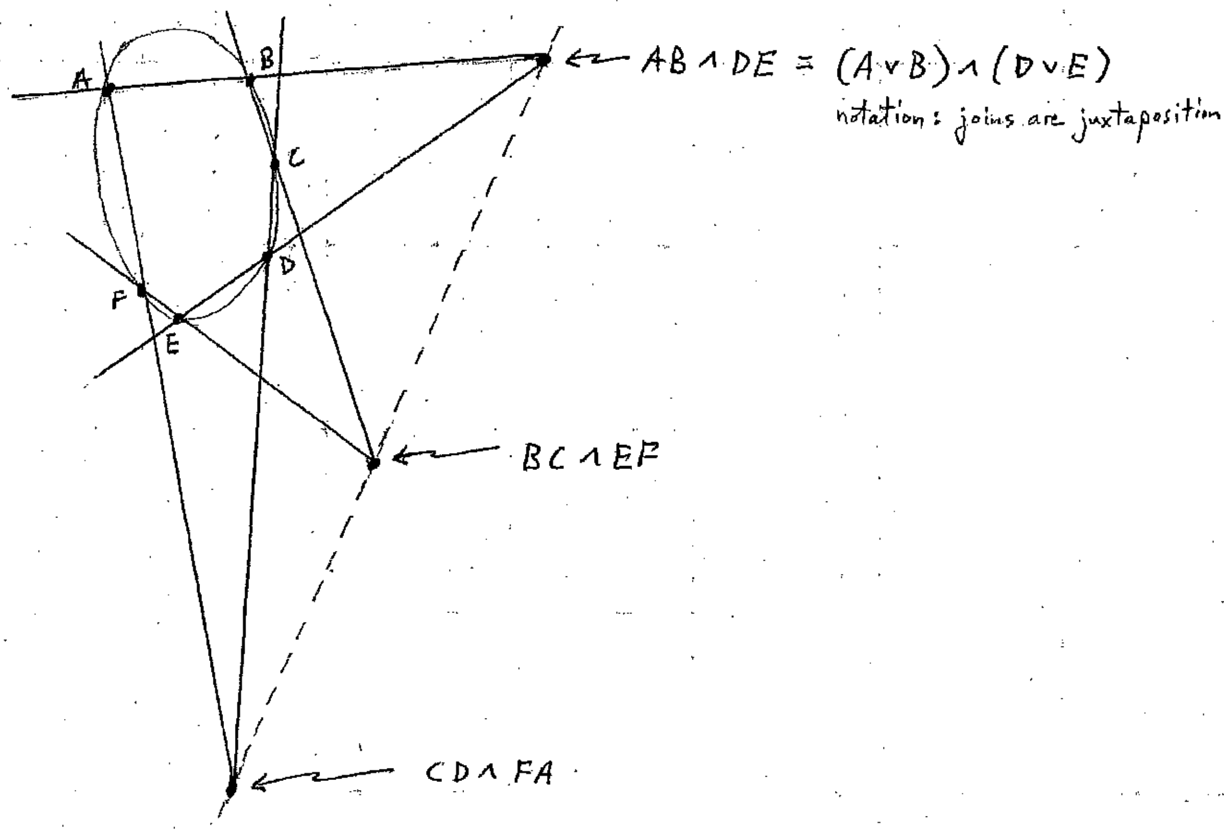
I tell you the truth. If I forget, how do I reconstruct it?
The secret reconstruction is:

Pappus' Theorem is a degenerate form of Pascal's Theorem.

And Pascal's Theorem is easier to remember.

Pascal's Theorem

Let me state this in old fashioned language.
You take a conic section and you take a hexagon inscribed in the conic section.



Pascal's Theorem: the 3 points $AB \wedge DE$, $BC \wedge EF$, $CD \wedge FA$ lie on a line

Now you say - "Prove it."
I say - "I don't remember the high school proof." I have to cheat again.

What is a conic section?
A conic section is the set of all points that satisfy a quadratic equation in homogeneous coordinates.

You have a homogeneous point x determined by 3 homogeneous coordinates:

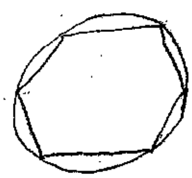
$$x = (x_0, x_1, x_2)$$

$$q(x_0, x_1, x_2) = a_{00}x_0^2 + a_{01}x_0x_1 + a_{02}x_0x_2 + \cdots + a_{22}x_2^2$$

The set of all points that satisfy $q(x_0, x_1, x_2) = 0$ is called a conic section.

So now I have to prove Pascal's Theorem.
And I say, from the point of view of projective geometry, one conic section is as good as another.
Because, by taking a change of coordinates, I can transform any conic section into any other.

Therefore, the assertion of Pascal's Theorem is invariant under linear changes of variables, since it's an assertion of projective geometry.
So you only need to prove it in one case.
So I take the cheapest possible case:



I take a circle ...
And inscribe a regular hexagon in it.

And then it's obvious. The 3 points $AB \wedge DE$, $BC \wedge EF$, $CD \wedge FA$ will meet at infinity, because they are parallel.

They all lie on the same line - the line at infinity.
This is true for this, so it's true for all of them.
That's it. End of the proof.

Except for one case.
Except for degenerate conic sections.

A conic section can be transformed into another conic section by a change of variable, if both conic sections are not degenerate.

Degenerate means that the quatric form:

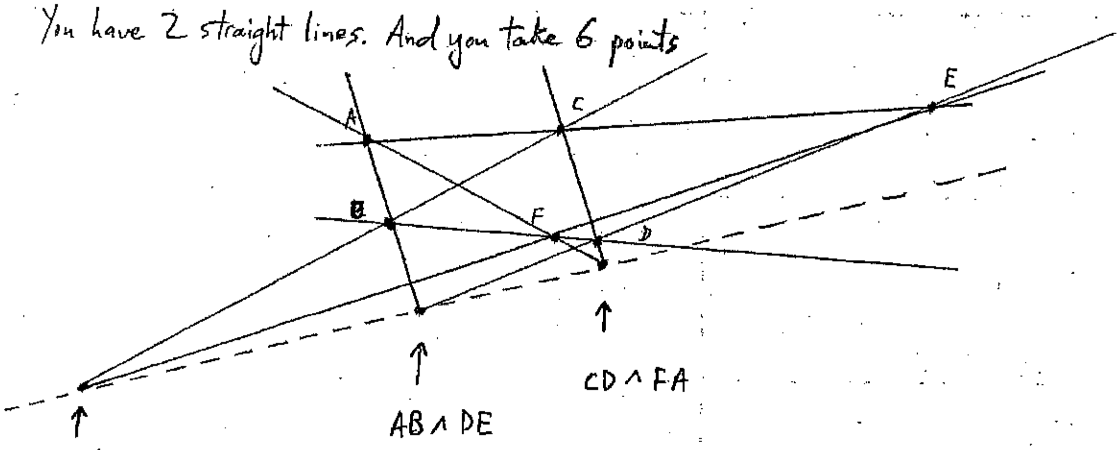
$$0 = q(x_0, x_1, x_2) = a_{00}x_0^2 + a_{01}x_0x_1 + a_{02}x_0x_2 + \dots + a_{22}x_2^2$$

is the product of 2 linear forms. That means the conic section is 2 lines.

By continuity, Pascal's Theorem remains true for degenerate conic sections (close an eye).
Pascal's Theorem for degenerate conic sections is Pappus' Theorem.
Now I remember Pappus' Theorem, But what I stated is not really true.
This continuity argument is phoney balony.
But at least I remember the statement.

Pappus' Theorem

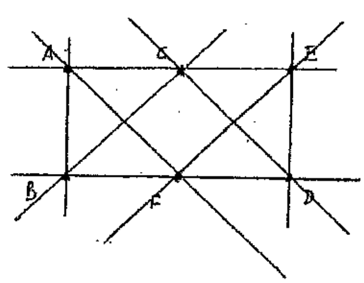
You have 2 straight lines. And you take 6 points



The 3 points $AB \wedge DE$, $BC \wedge EF$, $CD \wedge FA$ lie on a line.

But the proof we just gave, in spite of my phoney balony, is valid for non degenerat conic sections only.
So we need to prove this.
What do we do? We cheat.
How do I cheat?

We're in projective space. These lines don't know. They can be in any position I wish.
I can make changes of variables and place 3 points anywhere I wish.
So I now take the most favorable position that will give me a proof of Pappus' Theorem.



The points $AB \wedge DE$, $BC \wedge EF$, $CD \wedge FA$ meet at infinity, because they are parallel.
They all lie on the same line - the line at infinity.

This is the Theorem I was saying last time, Pappus' Theorem, that can be stated purely in lattice theoretic terms, using joins and meets.
You replace the points A, B, C, D, E, F with commuting equivalence relations.
But this statement is NOT true in every linear lattice.
Hilbert discovered:

Pappus' Theorem true in $L(V)$ only if V is a vector space over a commutative field.

He discovered this by analyzing von Staudt's reasoning very carefully.
So we now want to go through the main idea of the von Staudt - von Neumann Theorem [15.2].

von Staudt - von Neumann Theorem

Let's state it in this succinct way:

L = linear lattice

If every $x \in L$ is the sup of atoms and, for every $x \in L$, the set of complements is a non empty antichain and L is large enough ← (so you don't get a plane. You want something bigger than a plane so you can get Desargues' Theorem.)

then $\underbrace{L = L(V)}$.

then L is isomorphic to the lattice of subspaces of a vector space.

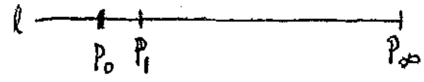
Key Ideas

Let's see the key ideas of the proof of Desargues' Theorem. A proof tour.
There's a whole volume of this - Baker's Principles of Geometry.

The key idea is this - how do you get addition and multiplication in a field, using joins and meets?
This is the basic insight.

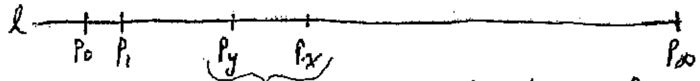
Let's see how to define addition.
We define addition on points on a line. Then we show that this addition of points on a line is commutative and associative.
This was the great turning point of geometry really - when they discovered you could do addition using joins and meets.

Addition - using joins and meets

(1) 

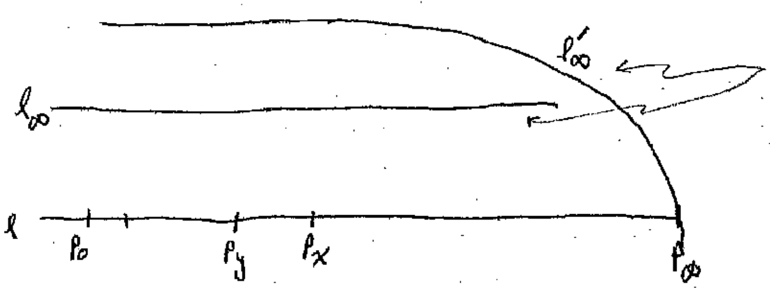
In order to define addition, you have to define which point is 0, 1, and infinity of your coordinate system.
Otherwise addition is not well defined.

(2)



We are given 2 points P_x and P_y. These are fixed forever.
I want to find P_{x+y}.

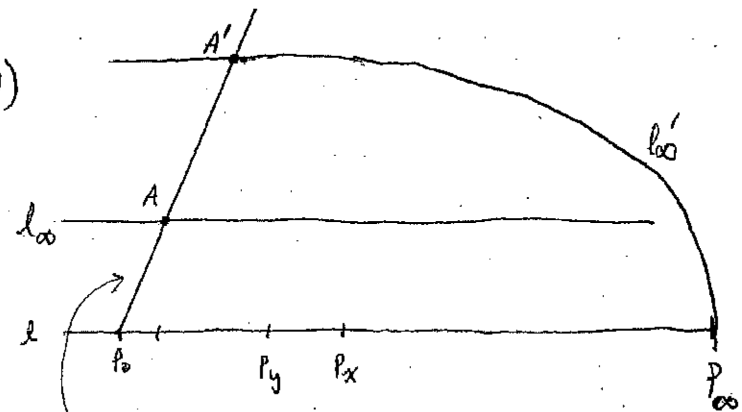
(3)



choose 2 lines l_∞ and l'_∞ through P_∞.

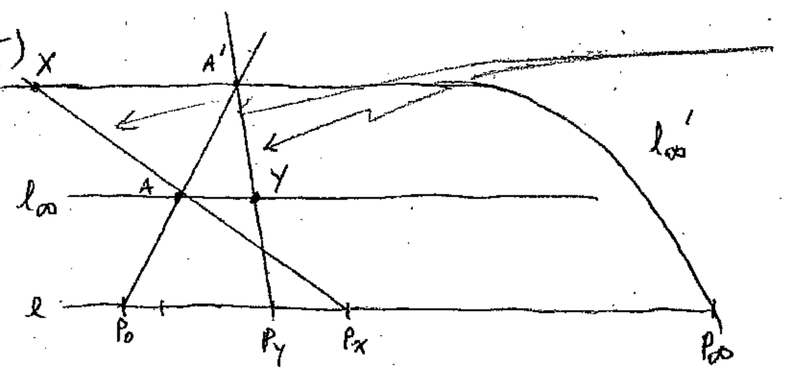
$l_\infty \parallel l'_\infty$

(4)



Take any line through P_0.
This line meets l_∞ at point A, and l'_∞ at point A'

(5)



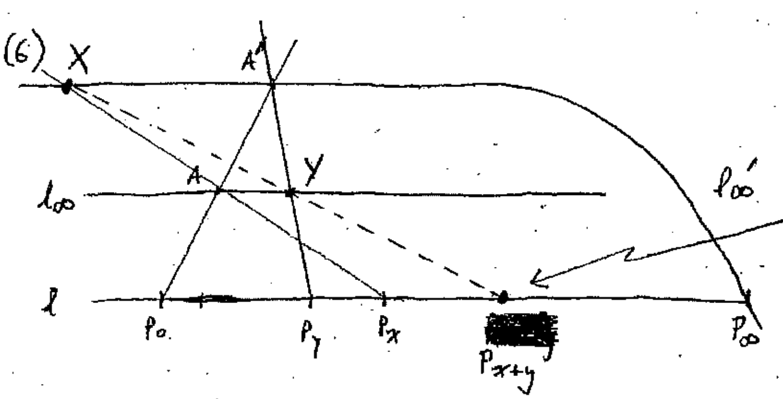
Construct lines $P_x \vee A$ and $P_y \vee A'$.

Set:

$X = (P_x \vee A) \wedge l'_\infty$

$Y = (P_y \vee A') \wedge l_\infty$

(6)



[redacted] $= (X \vee Y) \wedge l$
P_{x+y}

So P_{x+y} is really $x+y$, by construction.
Of course, you could do this without the point at infinity, but this becomes incomprehensible.
That's what the books do.

Where is the catch in this argument?
The catch in this argument is that it depends on the choice of the line through P_0 and l_∞ and l_∞' in step 4.
What if I took another line here?

So the theorem is that the point P_{x+y}, which you get at the end, is the <u>same</u>, no matter which line you choose in step 4. Provided it meets l_∞, l_∞'.
Why?

By certain applications of <u>Desargues' Theorem.</u>
Why is <u>addition</u> <u>commutative</u> and <u>associative?</u>
By certain other applications of <u>Desargues' Theorem.</u>

That's how addition comes out.

There is also a construction for <u>multiplication</u>, but I don't want to bother with it.
There's a similar construction for multiplication and you prove that <u>multiplication</u> is <u>associative</u> by 23 applications of <u>Desargues' Theorem.</u>
But, you can't prove multiplication is commutative.
To prove that <u>multiplication</u> is <u>commutative</u>, you need <u>Pappus.</u>
The first one to notice this was Hilbert.

That's the secret of the van Staudt - von Neumann Theorem.
This is it. Everything comes out of Desargues' Theorem.

Notice that we did not use P_1.
That's quite justified, as we only defined addition.
And you can not tell 1 from 0 by just + and −.
You have an Abelian group and it has a 0, which is the identity of the Abelian group.
To tell 1, you need the product. And I didn't do that.

<u>* Exercise 17.1</u>

In closing, let me give you another theorem of projective geometry, for which I do not know of an elementary proof. I know several non elementary proofs, using my tricks.
I would love it if you could get a high school proof of this.

Find a high school proof of Bricard's Theorem.

By the way, Bricard's Theorem, as I told you, has recently been proved by Catherine Yan to hold in all linear lattices.
It's a theorem about tetrahedra in space.

Bricard's Theorem

Given 2 tetrahedra $abcd$, $a'b'c'd'$ in 3 dimensional space.
Consider the following intersections:

$$aa' \wedge bcd = p_1$$
$$bb' \wedge acd = p_2$$
$$cc' \wedge abd = p_3$$
$$dd' \wedge abc = p_4$$

juxtaposition means join

The points p_1, p_2, p_3, p_4 are coplanar iff the following 4 planes meet at 1 point:

$$\left.\begin{aligned} &(bcd \wedge b'c'd') \vee a' \\ &(acd \wedge a'c'd') \vee b' \\ &(abd \wedge a'b'd') \vee c' \\ &(abc \wedge a'b'c') \vee d' \end{aligned}\right\}$$

plane $\wedge$ plane $\Rightarrow$ line
line $\vee$ point $\Rightarrow$ plane

I'd love to have a high school proof.

• Lattice theory is a very hard topic to tell you about in advance, w/o lying, because it's a really very broad subject and leads naturally into matroid theory.
So I will tell you just some bits in order to start with.
But you will see that it soon branches off into completely different and unexpected directions, which are projected by the very problems we saw.

It's a very rich and deep chapter of combinatorics.
We will have to go through a number of very delicate proofs.

Linear lattices lead naturally to normal subgroups, subgroups, ideals of a ring, subspaces of a vector space — all these separate fields.
Some day, people will develop the theory of linear lattices to the point where you only talk about linear lattices and you won't talk about these other things. The level of generality of linear lattices will be the right one.

• Matching Theory and Matroids (beg'g)

I'll tell you the dishonest definition of matching theory, which you will find in the books.

Given a relation $R \subseteq S \times T$ another relation $R' \subseteq S \times T$ is a partial matching (or partial transversal) when R' is a partially defined 1-1 function (isomorphism).

R' looks like :

In it's simplest form, matching theory is about the following questions:

Q: When does $R \geq R'$, where R' is a matching?
and How big can R' be?

The best possible situation would be that R' is a matching that is everywhere defined on S. In which case we say that R contains/has a matching.
If it doesn't have a matching, what's the maximum partial matching you can have?
And how do you determine this number?

That's the 1st approximation to matching theory, but I warn you this is just the beginning. The real interesting questions come later.

Let's proceed systematically. I've pieced together the theory from various sources, as general and systematically as possible. This dovetails naturally into matroid theory, without your even knowing it.

S, T finite sets.

We have for $A \subseteq S$, $|A|$ is a <u>measure</u>, namely:

$$|A \cup B| + |A \cap B| = |A| + |B|$$
$$|\phi| = 0$$

Any other function from sets to $\mathbb{R}$, with these properties, is called a measure, as we've defined it before [8.10].

In general,
<u>Set function</u> = function from sets to $\mathbb{R}$

This is totally deceiving the way it is generally used.
<u>Not</u> a function whose <u>values</u> are <u>sets</u>, as the term might indicate.
But functions from <u>sets to $\mathbb{R}$</u>.

Very little is known about set functions that are not measures.
That is what we will be up against.

family of all subsets of S

A set function μ defined on $P(S)$ is <u>submodular</u> when:

$$\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B), \text{ for all } A, B \subseteq S$$

Now I could redefine <u>matching theory</u> as the study of <u>submodular set functions</u>.

<u>Example</u>:

Recall that [3.2]:

$R(A \cup B) = R(A) \cup R(B)$ where $R(A) = \{b \in T : (a,b) \in R \text{ for some } a \in A\}$

$R(A \cap B) \subseteq R(A) \cap R(B)$

Set $\mu(A) = |R(A)|$

This is a <u>submodular set function</u> because:

$$\begin{aligned}\mu(A \cup B) + \mu(A \cap B) &= |R(A \cup B)| + |R(A \cap B)| \\ &\leq |R(A \cup B)| + |R(A) \cap R(B)| \\ &= |R(A)| + |R(B)| \\ &= \mu(A) + \mu(B)\end{aligned}$$

$$\therefore \mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B)$$

** Exercise 18.1

There is an interesting open question which ought to have been worked out.
And that I ought to have worked out, but I haven't.
Namely:

Characterize those submodular set functions that come from a relation in this way;

I've never really worked this out.
Roughly speaking, you have to satisfy a series of inclusion-exclusion inequalities.
That's a necessary and sufficient condition.
No one has written this out properly.

There are a tremendous number of submodular set functions, as we will see.
An enormous variety.
Now let's consider the submodular set function that will concern us in order to study the matchings of a relation.
That's called the deficiency of a relation.

The deficiency of R, say δ, is the set function:

$$\delta(A) = |R(A)| - |A|$$

δ is submodular.
Why? Because $|R(A)| = \mu(A)$ is submodular and minus the elements of A is modular.
A submodular plus a modular is submodular.

Tight set

A tight set is a set of minimum deficiency.

A subset $A \subseteq S$ for which $\delta(A)$ takes it's minimum value, say δ_0, is a tight set.

Observe that the deficiency of the null set is 0:

$$\delta(\phi) = |R(\phi)| - |\phi| = 0,$$

hence $\delta_0 \leq 0$

Now we have the first of a number of interesting theorems due to Ore:

Theorem 1

If A and B are tight sets, then so are:

a) $A \cap B$
b) $A \cup B$

Proof:

$$\delta(A \cup B) + \delta(A \cap B) \le 2\delta_0 \longleftarrow \begin{cases} \delta(A \cup B) \le \delta_0 \\ \delta(A \cap B) \le \delta_0 \end{cases}$$

Neither of these can be smaller than δ_0, because δ_0 is the minimum deficiency. If one is larger than δ_0, the other must be smaller than δ_0. But that can not be. Therefore:

$$\delta(A \cup B) = \delta(A \cap B) = \delta_0$$

Let me tell you an interesting fact, which we'll eventually squeeze to death. Tight sets form a distributive lattice.
So there's a minimum tight set (the intersection of all tight sets) and a maximum tight set (the union of all tight sets).

Corollary:

There is a minimum tight set N and a maximum tight set M.

N could be ϕ, of course

I haven't found much use for the maximum tight set. Perhaps you can find some.

Theorem 2

complement of the minimum tight set (A is disjoint from the minimum tight set)

If $A \subseteq N^c$ then $\delta(A) \ge 0$

Proof:

$$\delta(A \cup N) + \cancel{\delta(A \cap N)}^{\,0} \le \delta(A) + \underbrace{\delta(N)}_{\delta_0}$$

this can not be smaller than the minimum deficiency

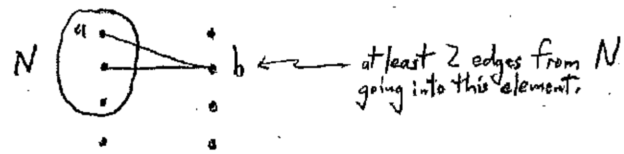
Since $A \cap N = \phi$, $\delta(A \cap N) = \delta(\phi) = 0$

$\delta(A \cup N) \ge \delta_0$

$\therefore\ \delta(A) \ge 0$

Theorem 3

Every element of $R(N)$ has marginals ≥ 2.



Proof:

Assume $b \in T$ has marginal 1, where $(a,b) \in R$, $a \in N$

It has to be at least 1, since $b \in R(N)$.
Let's assume it is exactly 1.
Thus, you have only one edge going to b. The one issuing from a.

Remove a from the minimum tight set:

$$|R(N-a)| \leq \underbrace{|R(N)| - 1}$$

at a minimum, you lose $b \in R(N)$.
If there are others with marginal $= 1$, that are related to a, then you lose these too.

By definition, the deficiency is:

$$\begin{aligned}
\delta(N-a) &= |R(N-a)| - \overbrace{|N-a|}^{|N|-1} \\
&\leq |R(N)| - 1 - (|N| - 1) \\
&= |R(N)| - |N| \\
&= \delta_0
\end{aligned}$$

So we have:

$$\delta(N-a) \leq \delta_0$$

But this contradicts that N is the minimum tight set, with the minimum deficiency.
Thus, our assumption is false.
b must have marginals ≥ 2.

Theorem 4

Let $R' = R - \{(a,b) \in R : b \in T\}$ for some $a \in N$

Then the minimum deficiency of R' equals $\delta_0 + 1$.

If you remove any point in the minimum tight set and remove all the edges issuing from this point, then the minimum deficiency of the remaining relation goes up by exactly 1.

Proof: Immediate from the preceding theorem (Theorem 3).

First, note that we've removed an element from the tight set:

$$|N - a| = |N| - 1$$

From Theorem 2, we note that every element of $R(N)$ has marginals ≥ 2. Thus, even after removing all the edges issuing from a, every element of the remaining relation remains covered.

$$|R'(N-a)| = |R(N)| \longleftarrow \text{unchanged}$$

$$\delta_{0_{R'}} = |R'(N-a)| - |N-a|$$

$$= \underbrace{|R(N)| - |N|}_{\text{this is the original } \delta_0} + 1$$

Therefore,

$$\delta_{0_{R'}} = \delta_0 + 1 \longleftarrow \text{increases by } 1$$

Theorem 5

Therefore, if we remove any $|\delta_0|$ points from N, we are left with a relation whose minimum deficiency is 0. It goes up by 1 each time we remove a point (Theorem 4) and we remove $|\delta_0|$.

Let $C \subseteq N$, $|C| = -\delta_0$ ← remember, $\delta_0 \le 0$.

Let $R'' = R - \{(a,b) : (a,b) \in R,\ a \in C\}$ ← from R, remove all vertices in C, as well as all attached edges.

Then the minimum deficiency of R'' equals 0.

Now, let's study for a while relations whose minimum deficiency is 0:

Let R be a relation whose minimum deficiency equals 0.

This means that for every $A \subseteq S$, we have:

$$|R(A)| \geq |A|$$

$$\left\{ \begin{aligned} \delta_0 &\le \delta(A) \\ &= |R(A)| - |A| \\ &\text{Given that } \delta_0 = 0: \\ &|R(A)| \geq |A| \end{aligned} \right\}$$

A relation satisfying this condition is said to satisfy the Hall condition.

↑ after Philip Hall

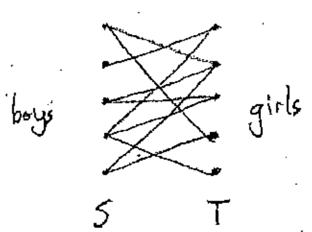
Theorem 6 - The Marriage Theorem

A relation R contains an everywhere defined matching R' iff it satisfies the Hall condition.

↑ everywhere defined 1-to-1 function

This is one of the most famous theorems of combinatorics. Before I prove it, let me give you some jazzy interpretations.
There are infinitely many applications of this Theorem.
Let's see a few.

The Classical Example (whence the name)

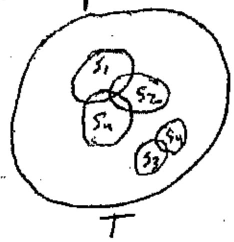


You have boys and girls.
And every boy knows some number of girls and every girl knows some number of boys.
And there is a dance. Ballroom dancing.
Then you want to match boys with girls so that every boy dances with a girl that he knows.

When is this possible?
It's possible if every subset of k boys collectively knows at least k girls, for every k.

The condition of the <u>Hall condition</u> is a great observation, because it's much easier to check the Hall condition than it is to find a matching.

Example: System of Distinct Representatives



Given a big set T and a family of subsets $S = \{S_1, S_2, \ldots, S_n\}$

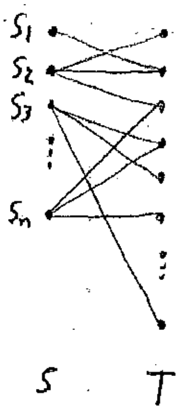
You consider these subsets as groups of people.

When can you find different (distinct) leaders for each group?

Want to pick, for each group, a leader so that the leaders are distinct. So the leader must be a member of the group.

This is called the <u>system of distinct representatives</u>.
You want 1 point as the representative of each subset.

This can be visualized immediately as a relation:



$R \subseteq S \times T$

edges represent the membership relations.

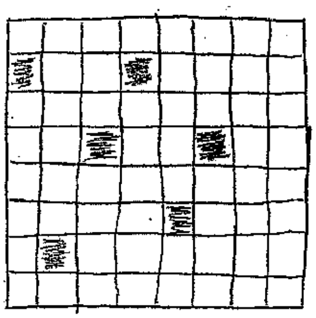
Since you have a relation, all you have to do is apply The Marriage Theorem to this relation and you have the <u>necessary</u> and <u>sufficient</u> condition for the <u>existence</u> of a system of distinct representatives.

Namely:

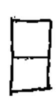
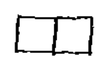
The union of any k sets must contain at least k elements

Since we don't have time for a non trivial example, let me give you a last trivial example and I'll give you a non trivial application next time.

Covering amputated chess board with dominos

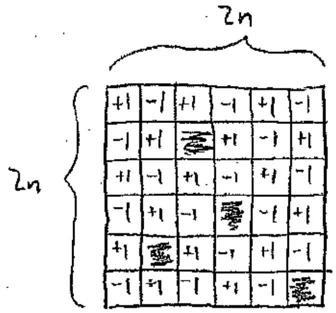


I have a chessboard.
I remove any number of squares at random (= ▩).
Then I have little domino pieces, each of which can be placed on the board vertically or horizontally. Each domino covers 2 squares:

 vertically  horizontally

When is it possible to cover the amputated chess board with domino pieces?

Let's suppose that the amputated chess board is a square of even sides.
Then you have to label adjacent squares +1 and −1.

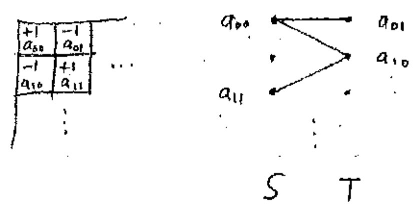


We form a relation, as follows:

$\boxed{+1} \in S$

$\boxed{-1} \in T$

If 2 points are adjacent, then there is an edge connecting them.

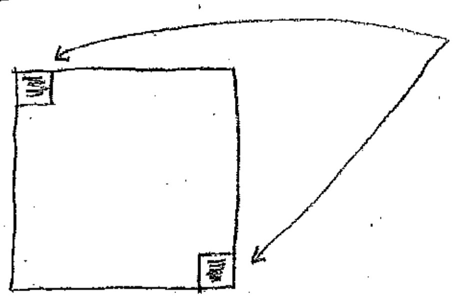


Now you see immediately that a covering by dominos is the same as a matching of this relation.

There is a covering by domino iff the Hall condition is satisfied.

You see immediately that if you cut out the two diagonal corners, the Hall condition is not satisfied.

Therefore,
You can not cover this board with dominos.



There is a famous story.
This was being explained to some big wig in Washington on the applications of mathematics.
And that person saying – "Yes, that's fine and dandy if you label the squares +1, −1.
What if you don't label them +1, −1? Then what happens?"
What do you say to that.

This actually happened, as I know the person to whom it happened.
It was to Professor Golomb of USC, who is the world's expert on covering chess boards.

Let me state a theorem that we will prove next time, as an immediate application of The Marriage Theorem.
And then we'll prove The Marriage Theorem. Then we'll go back to matching theory.

Birkhoff – von Neumann Theorem

This is in my book, by the way.
This stuff is done completely differently in my book.
You want another approach.

You take all $n \times n$ matrices, as follows:

$$\begin{array}{cc} & n \\ n & \left[\begin{array}{ccc} & & \\ & x_{ij} & \\ & & \end{array}\right] \begin{array}{l} \Sigma=1 \\ \Sigma=1 \\ \vdots \\ \Sigma=1 \end{array} \\ & \Sigma=1 \ \Sigma=1 \ \cdots \ \Sigma=1 \end{array} \qquad x_{ij} \geq 0$$

with the property that the <u>marginals</u> are <u>all</u> 1.
Namely, the sum of each row is 1.
And the sum of each column is 1.

A matrix with this property is called <u>doubly stochastic</u>.

So we consider the set of all doubly stochastic matrices.
That's the set of points in space of dimension n^2.
In fact, it's a convex polyhedron. It's a convex closed set in dimension n^2.
So the question is:

What are the <u>vertices</u> of this convex polyhedron?

The vertices are the points that are <u>NOT</u> <u>convex subrelations</u> of two other points (we'll define convex subrelations next time).

The Birkhoff – von Neumann Theorem tells you that the vertices of this polyhedron are <u>exactly</u> the <u>permutation matrices</u>.

Namely, the matrices, all of whose entries are 0 or 1.
Which means there is exactly one entry in each column and exactly one entry in each row.

This is an immediate consequence of The Marriage Theorem.
So next time, I'll give you Birkhoff's proof of this – which uses The Marriage Theorem.
And then, I can't resist the temptation of giving you von Neumann's never published proof that does <u>NOT</u> use The Marriage Theorem.

Matching Theory (cont'd)

Let's begin by reviewing the theory we began last time, which is the tip of the iceberg on matching theory.

Given $R \subseteq S \times T$ (finite)

generalization of this stuff to infinite sets is highly non trivial. There have been efforts in different directions – measure theoretic, trans finite, topological, etc.

Then we defined the deficiency of a relation:

$$\delta(A) = \delta_R(A) = |R(A)| - |A| \quad ; \quad A \subseteq S$$

It should really be written like this. The R (relation) is understood.

We have verified that the deficiency is a submodular set function. Namely:

$$\delta(A \cap B) + \delta(A \cup B) \leq \delta(A) + \delta(B)$$

This is the only example we've seen so far of a submodular set function.
I'll tell you now – there are lot's more coming up.
This is sort of a Micky Mouse example of a submodular set function.

Then we defined the minimum deficiency of the relation:

$$\delta_0 = \min_{A \subseteq S} \delta(A)$$

Sets B s.t. $\delta(B) = \delta_0$ are said to be tight.

Then we proved a number of theorems due to the Norwegian Øystein Ore (there are lots of Norwegians in this field, as you will see).

Theorem 1: If A and B are tight sets then so are $A \cup B$ and $A \cap B$.

The family of tight sets of a relation forms a distributive lattice.
There is, then, a minimum tight set (the intersection of all tight sets in the lattice).

Theorem 4: Let N be the minimum tight set of a relation.
If you remove any point in N and all the edges issuing from that point, you get a relation whose minimum deficiency increases by exactly 1.

Theorem 5: Removing any δ_0 points from N, you are left with a relation that has minimum deficiency 0.

Studying the structure of relations from this point of view, we study relations of minimum deficiency 0.

Hall condition

Minimum deficiency 0 means that $|R(A)| \geq |A|$ for all $A \subseteq S$.

Observe, by the way, that we could have defined the minimum deficiency of the inverse relation. Sooner or later we have to compare the minimum deficiency of R with the minimum deficiency of R^{-1}. That come's very easily.

Then we stated a theorem — no proof yet:

Marriage Theorem:

$R \supseteq R'$ where R' is an everywhere defined monomorphism (i.e., a matching or transversal) iff R satisfies the Hall condition.

You can match the elements of S to the elements of T in the relation $R \subseteq S \times T$ without any overlap iff R satisfies the Hall condition.

Before we prove this, let's look at an application (if you don't see an application, you don't care):

Application — Birkhoff — von Neumann Theorem

By the way, the most interesting application — which, unfortunately, in my book is given clumsily — is to prove the existence of Haar measure of compact groups. Now, if you look at the part of my book that I told you to, I prove studidly something about Haar, Bohr, and all those periodic functions. But the same arguments that are given in my book can be used to prove the existence of Haar measure on compact topological groups.
You can look that up. I won't give it here.

In $\mathbb{R}^{n^2}$ we consider the set of all doubly stochastic matrices $X = (x_{ij})$

Call this set C.

(i.e., $n \times n$ matrices s.t. $x_{ij} \geq 0$, $\sum_{i=1}^{n} x_{ij} = 1 \ \forall j$, $\sum_{j=1}^{n} x_{ij} = 1 \ \forall i$
Row and column marginals equal 1.)

First, observe that C is a closed convex set [13.3].
A convex combination of 2 doubly stochastic matrices is a doubly stochastic matrix.
In fact, it's actually a convex polyhedron, because it's defined by inequalities and the inequalities are the convex polyhedron.

C is a convex polyhedron.

Therefore, like all convex polyhedron, C has vertices.
The vertices are the points that are not convex combinations of any finite subset

Q: What are the vertices?

The answer is exactly those doubly stochastic matrices whose entries are 0 or 1.
This means they are permutation matrices. Namely, you start with the identity matrix I and permute the rows and the columns.

A: They are the permutation matrices.
That's the Birkhoff-von Neumann Theorem.

There are a number of very interesting applications.
So let's see 2 proofs of this Theorem.
First Birkhoff's proof, then von Neumann's.

Birkhoff's Proof:

In an obscure paper, published in Spanish, in an Argentine journal:

$$\begin{array}{cc} & T \\ S & \begin{bmatrix} x_{ij} \end{bmatrix} \end{array} = X$$

Take a doubly stochastic matrix X.
Suppose that the rows are S (the boys) and the columns are T (the girls).

$x_{ij} \geq 0$

$\sum_i x_{ij} = 1 \; \forall j, \; \sum_j x_{ij} = 1 \; \forall i$

The non zero entries of this doubly stochastic matrix defines a relation.
Namely, a row i is related to a column j if the corresponding entry is non zero ($x_{ij} \neq 0$).

$$R_X \subseteq S \times T$$

We show that:

(*) R_X, where X is doubly stochastic, satisfies the Hall condition.

Suppose we prove this assertion.
If we prove this assertion, then we can deduce the conclusion of the Birkhoff-von Neumann Theorem at once.
As follows:

Assuming (*), we apply The Marriage Theorem.
That means we have a matching of S to T.
But a matching means a set of entries which correspond to a permutation matrix whose entries are non zero.

By the Hall condition, we can find a subset of X of non zero entries s.t. no two of them are on a line (a line is a row or column).
Every line contains exactly one non zero entry.

Say the corresponding permutation matrix is P.
(In other words, replace the non zero entries of the subset of X found above by 1.
All other entries in P are 0.)

Let ε be the <u>minimum</u> of <u>non zero entries</u> in the subset of X found above.
Since $x_{ij} \geq 0$ and we are taking the minimum non zero entry in the subset of X found above, we know that $\varepsilon > 0$.

$\underbrace{X}_{\text{marginals } 1} - \underbrace{\varepsilon P}_{\text{marginals } \varepsilon}$ has non zero entries. ← $X - \varepsilon P$ has at <u>least one</u> additional zero entry than X.

$X - \varepsilon P$ is <u>NOT</u> doubly stochastic.
However, the marginals of this matrix are equal.
The marginals of X are 1 and the marginals of εP are ε.
So the marginals of $X - \varepsilon P$ are $1 - \varepsilon$.
Therefore:

$$Q = \frac{1}{1-\varepsilon}\left(X - \varepsilon P\right) \text{ is doubly stochastic.}$$

} From above, Q has <u>at least</u> one more zero entry than X.

Now we have the following:

$$X = \underbrace{(1-\varepsilon)\underset{\substack{\uparrow \\ \text{doubly stochastic matrix.}}}{Q}} + \varepsilon \underset{\substack{\uparrow \\ \text{permutation matrix.}}}{P}$$

} X is a <u>convex combination</u> of 2 <u>doubly stochastic matrices.</u>

I can perform the same trick on Q to obtain a convex combination of Q that includes a doubly stochastic matrix with at <u>least one additional</u> zero entry than Q.
I can do this recursively until all the entries are 0.

Therefore:
X is the convex combination of permutation matrices.

So we just need to prove assertion (*) [19.3], which we have assumed, and we can claim this result.

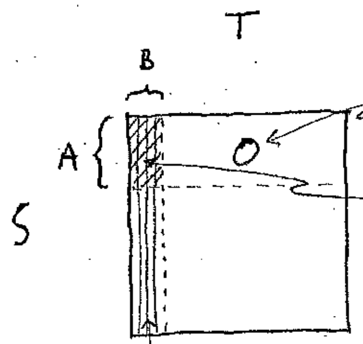
We need to show that R_X, where X is doubly stochastic, satisfies the Hall condition.

Suppose that the Hall condition is not true.

Then we have some subset A of S where:

$$R_X(A) = B \quad \text{and} \quad |B| < |A| \text{ (strictly)}, \text{ for some } A \subseteq S$$

Let's see what this means.



all 0, by definition of R_X

The entries here add up to as many as there are of A, namely $|A|$.

Because X is doubly stochastic, the sum of all rows of column vectors for B must equal exactly $|B|$.

This gives $|B| \geq |A|$.

But the assumption is that $|B| < |A|$.

So we have a contradiction and the Hall condition is satisfied.

And our proof is complete.

Birkhoff - von Neumann Theorem

Every doubly stochastic matrix is a convex combination of permutation matrices.

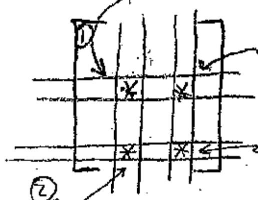
- <u>von Neumann's Proof :</u>

I don't know why this paper was never published.
It was transmitted orally – like the Odyssey.
Now I transmit it to you. And you will transmit it to your students. And so on and so on.
Let me do this first by gestures.

If X is a permutation matrix, we win.
So we might as well assume that it's not a permutation matrix.

That means there is an <u>entry</u> $0 < x_{ij} < 1$.

strictly



(1)

(4) which in turn requires some entry in this column strictly > 0 and strictly < 1.

(3) which in turn requires some entry in this row strictly > 0 and strictly < 1.

(2) Since this column must sum to 1, there must be another entry that is also strictly > 0 and strictly < 1.

And you keep going in this fashion until you get a <u>cycle</u>.
You keep going like this until you eventually get back to where you started.
You have a cycle of entries where each entry is strictly > 0 and strictly < 1.

Now what do I do?
I take the original entry of the cycle and <u>increase</u> it a little bit (ε).
The matrix is no longer doubly stochastic, so I decrease the next entry in the cycle by ε.
I continue around the cycle, in this fashion, alternatively increasing, then decreasing each entry in the cycle by ε.
We refer to this doubly stochastic matrix as $X_{+\varepsilon}$.

Now I take the original matrix X and the original entry of the cycle. This time I <u>decrease</u> it a little ($-\varepsilon$). The next entry in the cycle I increase, etc.
We refer to this doubly stochastic matrix as $X_{-\varepsilon}$.

Then we note:

$$X = \underbrace{\tfrac{1}{2} X_{+\varepsilon} + \tfrac{1}{2} X_{-\varepsilon}}_{\text{convex combination of doubly stochastic matrices}}$$

recursive applications on the subsequent $+\varepsilon$, $-\varepsilon$ doubly stochastic matrices eventually result in <u>permutation</u> matrices.

- **Exercise 19.1**

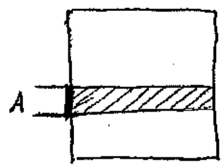
Write up von Neumann's Proof.

- Kultur

Whose taken real variables, functional analysis, etc.?

There is a continuous analogue of the doubly stochastic matrix.
Namely, a doubly stochastic probability measure.

You take the unit square.
Then you have a probability measure on the unit square.



Probability P of the events of the unit square add up to 1.

How do I make it doubly stochastic?
The probability of any rectangle that is formed is equal to the probability of the side * 1.
That's a doubly stochastic measure.

Now, again, it is obvious that the set of all doubly stochastic measures is convex.
Well, look at the properties that characterize the extremals.
What are the doubly stochastic measures that are NOT convex combinations?
This is a really cute invention that these guys did. Professor Douglas, now Provost of Texas A&M, and Professor Lindenstrauss of the Hebrew University in Jerusalem.
They found a way of characterizing extremals.
Extraordinary.
It goes like this:

A doubly stochastic probability is extremal iff functions $F(x,y) = f(x) + g(y)$ are dense in $L_1(P)$.

space of all integrable functions

This is the right way of saying the measure is very thin.
Because you can now calculate any function of two variables by summing the function of one variable relative to their probability.
The proof is given in my book. And it's very similar to von Neumann's Proof.

Muirhead's Inequality

I can't resist giving you an application of the work of von Neumann.
This is the most beautiful inequality there is.

Suppose we have n real variables:

$$x_1, x_2, \ldots, x_n \geq 0 \quad , \; x_i \in \mathbb{R}$$

It's been known, probably since the Greeks, that for any numbers:

$$\sqrt[n]{x_1 + x_2 + \ldots + x_n} \leq \frac{x_1 + x_2 + \ldots + x_n}{n}$$

geometric mean — arithmetic mean

This is the most famous inequality there is.
We want to generalize this.
This is how we do it.

Take a vector of exponents $\underline{a} = (a_1, a_2, \ldots, a_n)$, $a_i \in \mathbb{R}$

Define the $\underline{a}$-mean as:

$$[\underline{a}] = \frac{1}{n!} \sum_{\sigma} x_{\sigma_1}^{a_1} x_{\sigma_2}^{a_2} \cdots x_{\sigma_n}^{a_n}$$

↑ ranges over all permutations of the indices

Examples of $\underline{a}$-mean

$\underline{a} = (1, 0, \ldots, 0) \Rightarrow [\underline{a}] =$ arithmetic mean

$\underline{a} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) \quad [\underline{a}] =$ geometric mean

$\underline{a} = (1, 2, 0, \ldots, 0) \quad [\underline{a}] = \frac{1}{n(n-1)} \sum_{i,j} x_i x_j^2$

The problem is, given 2 $\underline{a}$-means, when is one less than the other?

Theorem (Muirhead) The most famous inequality that's not trivial.

We have $[\underline{a}] \leq [\underline{b}]$ for all $x_i \geq 0$ iff there exists a doubly stochastic matrix X for which $\underline{a} = X\underline{b}$.

For example, for geometric and arithmetic means, inequality is a special case.

If you take $\underline{a} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$ and $\underline{b} = (1, 0, \dots, 0)$, you find that $\underline{a} = X\,\underline{b}$,
where X is a doubly stochastic matrix.
It works.

Muirhead's Theorem gives you the <u>exact criteria</u> to get all possible inequalities, with which you can tease your friends.

Proof:

We'll see how <u>to use</u> the Birkhoff - von Neumann Theorem.
The proof consists of setting up your notation right.
Then it's obvious.

$$[\underline{a}] = \frac{1}{n!}\sum_{\sigma} x_{\sigma_1}^{a_1} x_{\sigma_2}^{a_2} \cdots x_{\sigma_n}^{a_n}$$

↑ ranges over all permutations of the indices.

I rewrite this by permuting the exponents, rather than permuting the indices of x:

$$= \frac{1}{n!}\sum_{\sigma} x_1^{a_{\sigma_1}} x_2^{a_{\sigma_2}} \cdots x_n^{a_{\sigma_n}}$$

↑ ranges over all permutations of $\underline{a}$ (i.e., the exponents)

This can be rewritten as:

$$= \frac{1}{n!}\sum_{\sigma} e^{\sum_{i=1}^{n} a_{\sigma_i} \log x_i}$$

Let $y_i = \log x_i$

$(\underline{a}, \underline{y}) =$ [redacted] $\sum_{i=1}^{n} a_i y_i$ ← dot product

Instead of permuting over σ (the a_{σ_i}), we use permutation matrices.
Let P = permutation matrix

$$= \frac{1}{n!}\sum_{P} e^{(P\underline{a},\, \underline{y})}$$

↑ ranges over all permutation matrices.

The assumption that Muirhead employs is that:

$$\underline{a} = X\,\underline{b}$$

where X is a doubly stochastic matrix.

$$[\underline{a}] = \frac{1}{n!}\sum_{P} e^{(P\underline{a},\, \underline{y})} = \underbrace{\frac{1}{n!}\sum_{P} e^{(PX\underline{b},\, \underline{y})}}$$

Next, we use the fact that X is doubly stochastic and simplify.
That we'll do next time.

Matching Theory (cont'd)

Let's continue with Muirhead's inequality.
Let's review the logical steps that we've been following.
We have stated, but NOT yet proved, The Marriage Theorem:

The Marriage Theorem

$$R \subseteq S \times T$$

$R \supseteq R'$ where R' is a matching iff for every subset $A \subseteq S$,

$$|R(A)| \geq |A|$$

i.e. $\delta(A) \geq 0$

Using The Marriage Theorem, we proved the Birkhoff - von Neumann Theorem.
This says if you take the set of all doubly stochastic matrices in n^2 dimensional space, this is a polyhedron (or polytope, whatever you'd like to call it), whose vertices are exactly the permutation matrices.
We saw that this was an immediate application of The Marriage Theorem.

"Permutation matrices are the only vertices (extremal points) in the convex set of doubly stochastic matrices."

I'd like to assign the following 1½ star problem.
By the way, why don't you do 2 one star problems.
I hear you are doing too little work in this course.
Two one star problems in the whole course, to be turned in. Instead of one one star problem.

---

* Exercise 20.1

I would work this out if I had the time, but I don't have the time to think about it.
I suspect that from the Birkhoff - von Neumann Theorem, you can deduce The Marriage Theorem. This is not just an intellectual exercise. I have an ulterior motive for assigning this problem. I always have ulterior motives.
That is, we have seen, in one of our Kultur asides, that there is a measure theoretic analogue of the Birkhoff - von Neumann Theorem. Namely, The Douglas - Lindenstrauss Theorem [19.7].

This states that if you take the set of all doubly stochastic measures on the square, that's a convex set, whose extremals are those measures for which L_1 of that measure is spanned by functions of the form $F(x,y) = f(x) + g(y)$.
It's a beautiful theorem.

So if we had a way of deriving The Marriage Theorem from the Birkhoff + von Neumann Theorem, then we would probably have a way of stating, for the first time known to man, a continuous analogue of The Marriage Theorem.
That's my ulterior motive.
So please work on it. You have lot's of time. What do you do with your time?
I am the one who has no time.

From the Birkhoff - von Neumann Theorem, deduce the Marriage Theorem.

By the way, there is a generalization of the Marriage Theorem in the transfinite sense, which is due to the great British combinatorist Nash Williams.
I might do this if you want to. I've never done this in class.
Where S and T are well ordered sets, there is a marriage theorem for that case.
This is extremely elegant.

Using the Birkhoff - von Neumann Theorem, we are now in the course of proving one of the most striking inequalities ever stated.
Inequality that encompasses all generalizations of the geometric, arithmetic mean inequality.
As you recall:

$$x_1, x_2, \ldots, x_n \geq 0 \ , \ x_i \in \mathbb{R}$$

$$\underbrace{\sqrt[n]{x_1 x_2 \cdots x_n}}_{\text{geometric mean}} \leq \underbrace{\frac{x_1 + x_2 + \cdots + x_n}{n}}_{\text{arithmetic mean}}$$

Don't you ever forget that. It's Mickey Mouse. High School. Known to the Greeks.

Muirhead's inequality is the ultimate generalization of the geometric arithmetic mean inequality.

Suppose we have $\underline{a} = (a_1, a_2, \ldots, a_n)$

Define the $\underline{a}$-mean as:

$$[\underline{a}] = \frac{1}{n!} \sum_{\sigma} x_{\sigma_1}^{a_1} x_{\sigma_2}^{a_2} \cdots x_{\sigma_n}^{a_n}$$

in general, you get cancellations here, of multiples of terms of the permutations.

ranges over all permutations of the indices (i.e., the set $\{1, \ldots, n\}$)

For example:

$\underline{a} = (1, 0, \ldots, 0) \implies [\underline{a}] =$ arithmetic mean

$\underline{a} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) \implies [\underline{a}] =$ geometric mean

Q: When is $[\underline{a}] \leq [\underline{b}]$ for all x_i?

Muirhead's inequality gives you the answer to the question:

A: Iff there exists a doubly stochastic matrix X s.t. $\underbrace{\underline{a} = X\underline{b}}$

(in some sense, this says $\underline{a}$ is an average of $\underline{b}$. Applying doubly stochastic matrices is a very suble way of averaging the entries.)

For example:

$$X = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \qquad \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix} = X \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

doubly stochastic

So, in the special case of Muirhead's Inequality, we get the special case of the geometric mean $\leq$ arithmetic mean inequality.

Before I go further into the proof, let me mention something that I should have mentioned before. This will be copied out of my book (Gian-Carlo Rota on Combinatorics) pp. 537–538.

Is there a simple criterion, given 2 vectors $\underline{a}, \underline{b} \in \mathbb{R}$, to know a priori whether there is a doubly stochastic matrix X where $\underline{a} = X\underline{b}$?

The answer is yes.

Remark

$\underline{a} = X\underline{b}$ for some doubly stochastic matrix X iff $\underline{a} \leq \underline{b}$ in the <u>dominance order.</u>

↑ we tacitly assume we have <u>extended</u> the dominance order [12.8–9] to vectors whose entries are real numbers other than integers.

You <u>reorder</u> the components of $\underline{a}$ and $\underline{b}$ s.t.:

$$a_1 \geq a_2 \geq \ldots \geq a_n$$

$$b_1 \geq b_2 \geq \ldots \geq b_n$$

as well you may, because the remark above is completely independent of the order of the entries, due to the permutations associated with the doubly stochastic matrix.

$\underline{a} \leq \underline{b}$ means that:

$$\left.\begin{array}{rcl} a_1 & \leq & b_1 \\ a_1 + a_2 & \leq & b_1 + b_2 \\ & \vdots & \end{array}\right\} \text{ all these are } \underline{\text{inequalities}}$$

$$a_1 + \dots + a_n = b_1 + \dots + b_n$$

and this last statement is an equality

To discuss the intuitive meaning of the dominance order [12.8-9], we try to understand the intuitive meaning of the covering relation.
For sake of the argument, let's take the special case where the entries are integers.
You have:

$$\underline{a} = (a_1, a_2, \dots, a_n), \text{ where } a_1 \geq a_2 \geq \dots \geq a_n$$
$$\underline{b} = (b_1, b_2, \dots, b_n), \qquad b_1 \geq b_2 \geq \dots \geq b_n$$

Then you draw the Ferrers matrix.

Example: $\underline{a} = (4,4,2)$ $\underline{b} = (5,4,1)$

$$\begin{array}{c|ccccc|} a_1 & 1 & 1 & 1 & 1 & \\ a_2 & 1 & 1 & 1 & 1 & \\ a_3 & 1 & 1 & & & \end{array} \qquad \begin{array}{c|ccccc|} b_1 & 1 & 1 & 1 & 1 & 1 \\ b_2 & 1 & 1 & 1 & 1 & \\ b_3 & 1 & & & & \end{array}$$

$\underline{a} \prec \underline{b}$

How do you get $\underline{a}$ from $\underline{b}$?
The covering relation is that you move one 1 unit down, w/o disturbing the fact that you have a Ferrers diagram.

$\underline{a} \prec \underline{b}$ is the covering relation.
The dominance order relations are the iterations.

So that is what the dominance order is about.
I stated before the unsolved two star problem to give a purely order theoretic characterization of the dominance order (exercise 12.1 [12.9]).
The dominance order has an involution $\perp$. It is one of the rare partial orders that has an involution.
No one has gotten a purely abstract characterization of this partial order as a lattice.

Historically, I think you ought to know that the dominance order arose first in the continuous case. And it was only later that people realized it could be used in the discrete.

It arose for continuous functions on the interval $[0,1]$.
I'll do it by gestures.
If you have a continuous function on the interval $[0,1]$, you can talk about rearranging the function – the continuous analogue of rearranging the entries of a vector.
In particular, you can define the notion of a <u>non increasing rearrangment</u> of the same function.
So <u>every</u> function has a non increasing rearrangement.
So you say:

functions $g \le f$ in the dominance order if $\int_0^x g(x)\,dx \le \int_0^x f(x)\,dx$

This has tremendous applications all over the place.
Statistics, for example.
So, it's a fundamental order relation.

Based on this Remark ($\underline{a} = X\underline{b}$ for some doubly stochastic matrix X iff $\underline{a} \le \underline{b}$ in the dominance order), it is very easy to work with Muirhead's Inequality.
Because all you really need to do is draw the Ferrers' diagrams and see if you can move units down.

Example: $\underline{b} = (3,2,1,0)$ $\qquad$ $\underline{a} = (3,1,1,1)$

$$\boxed{\begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & \\ 1 & & \\ & & \end{array}} \qquad \boxed{\begin{array}{ccc} 1 & 1 & 1 \\ 1 & & \\ 1 & & \\ 1 & & \end{array}}$$

$\underline{b} \succ \underline{a}$ so $\underline{b} \ge \underline{a}$ in the dominance order

Therefore, there exists a doubly stochastic matrix X where $\underline{a} = X\underline{b}$
(from the theorem you copied out of my book on pp. 537–538)

Then the Muirhead Inequality immediately says:

$$[\underline{a}] \le [\underline{b}]$$

The mean you form with $\underline{a}$ $\le$ the mean you form with $\underline{b}$

$$[\underline{a}] = \frac{1}{4!}\sum_{\sigma} x_{\sigma_1}^3 x_{\sigma_2} x_{\sigma_3} x_{\sigma_4} \le \frac{1}{4!}\sum_{\sigma} x_{\sigma_1}^3 x_{\sigma_2}^2 x_{\sigma_3} x_{\sigma_4}^0 = [\underline{b}]$$

Example: $\underline{a} = (4, 3, 2)$ $\underline{b} = (5, 3, 1)$

Factorials cancel and you get, where σ = all permutations of indices i, j, k:

$$\sum_{\sigma} x_i^4 x_j^3 x_k^2 \leq \sum_{\sigma} x_i^5 x_j^3 x_k$$

Kultur

$[\underline{a}] \leq [\underline{b}]$ is an equality between 2 symmetric functions of the variables $x_1, x_2, \ldots, x_n$. Consider $[\underline{b}] - [\underline{a}] \geq 0$

Hilbert's 17^{th} Problem

Suppose you have a polynomial $p(x_1, \ldots, x_n) \geq 0$ for all x_i

There is one good reason why a polynomial ≥ 0.
Namely, it's the square of another polynomial.
You can jazz this up.
It is the sum of squares of other polynomials.
So one reason why a polynomial ≥ 0 is because it's the sum of squares.

And it was noted very early in the game, particularly by Hilbert, that this is NOT true that if a polynomial ≥ 0, that it is the sum of squares. The polynomial is not necessarily the sum of squares.

$p(x_1, \ldots, x_n)$ is not necessarily the sum of squares.

Hilbert's 17^{th} Problem is:

when is a positive polynomial $\left(p(x_1, \ldots, x_n) \geq 0 \text{ for all } x_i \right)$ a sum of squares?
What condition has to be satisfied?
Hilbert found 3 cases.

Case 1: when you have a homogeneous polynomial of degree 2
If you have such a polynomial, then it's in quadratic form and the coefficients form a symmetric matrix.
And then the symmetric matrix can be diagonalized.
That means that every polynomial of degree 2 is the sum of squares.

So this case (quadratic polynomials) is true by matrix theory.

Case 2: This case comes from Emil Artin, father of Professor Artin; who established the following result:

Artin-Schrier

Given $p(x_1, \dots, x_n) \geq 0$ for all x_i, then:

$$p(\underline{x}) = \underbrace{\sum_{j=1}^{k} r_j(\underline{x})^2 \quad \text{where} \quad r_j(\underline{x}) = \frac{p_j(\underline{x})}{q_j(\underline{x})}}_{\text{always the sum of squares of rational functions}}$$

This was done using strictly first order mathematical logic. This was done in the 1920's.
It's not really what Hilbert asked for, but it's something.

Case 3: What if the polynomial is symmetric?

Q: If $p(x_{\sigma_1}, x_{\sigma_2}, \dots, x_{\sigma_n}) = p(x_1, x_2, \dots, x_n)$ then is it true that if $p(\underline{x}) \geq 0$ then $p(\underline{x})$ is the sum of squares.

A: No. But there are only a finite number of exceptions.

these were found by the Italian mathematician Procesi in 1974 (approximately).

There are a finite number of inequalities that are NOT sums of squares.
All the other ones are due to sums of squares.
This is a tremendous achievement.

So you see why I'm so concerned with Muirhead's Inequality.
This is what's in the background.

There's always a background you have to learn.
Lot's of people have worked on this stuff.
I could go on and on.
But I will simply stop now. End of Kultur.

At least you learn something about the big wide world and not just narrow things.
You must be narrow.
You must learn about everything.
Otherwise you'll end up in the gutter when fashions change.
You have to be prepared for when things shift around. Know what in the world.
You can't just know the algorithm from star alpha beta 2. You can't just make your living on that. You'll end up in a bad way if you do that.

Before proving Muirhead's Theorem, I have to bring in some stuff from 18.01 (Calculus).
We've talked about convex sets.
But what's a convex function?

convex function

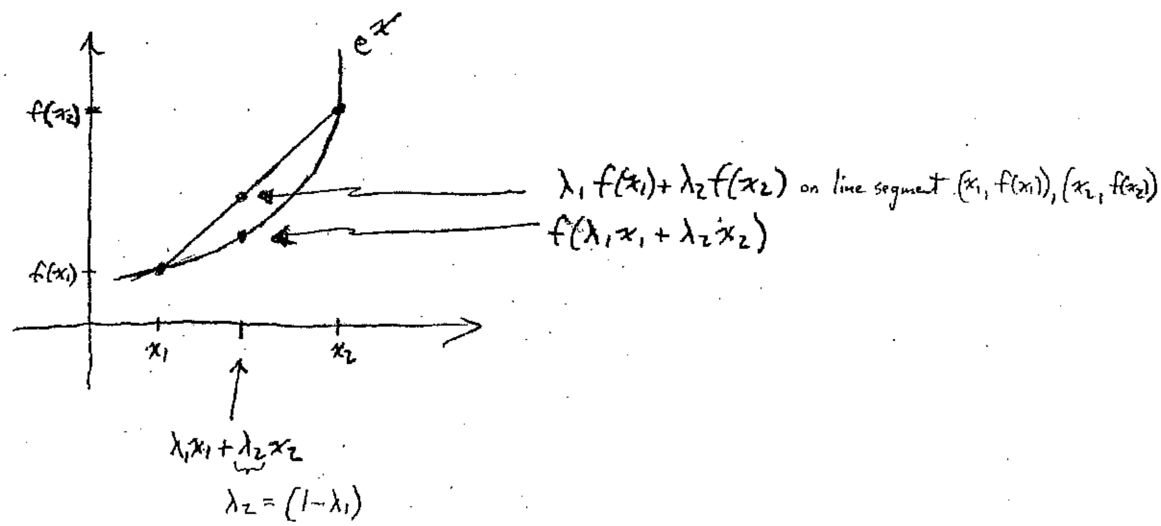
A function $f(x)$ of $x \in \mathbb{R}$ is said to be convex when:

$$f\left(\sum_i \lambda_i x_i\right) \leq \sum_i \lambda_i f(x_i) \quad \text{for all } x_i \text{ and}$$
$$\text{all } \lambda_i \geq 0 \text{ s.t. } \sum_i \lambda_i = 1$$

This is also known as Jensen's Inequality.

Example: e^x is a convex function

Why? Draw a picture in $\mathbb{R}^2$ to get the idea.



Proof of Muirhead's Inequality

$$[\underline{a}] = \frac{1}{n!}\sum_{\sigma} x_1^{a_{\sigma_1}} x_2^{a_{\sigma_2}} \cdots x_n^{a_{\sigma_n}}$$

$\uparrow$ ranges over all permutations of $\underline{a}$ (i.e., the exponents)

$$= \frac{1}{n!}\sum_{\sigma} e^{\sum_{i=1}^{n} a_{\sigma_i} \log x_i}$$

Let $\underline{y} = (\log x_1, \ldots, \log x_n)$

$$(\underline{a}, \underline{y}) = \sum_{i=1}^{n} a_i y_i \longleftarrow \text{dot product}$$

Instead of permuting over σ (i.e., the a_{σ_i}), we use permutation matrices.

P = permutation matrix

Then we can rewrite the above as:

$$= \frac{1}{n!}\sum_{P} e^{(P\underline{a}, \underline{y})}$$

$\uparrow$ ranges over all permutation matrices

Now we remember that $\underline{a} = X\underline{b}$.

$$= \frac{1}{n!}\sum_{P} e^{(PX\underline{b}, \underline{y})}$$

Now we use the Birkhoff - von Neumann Theorem.
X is a doubly stochastic matrix.
By Birkhoff - von Neumann, therefore X is the convex combination of permutation matrices.

$$X = \sum_{Q} \lambda_Q Q \quad , \quad \text{where } \sum_{Q} \lambda_Q = 1 \quad , \quad \lambda_Q \geq 0$$

$\uparrow$ ranges over all permutation matrices.

$$= \frac{1}{n!}\sum_{P} e^{\sum_{Q} \lambda_Q (PQ\underline{b}, \underline{y})}$$

At this point, we use the fact that e^x is a <u>convex function.</u>
This gives:

$$e^{\sum_Q \lambda_Q (PQ\underline{b}, \underline{y})} \leq \sum_Q \lambda_Q e^{(PQ\underline{b}, \underline{y})}$$

$$\leq \frac{1}{n!} \underbrace{\sum_P \sum_Q}\lambda_Q e^{(PQ\underline{b}, \underline{y})}$$

As P ranges all permutation matrices, Q ranges over all permutation matrices.

$$= \frac{1}{n!} \sum_Q \lambda_Q \underbrace{\sum_P e^{(PQ\underline{b}, \underline{y})}}$$

P ranges over <u>all</u> permutation matrices,
so this is independent of Q.
PQ is just another permutation, so we can
rewrite this as follows, where <u>R ranges over all permutation matrices.</u>

$$= \frac{1}{n!} \sum_Q \lambda_Q \sum_R e^{(R\underline{b}, \underline{y})}$$

Recall that $\sum_Q \lambda_Q = 1.$

$$= \frac{1}{n!} \sum_R e^{(R\underline{b}, \underline{y})}$$

What's this? It's exactly $[\underline{b}]$

$$= [\underline{b}]$$

$$[\underline{a}] \leq [\underline{b}],$$

Q.E.D.

This is the hard part.
The other half of this ~~iff~~ proof is trivial.

The program for the next few days is that we'll discuss the Marriage Theorem and its variants, Dilworth's Theorem, and a few applications.
Then we'll do the Marriage Theorem all over again using deficiencies. And that will lead us into matroids — by analyzing the notion of deficiencies of submodular set functions.
We'll be led to the study of matroids from the notion of submodular set functions.

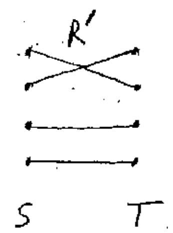
We'll cover a certain amount of basic material on matroids. Enough to get to the main matching theorems on matroids, which are an enormous strengthening of the Marriage Theorem.
Extremely power strengthening of the Marriage Theorem.
We will carry the theory of matroids that far.
We will not have time to do the geometric aspects of the theory of matroids, which are extremely interesting.
After that, I will have to switch over and start on geometric probability, as per list. We'll do geometric probability and then, hopefully, as much Möbius functions as we have time.
Geometric probability will serve as an introduction to Möbius functions.
This is an interesting challenge, by the way, to use geometric probability to introduce Möbius functions.
So that's the scheme of the content for the rest of the term.

The Marriage Theorem

You have seen this already stated several times now.
It's high time we prove it.

Given a relation $R \subseteq S \times T$
if for every $A \subseteq S$ we have $|R(A)| \geq |A|$ (i.e., deficiencies ≥ 0),
then there exists a matching for R (i.e., a relation $R' \subseteq R$ s.t. for every $a \in S$ there is exactly one element in R' of the form (a,b) and if $(a,b) \in R'$, $(c,d) \in R'$, if $c \neq a$ then $b \neq d$).

this is just a fancy way of saying that R' is a 1-to-1 function defined from S to T.
From each element of S there is exactly one edge issuing and no two edges go to the same element in T.



A graph of a 1-to-1 function everywhere defined on S.

Proof :

Let's consider first a simple and cute proof.
This is the next to simplest proof I know.
The simplest proof there is is the one that uses linear algebra. And I won't do it.
You ask Thomas Britz about it. I don't want to digress from here now.

Case 1: $|R(A)| > |A|$ for every $A \subset S$, $A \neq \emptyset$
($>$ strictly)

You pick an element of S, pick an element related to it, remove them, remove everything related to them, and consider the relation that remains.
When you remove them, the RHS of the above inequality goes down by at least one for the new relation. The inequality continues to be satisfied.
And you can proceed by induction.

Pick any $a \in S$, $b \in T$ s.t. $(a,b) \in R$

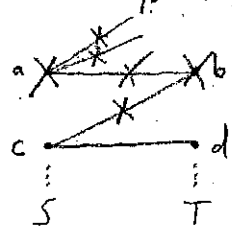
Remove a + all edges issuing from a.
Remove b + all edges issuing from b.

Let R'' be the restriction of R to $S-a$ and $T-b$.

For $B \subseteq S-a$, we have:

$$|R''(B)| \geq |B|$$

Since $|R(A)| > |A|$ for every $A \subset S$, each element of S is related to 2 or more elements of T.
After removal of all edges issuing from b, every element of $S-a$ is related to at least one element of T.
So equality is now possible in the inequality $|R''(B)| \geq |B|$, with the removal of all relationships with b. For example:



$B = \{c\}$
$|R''(B)| = |B|$

Continue by induction with the smaller relation R''.

case 2 : There exists a non empty $A \subset S$ s.t. $|R(A)| = |A|$

Then we proceed, as follows. A is smaller than S. Take a matching in A. Remove this matching. Then prove in the remaining sets that the Hall condition is still satisfied.

Consider $R''' = R|_A$ ← R restricted to A

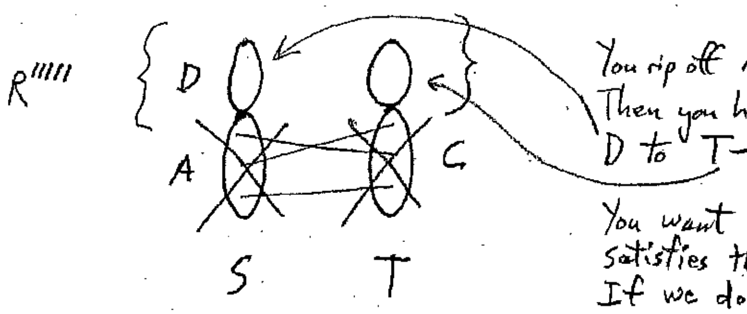
Then R''' also satisfies the Hall condition, and it's smaller than R.

Therefore, by induction, R''' has a matching, say R'''', from A to some set C of T.

Now I'll tell you what I do.
I remove everything in A and everything in C.
I can't use the elements in C anymore, they've already been matched.

Let R''''' be the restriction of R to $S-A$ and $T-C$, where C is the range of R''''.

The picture is like this :



You rip off A and you rip off C. Then you have a relation from D to $T-C$.

You want to show that this relation satisfies the condition of Hall. If we do that, we win. Because we can piece together the two and we get a matching of a smaller relation.

Need to show that R''''' satisfies the Hall condition:
Take subset $D \subseteq S-A$
Need to show:

$$|R'''''(D)| \geq |D|$$

Proof

$$|R'''''(D)| = |R'''''(D)| + \underbrace{|R(A)| - |A|}$$

we have, for this case, that $|R(A)| = |A|$
So we add 0, by adding and subtracting the same number of elements.

This is a measure of sets [8, 10], and we have:
$$\left\{ |R'''''(D)| + |R(A)| = |R'''''(D) \cup R(A)| + |R'''''(D) \cap R(A)| \right\}$$

$$= |R'''''(D) \cup R(A)| + |\underbrace{\cancel{R'''''(D) \cap R(A)}}_{\phi}| - |A|$$

This is empty, by the way we constructed D.

$$= |\underbrace{R'''''(D) \cup R(A)}| - |A|$$

$$R(D \cup A) = R'''''(D) \cup R(A)$$

disjoint by construction.

$$= |R(D \cup A)| - |A|$$

R satisfies the Hall condition.
So we have:

$$|R(D \cup A)| \geq |D \cup A|$$

$$\geq |D \cup A| - |A|$$

D and A are disjoint

$$= |D| + |A| - |A|$$

$$= |D|$$

Which gives:

$$|R'''''(D)| \geq |D|$$ ← We win, Finished. That's the end of the proof.

This proof has been arrived at with a lot of effort, Starting with simplifying the original proof.

Let's consider a variant of the Marriage Theorem.
Let's consider a theorem that is strongly related to the Marriage Theorem.
In fact, this is a consequence of the Marriage Theorem.
But to get it as a consequence of the Marriage Theorem, I need to do some fudging, which I don't like.

For the moment, you'll sense that the theorems are very closely related.
Then we'll see that they are variants of the same thing.

Dilworth's Theorem

In a sense, this is more elegant than the Marriage Theorem.
Although it's at the same level of depth.

P = finite partially ordered set

Dilworth's Theorem has to do with the following problem.
You want to partition the set P into blocks in such a way that every block is a chain. You want to do this as economically as possible.
So there is a minimum number of chains.
How many chains can you get away with in such a partition?

Can we get a rough bound?
Sure we can get a rough bound. You take the antichains of P, where every element of every antichain has to be in a different block.
Therefore, there must be at least as many blocks as there are maximum antichains.
Dilworth says that that is enough.

Theorem:
The minimum number of blocks in a partition of P into chains equals the maximum size of an antichain.

Proof (Tverberg)

This is even simpler than the proof given in my book.
If P has one element, the statement is trivial.
So we can proceed by induction.

You take P and take a maximal chain of P (i.e., a chain that can not be extended further).

Let C = <u>maximal</u> chain in P

Consider the partially ordered set $P-C$

There are two cases. Either the maximum size of an antichain in $P-C$ is one smaller than the maximum size of an antichain in P — and then we win by induction. Or else the maximum size of an antichain in $P-C$ is the same as the maximum size of an antichain in P — and then we have to argue.

case 1: the maximum size antichain of $P-C$ is one unit less than the maximum size antichain of P.
Proceed by induction.
induction on the maximum size antichain

case 2: the maximum size antichain of $P-C$ is the same as the maximum size antichain of P.

Pick a maximum size antichain of $P-C$:

Let $\{a_1, a_2, \dots, a_n\}$ be a maximum size antichain in $P-C$.

The maximum element of C has to be comparable to all the a_i, otherwise I would add it to $\{a_1, a_2, \dots, a_n\}$ and get a bigger maximum size antichain.

Let m = maximum element of C
So it's either greater than one of the a_i, or less than one of the a_i.
It's not equal to any of the a_i, because these are in the antichain $P-C$.

Suppose it's less than one of these, namely a_α:
Say $m < a_\alpha$
That's <u>impossible</u>, because C is <u>not</u> <u>maximal</u>.
I would immediately add a_α to C to get a larger set.

Hence $m > a_\alpha$ strictly greater than

Similarily, one argues that the <u>minimum</u> element (l) of C is strictly less than one of the a_i:

$l < a_\beta$

Now I perform the following split:

Let $U^+ = \{x \in P - C : x \geq a_i \text{ for some } a_i\}$

$U^- = \{x \in P - C : x \leq a_i \text{ for some } a_i\}$

Since $m > a_\alpha$, P is strictly greater than U^+.
Since $l < a_\beta$, P is strictly greater than U^-.

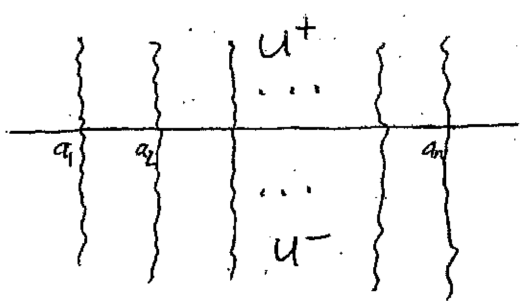
$$\left.\begin{array}{l} |U^+| < |P| \\ \\ |U^-| < |P| \end{array}\right\} \text{neither } U^+ \text{ nor } U^- \text{ is all of } P.$$

So U^+ and U^- are smaller sets and we can proceed by induction. By induction, the theorem already applies to U^+ and U^-.

Q: What's the biggest antichain of U^+?
A: $\{a_1, \ldots, a_n\}$

Q: What's the biggest antichain of U^-?
A: $\{a_1, \ldots, a_n\}$

What do you do?
You split U^+ and U^- each into n chains. And each of these n chains has to contain one of the a_i, otherwise you don't have enough chains. Then join the chains in U^+ and U^-.



Split U^+ and U^- into n chains each and match the chains.

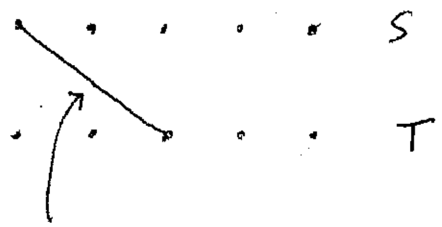
Forget C. C was just a prop to conclude that U^+ and U^- were strictly smaller than P. Once we conclude that U^+ and U^- are strictly smaller, we proceed by induction. Split U^+ and U^- and join the chains.

Now, let's derive the Marriage Theorem from Dilworth's Theorem.

2nd Proof of the Marriage Theorem

Corollary of Dilworth's Theorem : The Marriage Theorem, again.
How do we do this?
We are given a relation $R \subseteq S \times T$. You associate, with this relation, a partially ordered set. Like this. You write the elements of S on the top and the elements of T on the bottom. You say an element of T is greater than an element of S if there is an edge between them.

$R \subseteq S \times T$



P_R = partially ordered set associated to R ("T below S")

edge $\Rightarrow$ element s < element t

<u>Claim</u> : T is a maximum size antichain of the poset P_R.

<u>Assuming</u> this claim, you can apply Dilworth's Theorem to P_R.
Then we know, since T is a maximum size antichain, the minimum number of blocks in a partion into chains. And every element of S must be contained in one of those chains of $S \cup T$. Therefore, there must be as many two element chains as there are elements of S. This gives the matching.

So, <u>assuming the claim</u>, the <u>matching conclusion</u> follows immediately:

Existence of matching (Marriage Theorem) follows immediately from claim.
There must be as many two element chains as there are elements of S and that those two element chains give the desired matching.

<u>Proof of claim</u> [Gian-Carlo Rota on Combinatorics pp. 526-527]

Assume the claim is not true. Then there would be a <u>larger antichain U</u>:

$$U \subseteq P_R \quad \text{s.t.} \quad |U| > |T|$$

strictly

So $|T| - |U| < 0$

$$|T - U \cap T| = |T| - |U \cap T|$$

since S and T are disjoint, we have:

$$U = (U \cap S) \cup (U \cap T)$$

$$|U \cap T| = |U| - |U \cap S|$$

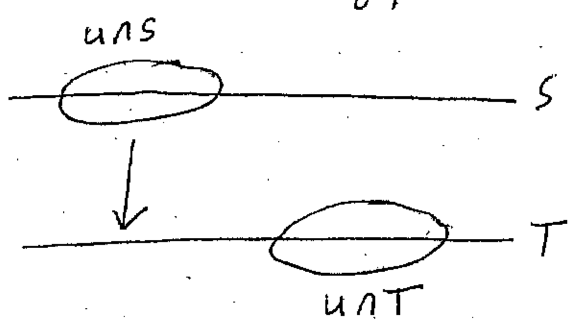
$$= \underbrace{|T| - |U|}_{<0} + |U \cap S|$$

$$< |U \cap S|$$

$$|T - U \cap T| < |U \cap S|$$

But $R(U \cap S) \subseteq T - U \cap T$

Why? Because of the following picture:



Since U is an antichain, no element of $U \cap S$ can be connected with an element of $U \cap T$.
Otherwise U would not be an antichain.

Therefore, $R(U \cap S)$ must be contained in the complement of $U \cap T$.
That's what I wrote.

So, we would have:

$$|R(U \cap S)| < |U \cap S| \longleftarrow$$

This contradicts the Hall condition for R.

We conclude, therefore, that the assumption is false and that the claim is true.

R has a matching.

That's the proof of the Marriage Theorem from Dilworth's Theorem.

- Now, let me tell you how Dilworth should come out of the Marriage Theorem.

  Assuming the Marriage Theorem, given partially ordered set P, define relation R_P as follows:

$$(a,b) \in R_P \quad \text{if} \quad a < b$$

strictly

Then by fudging the Marriage Theorem to R_P, out comes Dilworth's Theorem, But the proof I have is kind of inelegant.
We'll do it, in detail, using deficiencies.

If you come up with a proof of your own, I'd appreciate it.

- Next time, we'll do it all over with deficiencies. And we'll have some applications of Dilworth. There are some remarkable applications of Dilworth.
  Then, using the study of deficiencies, we'll introduce the concept of matroids.
  We'll see how the concept of a matroid comes out by analyzing the deficiency of a relation.

John Guidi
guidi@math.mit.edu

Dilworth's Theorem (conclusion)

P = finite partially ordered set

partition $\pi \in \Pi[P]$ where: every $B \in \pi$ is a chain
$|\pi|$ is minimum

Such a minimum equals the maximum size of an antichain of P.

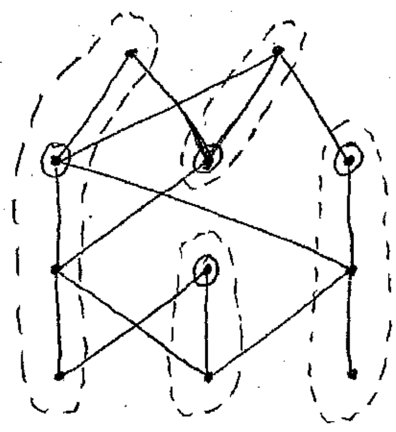
The minimum number of blocks in a partition into chains equals the maximum size antichain.

Last time we saw a very strict, short proof of this theorem.
I now repeat it, only by gestures.

You take the maximal chain and remove it.
See what's left.
If what's left has a maximum antichain that is smaller, you can proceed by induction.
If what's left has a maximum antichain of the same size, then we see that the chain that we removed has at least one element above this maximum antichain and at least one element below this maximum antichain. That means you can split P into P^+ and P^- which are strictly smaller. Therefore you can apply induction on P^+ and P^- obtaining two partitions of chains U^+ and U^-.
U^+ has minimal elements that are elements of the maximum antichain.
U^- has maximal elements that are elements of the maximum antichain.
So you can match the chains of U^+ and U^-. And this gives the number of chains.
That's the proof we saw last time.
We will shortly see another proof, based on deficiency concepts.

Example of Dilworth's Theorem

Hasse diagram of a partially ordered set



A maximal size antichain has 4 elements. So Dilworth tells you there has to be a partition of this partially ordered set into 4 chains.

One maximal size antichain indicated with ⊙ elements.
4 chains illustrated with dotted lines.

Now let's take a non trivial example of Dilworth's Theorem.
Let's take a Boolean algebra.

Example of Dilworth's Theorem – Boolean algebra

$\underbrace{P(S)}$, S finite

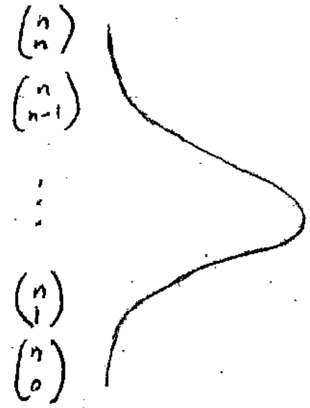
The Boolean algebra of all subsets of a set.
We want to partition the Boolean algebra into a minimum number of chains.
How do we do it?

We do it by the Greene–Kleitman bracketing algorithm.
In order to use this algorithm, we need to know the maximum size antichain of $P(S)$.

What is the maximum size antichain?

$|S| = n$

If you take the Hasse diagram of $P(S)$, it's a ranked, partially ordered set. And the elements of each rank are the subsets with 1 element, 2 elements, etc. If you count the elements of each rank, it's equal to the binomial coefficients – the number of subsets of k elements.



It's well known that the binomial coefficients increase to a maximum and then decrease. And if you normalize, then you get the bell shaped curve.
That's called the Central Limit Theorem.

The maximum binomial coefficient is either the middle one, when n is odd. Or the two middle ones, when n is even.
This can be checked by a simple algebra computation.

So, if n is odd, you have an antichain with as many elements as the maximum of the binomial coefficients. If n is even, as many as the two maxima.

But how do you know that's the maximum sized antichain?

How do you know that you can't combine things together?
You have to prove it. Sperner's Theorem (not to be confused with Sperner's Lemma).
This stuff is now in books, but we have to do it because it's important material.

Sperner's Theorem

The maximum size antichain of $P(S)$, $|S| < \infty$, $|S| = n$
has $\binom{n}{\left[\frac{n}{2}\right]}$ elements.

$\left[\frac{n}{2}\right]$ ← nearest integer

This is an extremely important result.
Like many results that we are considering in this chapter on combinatorics, it's important not only because of what it says, but because of all the conjectures it has led to. Remind me to tell you some.

So, it's a springboard. Once you get there, you ask similar questions about powers of partially ordered sets.

To prove this, we have to prove the famous LYM inequality – also found in all the books.

Proof.

Follows from LYM inequality (Lubell, Yamamoto, Meshalkin).

Let $\mathcal{U}$ be an antichain of $P(S)$.

Let $\mathcal{U}_k = \mathcal{U} \cap \underbrace{P_k(S)}$,

(this is standard notation for a family of subsets of S with k elements.)

so that $\mathcal{U}_k$ are the non-empty blocks of a partition of $\mathcal{U}$.

Then
$$\sum_{k=0}^{n} \frac{|\mathcal{U}_k|}{\binom{n}{k}} \leq \underline{1}$$

Proof of LYM inequality

How many complete chains are there in $P(S)$?

maximum size chain. A chain that you can not increase the size of.

There are $n!$ complete chains in $P(S)$.

Why?

Because the only way to get a complete chain is to start with the null set. I add one element, then I add another element, then I add another element, etc. until I have n elements.

How many ways can I do this?

As many ways as I can order the n elements.

So it's $n!$, the number of permutations.

So the number of chains is $n!$.

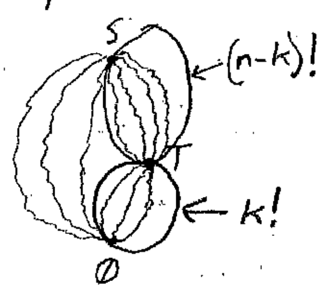
Now, I want to refine this.

Suppose you have a subset T:

$$T \subseteq S \quad , \text{ say } |T| = k$$

Now I ask the following question:

How many complete chains in $P(S)$ pass through T?



There are $k!(n-k)!$ complete chains in $P(S)$ containing the set T.

Why? For the following reason.

Q: How many chains are there from the null set to T?

A: $k!$

Q: How many chains are there from T to S?

A: $(n-k)!$

Let $\mathcal{U}$ be any antichain.

Any complete chain can meet the antichain in at most 1 point.
So how many chains meet some element of the antichain?
Let's count them.

Again, $\mathcal{U}_K$ = set of all sets in $\mathcal{U}$ with K elements

There are $k!(n-k)!$ complete chains that go to any sets with k elements.
Multiply by size of $\mathcal{U}_K$ and you get the most number that meet this antichain.
The sets $\mathcal{U}_K$ are disjoint, so you add it all up.
It's simple addition.
Marble counting.

The number of complete chains meeting the antichain $\mathcal{U}$ is <u>at most</u>:

$$\sum_{K=0}^{n} |\mathcal{U}_K|\, k!\,(n-k)!$$

And we just said that the total number of complete chains in $P(S)$ is $n!$. So we have:

$$\sum_{K=0}^{n} |\mathcal{U}_K|\, k!\,(n-k)! \le n!$$

Divide both sides by $n!$ and you get the binomial coefficient on the LHS:

$$\sum_{K=0}^{n} \frac{|\mathcal{U}_K|}{\binom{n}{k}} \le 1$$

That's the LYM inequality.

Proof of Sperner's Theorem

Now, using the LYM inequality, let's prove Sperner's Theorem.

Well, I just said that the maximum of the binomial coefficient is reached when k is $\left[\frac{n}{2}\right]$. See [22.2].

$$\sum_{k=0}^{n} \frac{|U_k|}{\binom{n}{\left[\frac{n}{2}\right]}} \leq \underbrace{\sum_{k=0}^{n} \frac{|U_k|}{\binom{n}{k}} \leq 1}_{\text{Sperner's Theorem}}$$

$$\underbrace{\sum_{k=0}^{n} |U_k|}_{\substack{\text{this is the size}\\ \text{of } U\text{, because } U\\ \text{is the partition into the } U_k\text{.}}} \leq \binom{n}{\left[\frac{n}{2}\right]}$$

$$|U| \leq \binom{n}{\left[\frac{n}{2}\right]}$$

Q.E.D.

That's the proof, in style.
So, in conclusion, the maximum size antichain in a Boolean algebra is exactly what we think it ought to be.

Now let me tell you of a conjecture of mine that I made 35 years ago.

Conjecture (Rota)

Now I look at the lattice of partitions.

Take $\Pi[S]$

Partitions are ordered by refinement.
This lattice also splits according to levels.
The top level is the partition with 1 block.
The next to the top element is the partition with 2 blocks.
The bottom element is the partition with as many blocks as there are elements of S.

The elements at each level are the number of partitions with k blocks,
And these are the Stirling numbers of the 2^{nd} kind.

The number of Π with $|\Pi| = k$ equals $S(n,k)$ = Stirling number of 2^{nd} kind.

And then you can take a table of Stirling numbers of the 2^{nd} kind.
And, sure enough, they behave like the binomial coefficients.
They increase to a maximum and then they go down.

So it becomes natural to conjecture that the maximum antichain in the lattice of partitions is equal to the maximum of the Stirling numbers of the 2^{nd} kind.

Is the maximum size antichain in $\Pi[S]$ equal $\max_{0 \le k \le n} S(n,k)$?

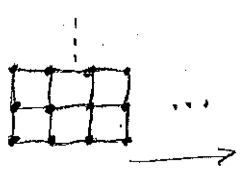
Fortunately, I stated this in the form of a question.
The answer was found, 20 years later, to be No.
But the first counterexample has 10^{10} elements.
In other words – this conjecture is not true. But the smallest set for which it is not true has at least 10^{10} elements. So I'm kind of excused.

This was done probabilistically by Roger Canfield from University of Georgia.

We still don't know the reason why this conjecture is not true.
If you look at the Hasse diagram, what is it that makes it not right?
We still don't know, to this day, the real reason why this does not work.
Something happens when the set S is very large that can not happen before.

Exercise 22.1 (required)

Let's take the partially ordered set $\mathbb{N} \times \mathbb{N}$. It's an infinite partially ordered set.
It looks like this, with the covering relations.



partial ordering goes this way

This is a nice partially ordered set. It's a lattice.

Gordon's Lemma

Every antichain of this $\mathbb{N} \times \mathbb{N}$ lattice is finite.

Prove this as an exercise.

Kultur remarks.
For those of you who know some commutative algebra, Gordon's Lemma is equivalent to the Hilbert Basis Theorem. You can derive it from Gordon's Lemma, Gordon didn't have the concept of a partially ordered set.

The Young Lattice

This is a very nice distributive lattice.
How does a distributive lattice arise?
I remind you [12.3] that a good way of getting a distributive lattice is to take all order ideals of a partially ordered set.

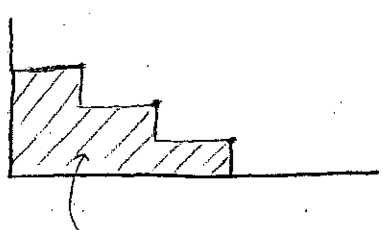
If you have a finite distributive lattice, then it's very easy to prove (you can find this in Stanley's book and my book) that every finite distributive lattice can always be represented as the lattice of order ideals of a partially ordered set.

So finite distributive lattice is the same as lattice of order ideals of a partially ordered set. This is due to Birkhoff.

Now we go to infinite case, but nobody wrote about this.
That is, the profinite point of view.
It's far from trivial.

The Young Lattice is the lattice of order ideals of $\mathbb{N} \times \mathbb{N}$

What does it look like?



An order ideal means you take a certain number of elements and everything below them.

Observe that the order ideal is also a partition of the dominance order. So you can have a bigger order ideal, with the notion of containment.

Take the elements of the sublattice $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$
Finite – just for the sake of the argument.
So you can easily obtain all order ideals of the Young Lattice.

The <u>open problem</u> is to find the Dilworth decomposition of the distributive lattice, as well as the Sperner number (the maximum size).

This is extremely difficult

If you want to find examples of Dilworth decompositions, you can look at my book [GCR on Combinatorics, pp. 563 - 565]. where Metropolis and I have found the Dilworth decomposition of the lattice of faces of the n-cube for all n. It took us the whole summer. That's something I don't like to review. It's a nightmare.

This was immediately generalized, as soon as we published it. And you'll find [ibid, pp. 567 - 570] the generalization. This generalization is as far as the technique that we developed can be carried.
This technique <u>does not work</u> for the Young Lattice.

So, if you want to become famous, find the Dilworth decomposition of the Young Lattice.

Now, let's go back to $P(S)$.
Let's find the decomposition into chains of $P(S)$, now that we know what the maximum size antichain is.
That's the Greene - Kleitman bracketing algorithm. (Greene was a postdoc at MIT a long time ago. He was my first postdoc in combinatorics.)

<u>Greene - Kleitman Bracketing Algorithm</u>

<u>Partition</u> of $P(S)$ into $\binom{n}{[\frac{n}{2}]}$ chains.

Completely explicit.
It goes like this. Some things in combinatorics are best understood by example.

Take $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$
$T = \{1, \quad 3, 4, \quad 7, 8 \quad\}$, $T \subseteq S$

Given this subset T, you want to know to which chain in the Greene - Kleitman partition of $P(S)$ does T belong.

The Greene-Kleitman algorithm gives you the criterion to tell exactly which chain it belongs to.
And you see immediately the number of chains is what it needs to be.

You do it like this.
Write the elements of S. Underneath, write a right parenthesis under every element of the subset $T = \{1, 3, 4, 7, 8\}$. Then, write a left parenthesis under every remaining element.

1	2	3	4	5	6	7	8	9
)	(	)	)	(	(	)	)	(

Now, you match parentheses.

This is called the bracketing according to the subset T.
Greene and Kleitman tell you that the chain to which T belongs in the Dilworth decomposition is exactly the chain containing all subsets which have the same bracketing structure.

Let's see another subset that has the same bracketing structure:

$T' = \{3, 7, 8\}$

1	2	3	4	5	6	7	8	9
(	(	)	(	(	(	)	)	(

T and T' have the same bracketing structure.

The following subsets have the same bracketing structure:

$\{3, 7, 8\}$, $\{1, 3, 7, 8\}$, $\{1, 3, 4, 7, 8\}$, $\{1, 3, 4, 7, 8, 9\}$

Why? Start with subset $\{3, 7, 8\}$

1	2	3	4	5	6	7	8	9
	(	)		(	(	)	)	

All other elements have a left parenthesis.
In order not to change the bracketing structure, we can run right parentheses from left to right. This ensures there are no matching brackets.
That's how we get these sets. And these sets with the same bracketing structure clearly form a chain.

So the subset T is now identified with a chain. You have sets that go from k elements to $n-k$ ($|T| = k$, $|S| = n$).

So it's symmetric. Therefore there has to be an item in the middle.
It's a complete chain.
That's the end of the proof.
Because any two chains are disjoint because the two chains have different bracketing structures.

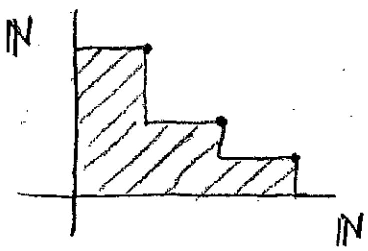
So we have disjoint chains. Any one of them contains an iten in the middle.
And they run from T by n-k.
Therefore, that's it.
That's the decomposition into chains.

Now read my thing with Metropolis, which is a nightmare, if you want to see how to jazz this up.

I was going to do some more matching theory, but it's going to take so much time that I'm going to sketch it and leave the details as required problems.

Let's start with some Kultur,
Last time, we discussed the LYM inequality.
And the Greene – Kleitman Bracketing Algorithm, whereby you partition the Boolean algebra of subsets of a finite set into chains.

And I mentioned the problem of the Young Lattice – the lattice of order ideals of $\mathbb{N} \times \mathbb{N}$.



the order ideal consists of taking points and taking everything underneath.

I mentioned that a very important open problem is the problem of finding a Dilworth partition of order ideals of the Young Lattice.

partition into blocks of chains

You have to find first the maximum antichain of the Young Lattice,
That's already non-trivial.
Then you have to find the blocks,

But, what is interesting is something else,
What is interesting is to consider a complete chain in the Young Lattice and what it looks like,

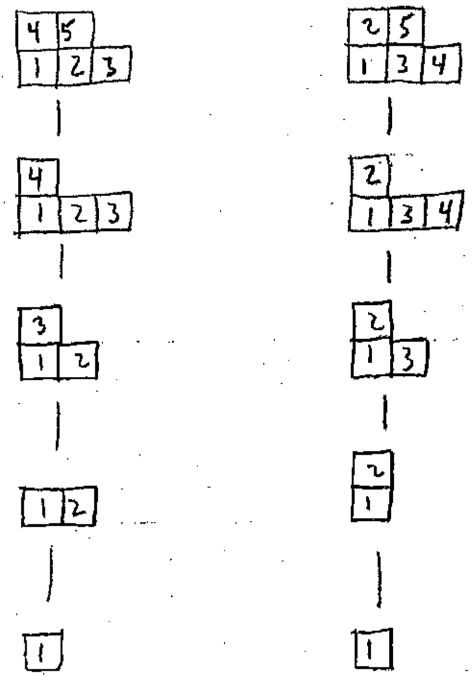
What does a complete chain in the Young Lattice look like?

Let's take a simple case. We'll talk about squares as elements, instead of vertices.
This is the order ideal that corresponds to the Ferrers matrix $\begin{pmatrix} 1 & 1 & \\ 1 & 1 & 1 \end{pmatrix}$:

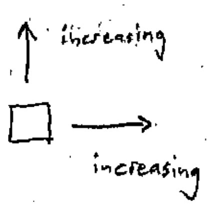
Now I want to take the complete chain, starting from the empty set, and running to this order ideal.
Let's see what this looks like.
It's a very educational experience.
There are many such complete chains
We start we this square.

As we add squares, label the squares in the order we add them.

Examples:



What characterizes these complete chains?
If you look at the top of a chain and the way it has been filled by integers, that characterizes the chain completely.
So the chain is completely determined by the top element filled with integers.
The way the top element is filled is not arbitrary. What's the condition?
The condition is that the integers going to the right along any row are in increasing order and the integers going up along any column are in increasing order:



Conversely, if you take the shape of the top element and fill it, in any way, with the integers 1 to the number of squares in the shape, subject to these two conditions (integers going up a column are in increasing order and integers going to the right along a row are in increasing order), you get a complete chain in the Young Lattice.

Take the 2^{nd} complete chain above.
We have the matrix:

$$\begin{pmatrix} 2 & 5 & & \\ 1 & 3 & 4 & 0 \\ & 0 & & \end{pmatrix}$$

Standard Young Tableau

The matrix has exactly n non-zero entries and consists of integers from 1 to n.
The entries on each row, from left to right, are in increasing order.
And the entries on each column, going bottom up, are in increasing order.

This is a <u>Ferrers matrix</u>, in <u>inverse form</u>, according to its shape.

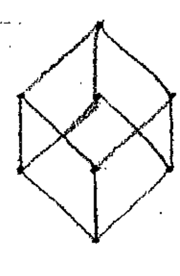
These <u>objects</u> are called <u>Standard Young Tableaux</u> (or Standard Young Diagrams).
these matrices

We just saw the simplest situation where Standard Young Tableaux arise.
Standard Young Tableaux are endemic in combinatorics.
So you want to know what Standard Young Tableaux are.

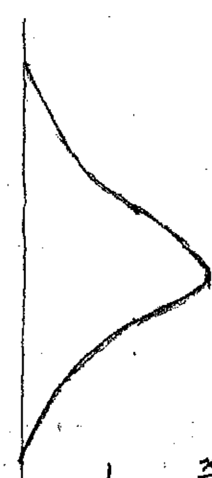
It is non-trivial to count how many Standard Young Tableaux there are of a given shape (i.e., how many chains there are).

- <u>*** Exercise 23.1</u> (Thesis problem)

Suppose we take a finite set S and examine the number of elements in each level of $P(S)$.
For example, the n-cube has:



level	count of number of elements
n	$\binom{n}{n}$
$n-1$	$\binom{n}{n-1}$
$\vdots$	$\vdots$
1	$\binom{n}{1}$
0	$\binom{n}{0}$



$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

If you normalize this property, with the binomial coefficients, this becomes $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. That's called the Central Limit Theorem of probability.

My problem is to construct a continuous lattice where the levels are exactly equal to this.
To take the limit of Boolean algebra, in such a way as to get a continuous lattice, where the levels are exactly equal to this.

I'm sure that this exists.

That would enable us to work with this continuous lattice as a continuous Boolean algebra with e^{-x^2} as the analogue of a measure of a partition.

Matching Theory (conclusion)

Suppose P = finite partially ordered set.

Define a relation R_P as follows:

$$x R_P y \quad \text{if} \quad \underset{\uparrow \text{strictly}}{x > y}$$

Remember, a partial ordering really a relation.

Now we study the deficiency of this relation.
Remember, we started matching theory by proving some results about deficiency.
Let me remind you [18.3 - 18.7]:

We took a relation and considered the deficiency.
We took the minimum deficiency.
Then we proved that if you take the absolute value of the minimum deficiency, you can remove from the relation the number of elements equal to the absolute value of the minimum deficiency.
You get a relation that has zero deficiency and, therefore, has a matching.

So the matching of a relation is obtained by removing a number of judiciously chosen set of elements equal to the absolute value of the minimum deficiency.
That's what we proved before.

Now let's apply this to partially ordered sets.
We want to study the minimum deficiency of this relation R_P.
A very interesting result comes out.

Theorem

The sets of minimum deficiency of the relation R_P are the order ideals of P whose set of maximal elements is a maximum size antichain.

Let's see why this is so:

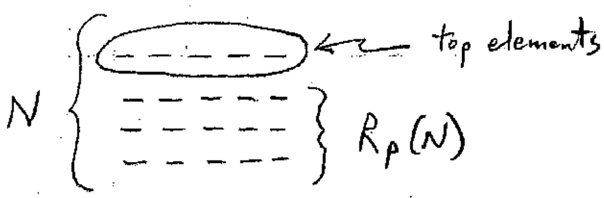
Let N = set of minimum deficiency

I claim that this has to be an order ideal.

if $x \in N$, $y < x$ then $x \, R_P \, y$

The deficiency of N is:

$$\delta(N) = |R_P(N)| - |N|$$



Suppose N is an order ideal including the top elements.
$R_P(N)$ are all elements strictly below the top elements.
The difference in absolute value gives minus the number of elements in the top antichain.

If N is an order ideal then $\overbrace{|R_P(N)| - |N|}^{\delta(N)}$ = minus number of maximal elements of N

So if you want a minimum deficiency, you want a maximum antichain on top

Hence, the conclusion:

$$\delta_0(N) = \min\Big(|R_P(N)| - |N|\Big) = -\max\Big(\text{number of maximal elements of } N\Big)$$

maximum size antichain

So now we have an interesting conclusion,
We found the relation where the set of minimum deficiency corresponds to the order ideal that has the maximum size antichain.

We've shown before that the intersection and union of sets of minimum deficiency are a set of minimum deficiency [18.4].

There is a theorem, due to Dilworth, that if you take the union and intersection of order ideals with maximum size antichains, you again get order ideals with maximum size antichains.
This is a non-trivial fact.

Theorem

If N_1 and N_2 are order ideals whose sets of maximal elements are maximum size antichains, so are:

$$N_1 \cup N_2 \quad \text{and} \quad N_1 \cap N_2$$

Exercise 23.2 (required)

From this fact, applied to the partially ordered set of the Boolean algebra of subsets of a set, you can get a new proof of Sperner's Theorem. [22.3-22.6].

Get a new proof of Sperner's Theorem, using this fact.

Exercise 23.3 (required)

From the theorem about minimum deficiency sets [23.5], get a new proof of Dilworth's Theorem [21.5-21.7], using the main matching theorem we proved before (i.e., the Marriage Theorem).

R^* = inverse relation

The notation R^{-1} for the inverse relation is bad.
For once, I want to change the notation. You should change this in your notes. [2.4]
This was a terrible mistake. I don't know why I did that.
Why is this notation misleading?

Because $R^{-1} \circ R \neq I$

the composition of the inverse relation with the relation is not the identity.

So it's stupid to use R^{-1} as the inverse relation.

It's better to use R^* for the inverse relation.

Exercise 23.4 (required)

Suppose we have a relation R:

$$R \subseteq S \times T \text{ , where } \delta_R = \min \delta(A), \; A \subseteq S$$

We can also define the deficiency of the inverse relation R^*.
What's the relationship the two minimum deficiencies?
The theorem is that they are equal.

Prove that $\delta_R = \delta_{R^*}$

Not hard.
This is an interesting fact.

---

* Exercise 23.5

This is a fairly deep matching theorem that gives you the detailed structure of a relation.

$$R \subseteq S \times T \; , \; |S| = |T| \text{ for simplicity (not really required)}$$

You've already suspected that a relation is sort of a combinatorial analogue of a matrix.

So now, you want to prove the following matching theorem.
There are partitions of the sets S and T as follows:

$$S = N_S \cup R^*(N_T) \cup S_1 \quad \leftarrow \text{union of disjoint sets}$$

$$T = N_T \cup R(N_S) \cup T_1 \quad \leftarrow \text{union of disjoint sets}$$

already non-trivial statements

where:

N_S = minimum set of minimum deficiency of R

N_T = minimum set of minimum deficiency of R^*

such that:

(1) $\underbrace{R\big|_{S_1,T_1}}$ has deficiency 0.

R restricted to S_1 and T_1.
In other words, you take only those edges in the relation R that go from S_1 to T_1.

Since deficiency equals 0, you have a matching.

(2) Every tight set [18.3] of R is of the form:

$$N_S \cup A \quad , \text{ for } A \subseteq S_1$$

$$\left(\text{so that } \delta_R(A) = 0\right)$$

Now let's look at this from the point of view of incidence matrices.

Say $S_1 = T_1 = \emptyset$ for simplicity.
Then, from the preceeding statement about partitions, we have:

$$\begin{array}{cc|cc}
 & & \multicolumn{2}{c}{T} \\
 & & N_T & R(N_S) \\
\hline
\multirow{2}{*}{S} & R^*(N_T) & \text{stuff} & 0 \\
 & N_S & 0 & \text{stuff}
\end{array}$$

We can split this further by taking, from N_S, a subset equal to the size of the minimum deficiency of R.

Take a subset D_S of N_S with $|D_S| = |\delta_R|$.

minimum deficiency of relation R

If you remove this from N_S, the remaining has a matching, by the main matching theorem proved before.

Similarily, take a subset D_T of N_T with $|D_T| = |\delta_{R^*}|$.

$$\begin{array}{rc} & \begin{array}{cccc} & T & & \\ N_T - D_T & D_T & R(N_S) & T_1 \end{array} \\ \begin{array}{r} R^*(N_T) \\ S \quad N_S - D_S \\ D_S \\ S_1 \end{array} & \begin{bmatrix} \underline{\text{matching}} & \text{stuff} & 0 & \text{stuff} \\ 0 & 0 & \underline{\text{matching}} & 0 \\ 0 & 0 & \text{stuff} & 0 \\ 0 & 0 & \text{stuff} & \underline{\text{matching}} \end{bmatrix} \end{array}$$

This is the universal decomposition.
Every matrix, whatsoever, has non zero entries that must be arranged this way. The most canonical, general form is this matrix.
This is the maximum you can do onto a matrix without using linear algebra.

That's the end of the problem.
Prove it.

It's not hard. The only hard part is parts (1) and (2). The rest is easy.

This is a very useful decomposition.
Observe that R and R^* have the same minimum deficiency. [23.7 exercise 23.4]
And these relations always have a matching, as indicated by this matrix.

By the way, this decomposition can also be used to prove Dilworth's Theorem.

Next time, we start on matroids. We're going to do it the following way.
I'm going to use an unusual mathematical device.
I'm going to give a description of the field, without proof. So you see what it is like if you go to the zoo.
Then, once you get a feeling for that zoo, we'll fill in some of the proofs.

Matroids

Matroids turn people off. People are scared of them.
When I wrote my book on matroids, I changed the name. I called it "Combinatorial Geometries" - but it didn't take. They said "that's really matroids, isn't it?"

So, what are matroids?
Let me tell you a dirty little secret.
I've worked on matroids most of my career.
And I don't really know what matroids really are.

Why? Because of all mathematical structures that I know, matroids are the one structure that has the most different definitions.
Completely different definitions that are equivalent.

Because of this variety of definitions, some people think it's this. Some people think it's that.

Different people think it's a generalization of: graphs, projective geometry, 4 color theorem, matching theory, combinatorial topology, invariant theory.

We'll approach matroids from the point of view of matching theory.
And try and get, as quickly as possible, to a very powerful generalization of the Philip Hall matching theorem - the Marriage Theorem.

The Marriage Theorem has an incredible variety of applications.
For example, this theorem tells you when you can find a matching in a relation.
Or a set of distinct representatives in a family of sets.
Suppose you want partial representatives?
Suppose you want to double your representatives?
Suppose you want your set of representatives to be repeated in ways that you prescribe?

Then, matroid theory gives you an automatic way to solve all these problems by getting generalized Hall conditions for each case.
So, in this sense, it's extremely powerful.

So let me show you a kind of matching theorem that matroid theory might lead to, by way of giving you a bird's eye view.
Then, gradually, we'll work up to a definition.

The Theory of Matroids

One day, Professor Alfred Horn of UCLA had an idea.
He said there's an analogy between the Boolean algebra and the lattice of subspaces of a vector space.
It's one of the great analogies of mathematics. Half of mathematics is based on this analogy.

I. Boolean Algebra

$$P(S)$$

$$r(A) = |A|$$

in $P(S)$, you have the rank function that is the number of elements of a subset.

II. Lattice of Subspaces

$$L(V)$$ ← which we represent as a projective geometry using homogeneous coordinates, for visualization purposes.

$$r(W) = \dim(W)$$

Here, the rank function of a subspace W is the dimension of W.

$r(\cdot) = 1$ for atoms.
So the atoms of the lattice are lines.
(They are points in the representation as projective space – dimension goes down by 1.)

Let's go one step further and consider the Triality Principle.
This states that there are 3 lattices in the world and that all other lattices are sort of inf of these.
These 3 lattices are I. Boolean Algebra, II. Lattice of Subspaces, and the third is the lattice of partitions:

III. Lattice of Partitions

$$\Pi[S]$$

$$r(\pi) = n - |\pi| \quad , \; n = |S|$$

↑ number of blocks in partition π

So, if a partition π is an atom, that means that π has one block with 2 elements and all other blocks have one element.

Now, for $P(S)$, we have the Marriage Theorem.
Let me state it in the form of distinct representatives.

Distinct representatives

Given a family $A_1, A_2, \dots, A_K \subseteq S$
we can find a subset $\{x_1, x_2, \dots, x_K\}$ with $x_i \in A_i$
iff

when I say subset, that implies that no two x_i are equal. That's not a set - that's a multiset. So the x_i are automatically distinct

$$|A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_j}| \geq j,$$

for all subfamilies $A_{i_1}, A_{i_2}, \dots, A_{i_j}$

This is a restatement of what we stated in terms of relations.
Because a family of subsets, as I've said many times before, defines a relation.

Most often, the Marriage Theorem is stated in this form.
The Marriage Theorem in terms of relations is slightly more general, because it allows two identical sets.

Now, Professor Horn said - "Gee, what if we try this?"
Instead of the A_i being subsets of S, let's suppose that the A_i are subsets of a vector space.

Say $A_i \subseteq V \leftarrow$ vector space

The A_i are sets of vectors.

Then, it doesn't just make sense to find a set of distinct representatives.
You may ask for a stronger condition.
You may ask for $\{x_1, \dots, x_K\}$ to be locally distinct, but linearly independent.

Q: When is there a subset $\{x_1, \dots, x_K\}$ with $x_i \in A_i$ that is linearly independent?

subset means the x_i are distinct.

The answer is strikingly simple:

A: Iff

$$\dim(\mathrm{span}(A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_j})) \geq j$$

for all subfamilies $A_{i_1}, A_{i_2}, \dots, A_{i_j}$

And this is what Professor Horn proved.
If the A_i are sets of vectors, you replace the $|\cdot|$ with $\dim(\mathrm{span}(\cdot))$ in the iff clause.
Little did he know that there is a more general theorem of this specific case.

This is a beautiful result, with extremely interesting applications of independent representatives.

You've got loads of points, lines, planes - overlapping in funny ways - over a finite field, say. You can pick independent representatives using the necessary and sufficient conditions.

- Philosophy

Let's use the Triality Principle.
The study of $P(S) \to$ set theory
$L(V) \to$ linear algebra
$\Pi[S] \to$ some sort of generalized linear algebra that I've been insisting about for years, which is only partially developed. If we completely understood this, then we would solve the problem of coloring of graphs.
The study of this not completely understood generalization of linear algebra of this lattice is intimately connected to the coloring of graphs, as I promise to show you soon.

So now we observe that we have the following results:

$P(S) \to$ Marriage Theorem
$L(V) \to$ Horn's Theorem
$\Pi[S] \to$?

is there a similar result here?
Well - what do we mean by <u>linearly independent</u>?
We have to define a generalization of the notion of linear independence that goes with the lattice of partitions.
Unless we have that, we can't state this.

I could tell you what the answer is, but, at this point, we might as well go one step further and develop the <u>abstract theory of linear independence</u>, which is the <u>theory of matroids</u>.

We will see that $\Pi[S], L(V), P(S)$ results are special cases of the abstract theory of linear independence.

A <u>rank function</u> r is a set function on S, taking ≥ 0 integer values with the following properties:

(1) $r(\emptyset) = 0$

rank of the null set equals 0

(2) it is increasing. Namely:

if $A \subseteq B$ then $r(A) \leq r(B)$

(3) $r(x) = \begin{cases} 0 \\ 1 \end{cases}, \quad x \in S$

x a single point.
I should really write $r(\{x\}) = \begin{cases} 0 \\ 1 \end{cases}$ but I don't like to write braces.

(4) it is a submodular set function

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

A <u>matroid</u> is a finite set S endowed with a rank function.
A matroid is an assignment of a rank function to a set.
This <u>rank function</u> must be thought of as a <u>generalization</u> of <u>dimension</u>.

Our objective will be to show that most of the properties of dimension hold true.

Now you say, if it's dimension, how come we have inequality here?
We have some easy consequences.

<u>Proposition 1</u>

$$r(A \cup x) \leq r(A) + \begin{cases} 0 \\ 1 \end{cases}, \quad x \notin A$$

<u>Proof</u>: In the submodular inequality (property 4 above), set $B = x$.

$$r(A \cup x) + \underbrace{\cancel{r(A \cap x)}}_{r(\emptyset)=0} \leq r(A) + \underbrace{r(x)}_{r(x) = \begin{cases} 0 \\ 1 \end{cases}}$$

$$r(A \cup x) \leq r(A) + \begin{cases} 0 \\ 1 \end{cases}$$

Theorem 1 (The Whitney Property)

Hassler Whitney was one of the greatest mathematicians of this century.
He invented lots of things. He invented matroids, He invented tensor products, cohomology — you name it.
He belonged to the Whitney family - The Whitney Museum, Eli Whitney.

If $r(A \cup x) = r(A)$ and $r(A \cup y) = r(A)$

then:
$$r(A \cup x \cup y) = r(A)$$

Proof We assume $x, y \notin A$, otherwise the statement is rather trivial.
In the submodular inequality, replace A and B, as follows:

$$A \leftarrow A \cup x$$
$$B \leftarrow A \cup y$$

$$r((A \cup x) \cup (A \cup y)) + r((A \cup x) \cap (A \cup y)) \leq r(A \cup x) + r(A \cup y)$$

$$r(A \cup x \cup y) \quad + r(A) \quad \leq \underbrace{r(A \cup x)}_{r(A)} + \underbrace{r(A \cup y)}_{r(A)}$$

by assumption

$$r(A \cup x \cup y) \quad \leq r(A)$$

From the increasing property (property 2):
$$A \subseteq A \cup x \cup y \Rightarrow r(A) \leq r(A \cup x \cup y)$$
Therefore, we must have equality:

$$r(A \cup x \cup y) = r(A)$$

Theorem 2 (Extended Whitney Property)

If $r(A \cup x) = r(A)$ for every $x \in B$, $A \cap B = \emptyset$

then $r(A \cup B) = r(A)$

Proof

If $B = \{x, y\}$, it's the previous theorem (The Whitney Property).

Say $B = \{x, y, z\}$

From the assumptions:

$r(A \cup x) = r(A)$ $\quad$ $r(A \cup y) = r(A)$ $\quad$ $r(A \cup z) = r(A)$

take all pairs and apply Theorem 1 to each pair.

$r(A \cup x \cup y) = r(A)$ $\quad$ $r(A \cup x \cup z) = r(A)$ $\quad$ $r(A \cup y \cup z) = r(A)$

$= r(A \cup x)$ (given) $\quad$ $= r(A \cup x)$ (given)

Let $A' = A \cup x$:

$r(A' \cup y) = r(A')$ $\quad$ $r(A' \cup z) = r(A')$

apply Theorem 1 (the Whitney Property)

$$r(A' \cup y \cup z) = r(A')$$

$$r(A \cup x \cup y \cup z) = r(A \cup x)$$

$$= r(A) \quad \text{(given)}$$

$$r(A \cup B) = r(A)$$

This proves it for 3 elements.

An inductive argument proves it for arbitrary sized set B.

Proposition 2

$$r(A) \leq |A|$$

Why?
From the definition and proposition 1, we have:

$$r(\emptyset) = 0$$

$$r(A \cup x) \leq r(A) + \begin{cases} 0 \\ 1 \end{cases}, \quad x \notin A$$

Starting with the null set, everytime you add an element x, the rank goes up 1 or 0.
So you can't go up more than the size of the set.

We say that $I \subseteq S$ is <u>independent</u> if $|I| = r(I)$.

↑ relative to the matroid we have chosen.

Observe that this is what happens in linear independence.
A set of vectors is linearly independent if the dimension of the subspace they span is equal to the number of vectors.
So the above statement kind of checks.

Proposition 3

If I is independent and $J \subseteq I$
then J is independent.

<u>Proof</u>

$$r(I) = r(J \cup (I-J))$$

Now we apply the submodular property of a rank, which gives:

$$r(J \cup (I-J)) + \underbrace{r(J \cap (I-J))}_{\substack{J \cap (I-J) = \emptyset \\ r(\emptyset) = 0}} \leq r(J) + r(I-J)$$

$$r(I) \le r(J) + r(I-J)$$

Since I is independent, $r(I) = |I|$:

$$|I| \le \underbrace{r(J)}_{r(J) \le |J|} + \underbrace{r(I-J)}_{r(I-J) \le |I-J|}$$

Both, from Proposition 2.

The only way the inequality $|I| \le r(J) + r(I-J)$ can be satisfied is with the equalities:

$$\underbrace{r(J) = |J|}_{J \text{ is independent}} \quad \text{and} \quad r(I-J) = |I-J|$$

Q.E.D.

Theorem 3 (Exchange Property)

First stated by the great German mathematician Steinitz, who came up with the theory of transcendental extensions of fields.

If I and J are independent sets (relative to a given matroid) and $|I| < |J|$

then there exists $x \in J$, $x \notin I$ such that

$I \cup x$ is independent

Proof

Suppose the conclusion was not true. That $I \cup x$ is not independent, for all x.

If not true, then

$$r(I \cup x) = r(I), \quad \text{for all } x \in J$$

Otherwise, if $I \cup x$ were independent, $r(I \cup x) = |I \cup x| = |I|+1 = r(I)+1$. In other words, if $I \cup x$ were independent, the rank would have to go up by 1, for all x.

By the Extended Whitney Property, that means:

$$r(I \cup J) = r(I)$$

But, by the increasing property of rank:

$$J \subseteq I \cup J \Rightarrow r(J) \leq r(I \cup J)$$

This gives:

$$r(J) \leq r(I \cup J) = r(I)$$

And, since it is given that I and J are independent:

$$|J| = r(J) \leq r(I) = |I|$$

Which gives us our contradiction, since $|I| < |J|$.

A <u>basis</u> is a <u>maximal independent set.</u>

The set corresponds to a vector space.
And, in a vector space, a basis is a maximal independent set.

<u>Theorem 4</u>

Any two bases of the same matroid have the same number of elements.

If B_1 and B_2 are bases of the same matroid
then $|B_1| = |B_2|$.

<u>Proof</u>

I do this by gestures.
If not, one is smaller than the other.
So you have two independent sets, one smaller than the other.
According to the previous theorem, that means you can pick an element from the bigger one and join it to the smaller one.
That means the smaller one is <u>not</u> maximal.
Thus, it's not a basis.

- $\underbrace{(A, r)}_{\text{restrict } r \text{ to } A}$ is a matroid called the <u>restriction</u> to A.

<u>Proposition 4</u>

- In (A, r), every maximal independent set has size $r(A)$.

Any two bases have the same number of elements. Namely, the rank of A.

<u>Proof</u>

Let I = a maximal independent set in A.

That means:

For every $x \in A - I$,

$$r(I \cup x) = r(I)$$

otherwise, you'd have a <u>bigger</u> maximal independent set.
If I is a maximal independent set, then no matter what you add, the rank can not increase.

Therefore, by the Extended Whitney Property (Theorem 2):

$$r(A) = r(I)$$

And, since I is an independent set, $r(I) = |I|$

$$= |I|$$

- <u>Proposition 5</u>

If r is a rank function and $A \subseteq S$,
then r_A defined as $r_A(B) = r(A \cup B) - r(A)$
where: A is fixed and B variable
↑ i.e., any set B.
is also a rank function, called the <u>contraction</u> by A.

Proof
Show that the properties that define a rank function hold. [24.5]

(1) $r_A(\emptyset) = r(A \cup \emptyset) - r(A)$
$= r(A) - r(A)$
$= 0$ ✓

(2) if $B \subseteq C$ then $r_A(B) \leq r_A(C)$
This is trivial to show. ✓

(3) $r_A(x) = \underbrace{r(A \cup x)} - r(A)$
from Proposition 1, $r(A \cup x) \leq r(A) + \begin{cases} 0 \\ 1 \end{cases}$
$\leq \cancel{r(A)} + \begin{cases} 0 \\ 1 \end{cases} - \cancel{r(A)}$
$= \begin{cases} 0 \\ 1 \end{cases}$ ✓

(4) $r_A(B \cup C) + r_A(B \cap C) \overset{?}{\leq} r_A(B) + r_A(C)$
All we have to do is write this out long hand, using the definition:

$$r(\underbrace{A \cup B \cup C}_{(A\cup B)\cup(B\cup C)}) - \cancel{r(A)} + r(\underbrace{A \cup (B \cap C)}_{(A\cup B)\cap(B\cup C)}) - \cancel{r(A)} \overset{?}{\leq} r(A \cup B) - \cancel{r(A)} + r(A \cup C) - \cancel{r(A)}$$

$$r((A \cup B) \cup (B \cup C)) + r((A \cup B) \cap (B \cup C)) \overset{?}{\leq} r(A \cup B) + r(A \cup C)$$

This is just a specific case of the submodular law.
So we remove the question marks all the way back. ✓

So, given a rank function and a set, there are two rank functions you can derive from it: the restriction to a set and the contraction by a set.
These have the following correspondences:

restriction $\Rightarrow$ subspace
contraction $\Rightarrow$ quotient space

John Guidi
guidi@math.mit.edu

Theory of Matroids (cont'd)

Let's begin by reviewing the theory of matroids:

A matroid is a pair, consisting of a finite set S and a set function r, called the rank function.

The rank function is a function from the subsets of S to the non-negative integers, with the following properties:

(1) $r(\emptyset) = 0$

(2) if $A \subseteq B$ then $r(A) \leq r(B)$

(3) $r(x) = \begin{cases} 0 \\ 1 \end{cases}$ if $x \in S$

(4) r is submodular

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

Then we proceeded to develop some elementary properties of matroids. To wit, we showed:

$$r(A) \leq |A|$$

We say that:

$I \subseteq S$ is independent when $r(I) = |I|$

And we showed that the family of independent sets has the following properties:

(a) any subset of an independent set is independent

(b) if I and J are independent and $|I| < |J|$ then there exists $x \in J - I$ such that $I \cup x$ is independent. (The Exchange Property)

Just like in linear algebra.

Exercise 25.1

$\mathcal{I}$ = family of subsets satisfying (a) and (b) above.

Under these conditions, we define a rank function to be the maximum size of an independent subset of A.

Define $r(A)$ = maximum size of an element of $\mathcal{I}$ contained in A.

Prove that r is a rank function.

$\mathcal{I}$ satisfying (a) means:

if $I \in \mathcal{I}$ and $J \subseteq I$

then $J \in \mathcal{I}$

This is the first of the many possible alternative definitions of matroids.
And you begin to see why matroids are sort of strange.
Because you can take any concept defined and you can use that concept to give a new definition of a matroid.
So there are infinitely many definitions.
People keep discovering new ones.
Some people like one better than the other,
And they quarrel that one is better than the other.

We also saw that:

A basis is a maximal independent subset

And we showed that a basis has the following properties:

(α) any two bases have the same size

(β) if B_1 and B_2 are bases, $x \in B_1$, there exists $y \in B_2$ such that:

$$(B_1 - x) \cup y \text{ is a basis}$$

This is an immediate consequence of Proposition 3 of independent sets [24.8]

Exercise 25.2

Given a set S and a family $\mathcal{B}$ of subsets satisfying (α) and (β) above.

Say I is independent if I can be extended to an element of $\mathcal{B}$.

Then there exists a unique rank function for which the element of $\mathcal{B}$ exists:

Then $\mathcal{B}$ are all the bases of some matroid.

In other words, you can axiomitize matroids in terms of bases.

Give this proof.

We saw, last time, the Whitney Property about the rank function of a matroid. [24.6]

The Whitney Property

If $r(A \cup x) = r(A)$ and $r(A \cup y) = r(A)$
then $r(A \cup x \cup y) = r(A)$

We saw that this was an easy consequence of the submodularity and increasing properties.

And we saw that the Whitney Property implies the Extended Whitney Property:

Extended Whitney ~~Property~~

If $r(A \cup x) = r(A)$, for all $x \in B$,
then $r(A \cup B) = r(A)$

Now let me state the Theorem of Whitney, which in a sense is the converse of this.

Theorem (Whitney)

Let μ be a set function s.t.

(1) $\mu(\emptyset) = 0$

(2) $\mu(A \cup x) = \mu(A) + \begin{cases} 0 \\ 1 \end{cases}$, for $A \subseteq S$ and $x \in S$

(3) if $A \subseteq B$ then $\mu(A) \leq \mu(B)$

(4) μ has the Whitney property

Then μ is a rank function. (See also [26.4])

so μ is submodular

Exercise 25.3

Prove Whitney's Theorem.
The proof is deferred, because it's dull.
I tried to get a cute proof last night, but I couldn't get it.
Please try and get a cute proof of this.
There should be a one line proof, but I don't have it.
I have an induction proof.

At any rate, this is one way of checking the structure of a matroid.
At this point, let us see examples of matroids.

There are two kinds of examples.
There are the intended examples and the unintended examples.

some extremely weirdo structures turn out to be matroids.

It is the unintended examples that make the theory interesting.
If you just had the intended examples, it would just be linear algebra.

Let's look at the intended examples first.
There are three:

(1) sets of points in projective space
(2) arrangement of hyperplanes
(3) graphs

Example 1 – projective space of dimension n

For those of you who know some algebra, this can be projective space over any field.
In particular, the interesting case is the field with two elements.
The projective space over a field with two elements is a very important example.

Take any finite subset $S \subseteq \mathbb{P}^n$

Then, on that, define a rank function, as follows:

$$r(A) = \dim(\text{span}(A)) + 1 \quad , \quad A \subseteq S$$

why +1?
Because the rank of a point we want to be 1.

So we go back to the origins of projective space, which is really subspaces of a vector space.

I claim that this is the rank function of a matroid.

And you say what I call dimension has equality.
But, I say when you take a subset S, then the equality fails.

Let me tell you, intuitively, what's going on here.
Suppose you take the rank of $A \cup B$:

$$\begin{aligned} r(A \cup B) &= \dim(\text{span}(A \cup B)) + 1 \\ r(A) &= \dim(\text{span}(A)) + 1 \\ r(B) &= \dim(\text{span}(B)) + 1 \end{aligned}$$

You are tempted to write:

$$r(A \cup B) + r(A \cap B) = r(A) + r(B)$$

equality

But that would be wrong!

Why?
From $r(A \cup B)$, you get the term (ignore the +1, which cancels):

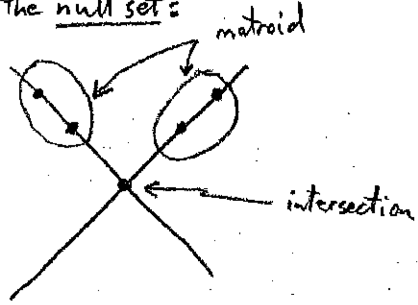
$$\dim(\text{span}(A \cap B))$$

Note that:

$$\text{span}(A \cap B) \subseteq \text{span}(A) \cap \text{span}(B)$$

In general, this will not be equal.
There may not be enough points to go around.

For example, you may have two points on a line, and another line with two points. The matroid is these 4 points. But the intersection of these is the null set:



It turns out that the best you can have is inequality.
In other words:

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

inequality

That's the intuitive argument.
Now, rigorously, let's use Whitney's Theorem to prove that our r, so defined, is a rank function. That way we don't have to reason about intersections not being big enough, etc.

- By Whitney [25.4], you immediately see that the rank I have defined:

$$r(A) = \dim(\text{span}(A)) + 1$$

satisfies conditions 1 and 3.

For condition 2: if you add a point in the span of A, the dimension does not change.
if you add a point not in the span of A, the dimension goes up by 1.

Finally, we check that our function satisfies the Whitney Property:

We are given that:

$$r(A \cup x) = r(A)$$

With our function, this means that:

$$\dim(\text{span}(A \cup x)) = \dim(\text{span}(A))$$

This means that x is in the span of A, by linear algebra.

$$x \in \text{span}(A)$$

Similarly, we have:

$$r(A \cup y) = r(A) \Rightarrow y \in \text{span}(A)$$

So, the Whitney Property is fairly trivial.
If $x \in \text{span}(A)$ and $y \in \text{span}(A)$
then:

$$x \cup y \in \text{span}(A)$$

and $r(A \cup x \cup y) = r(A)$

Therefore, by Whitney's Theorem, all the properties are satisfied and we conclude that our function is, indeed, a rank function.

- Note that you have to distinguish between span in the sense of linear algebra (vector space) and span in the sense of projective space.
This is a classic story. It comes out in my book.

If A is a single point, then the span, in the projective space, will be the point:

A = a point, then $\text{span}(A) = A$
$\dim(\text{span}(A)) = 0$

I add one to the function because I want the rank to be 1 for a point:

$$r(A) = \underbrace{\dim(\text{span}(A))}_{0} + 1$$

$$= 1$$

If you want span in the linear algebra sense, then the span of every point is a line, because you have homogeneous coordinates.
The reason we take projective space is that we like to take the span of points as points.

This is the classic story when you switch between vector space and projective space. You get into this crisis where points (in projective space) are really lines (in vector space). Remember that points are given by homogeneous coordinates.
This is an old story.

When interpreting dimension in the projective sense, then the dimension of a set of points is that of the span of all the linear combinations of the affine set.
For example, if you take two points p and q in <u>projective space</u>:

The span of these two points is the set of all points satisfying:

$$\lambda p + (1-\lambda) q \quad , \quad \lambda \in \mathbb{R}$$

So the <u>span</u> is a <u>line</u>. And the dimension is 1.

$$\left.\begin{array}{l}\text{The dimension of a point is } 0. \\ \text{The dimension of a line is } 1.\end{array}\right\} \text{ in projective space}$$

Example 2 – arrangements of hyperplanes

This is mathematically identical to the preceeding example, but psychologically quite different.

You all know that the dual of a vector space is a vector space.
And an element of the dual of a vector space is a hyperplane.

An arrangement of hyperplanes is a set of hyperplanes, whose elements are dual of the vector space.
And you define the rank, as you did before, in the dual.

Let $\mathcal{H}$ = set of finite hyperplanes

A hyperplane H has dimension $n-1$ in projective space.

If you use homogeneous coordinates, parallel hyperplanes have different homogeneous coordinates, because the coordinate at infinity is different.
Parallel hyperplanes meet at infinity.
(to do linear functions, you have to go back to the vector space)

So, mathematically, you consider the hyperplanes as points in the dual space.
And then you can define a rank function....

But, let's pretend you don't know that.
Since the hyperplane has dimension $n-1$, we define the rank of a hyperplane to be 1:

$$r(H) = 1$$

Then we consider the rank of a set of hyperplanes $r(\{H_1, H_2, \dots, H_k\})$.
If this were a set of points, it would be the dimension of the span.
How do you "dualize" that?
You take the intersection of the hyperplanes, then subtract the dimension of the intersection from n:

$$r(\{H_1, H_2, \dots, H_k\}) = n - \dim(H_1 \cap H_2 \cap \cdots \cap H_k)$$ codimension

When you think about it, this is just "upside down" linear algebra. You don't say anything, but people like to think this way. When you think hyperplanes, you think different questions. For example, a good question to ask is:

Q: Given a set of hyperplanes, how many regions of space are determined by these hyperplanes?

A: This is a number that is computed with a matroid.
A very reasonable computation, done by a student at MIT. Zaslowsky?

Example 3 - graphs

This is the original example of a matroid.

Let's do some ideal history.
This is how Whitney should have thought about this.
Not the way he thought about it, but the way he should have.

I told you, at the begining of this chapter on matroid theory [24.2], that there are three major lattices:

(1) lattice of the subsets of a set
(2) lattice of the ~~subsets~~ spaces of a vector space
(3) lattice of partitions

For subsets of a set, we have trivial matroids.
For ~~subsets~~ spaces of a vector space, we have these rank functions that are non-trivial.
Matroids of partitions are the most interesting - and the least understood.
That's the graph coloring problem. Matroids latch on to graph coloring.

What did we do in the case of a vector space?
We worked in projective space because we like to deal with points.
But, they are really lines in vector space.

Here, we do the same thing.
We take the set of all atoms in the lattice of partitions.
And we pretend these are points.
And we see there is a matroid structure defined on the set of all atoms.

Let $\Pi[T]$ = family of all partitions of the set T

S = set of all atoms of the lattice $\Pi[T]$

Now, we define a rank function on this set of atoms.
I'll tell you what it is and then we'll check that it is, indeed, a rank function.

Recall that in $\Pi[T]$ that there is a rank - the lattice rank [11.2]. & [12-6]
Namely:

$r_\ell(\text{atom}) = 1$

The zero element is the partition with as many blocks as there are elements of T. An atom covers the zero element, so atoms are partitions that have one block with 2 elements and all the other blocks have 1 element.

The rank of any level is:

$r_\ell(\text{level}) = n - (\text{number of blocks of partitions of the level})$

Let $A \subseteq S$

Set $\mu(A) = r_\ell(\vee A)$
$\uparrow$ sup

I claim that this defines a matroid.

Let's show that μ defines a rank function, by using Whitney's Theorem [25.4].
This matroid involves rank in two senses. One is the rank μ of the matroid.
The other is the lattice rank r_ℓ. ↳ defined on set of atoms

Conditions of Whitney's Theorem:

(1) $\mu(\emptyset) = r_\ell(\vee \emptyset)$
$= 0$ ✓

(2) $\mu(A \cup x) = \mu(A) + \begin{cases} 0 \\ 1 \end{cases}$

x is an atom in the lattice $\Pi[T]$.

A is a set of atoms.
What happens when you add an extra atom (x) to this set of atoms (A)?
Either:

a) the number of blocks remains the same:

$$\mu(A \cup x) = \mu(A) + 0$$

-or-

b) the new atom x joins two blocks that were not previously joined. In which case the number of blocks goes down by 1. So the rank, then, goes up by 1:

$$\mu(A \cup x) = \mu(A) + 1$$

So this checks. ✓

(3) if $A \subseteq B$ then $\mu(A) \leq \mu(B)$
obvious ✓

(4) The Whitney Property.

Suppose that $\mu(A \cup x) = \mu(A)$

This means that:

$$r_\ell(\vee A \vee x) = r_\ell(\vee A)$$

But what does this mean, lattice theoretically?
It means that x is underneath sup of A:

$$x \leq \vee A$$

Similarly, $y \leq \vee A$. And $x \vee y \leq \vee A$
And, therefore, it follows that:

$$r_\ell(\vee A \vee x \vee y) = r_\ell(A)$$

$$\mu(A \cup x \cup y) = \mu(A) \quad \checkmark$$

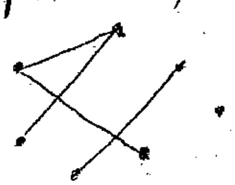
So the Whitney Theorem conditions are satisfied.
Therefore, we have a matroid.

- Now you say: "What does this have to do with graphs?"
Good question.

You remember we said that given any matroid and any subset $A \subseteq S$ we restrict the matroid to A and we get a matroid, trivially. [24.11]

If we do this restriction here, we get a matroid, trivially.
But the interpretation of it is non-trivial.

Take a set of atoms.
An atom has a 2 element block.
So I represent it by an edge.



T

A set of atoms is a set of edges.
That's called a graph.

So, any set of atoms is a graph.
And we just said that the restriction of every subset of a matroid is a matroid.
Therefore: Every graph defines a matroid.

• Now you say: "This is fine. But how do I visualize it?"
Fine. Let's see the classical way of visualizing it.

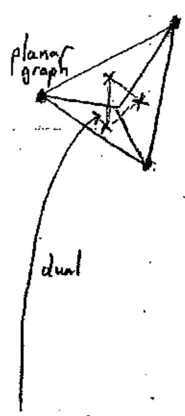
Any graph defines a matroid.
The graph, to repeat, is interpreted as a set of atoms in the lattice of partitions.
How can we <u>visualize</u> this matroid?

To visualize it, we associate to every matroid a lattice, just like the lattice of partitions.
So, you want to associate a lattice to the matroid obtained by taking a subset of T.
This is called the <u>lattice of contractions</u> on a graph.
And we'll see this next time.

• Since we have 3 minutes left, let me mention the original motivation of Whitney, which we will come to.
His thesis advisor, Professor Birkhoff at Harvard, gave him, as a thesis problem, solve the 4 color conjecture.
He was his best student. And he did the best he could.

In fact, his paper, called "The coloring of graphs", is still remarkable today.
Because, implicitly, he discovered the concept of the Hopf algebra.

He hit upon a difficulty with the 4 color conjecture that is concerned with the planar graph:



That you can always color the vertices with any one of 4 colors, in such a way that 2 adjacent vertices never have the same color.

They say this is proved.

But all the proofs are by computer and they always have an error.
Later corrected by another proof, which turns out to have an error.

That's the way it turns out, so far.

So Whitney said, the big thing about planar graphs is that every planar graph has a dual graph.
Namely, you place a little x in the middle of each region.
And you join two crosses when you go across the region.

Then he said, that's funny, if the graph is not planar, then you don't have a dual graph.
So he invented the generalization of a graph, called the matroid.
And he showed that every graph has a dual matroid, which is a planar graph.

That's how he invented matroids.
By associating dual objects with graphs.

For the coloring problem, we need this concept of dual matroid, which is coming.

Next time, we'll discuss dual matroids, with these obvious examples.
And we'll do Rado's Theorem, which is the generalization of Hall's Theorem of the matroid.

Then we'll start kicking in with the non-standard examples of a matroid.
Let me give you a hint of what's coming.

Suppose you have a relation:

$$R \subseteq S \times T$$

Then, you can define a set function:

$$\mu(A) = |R(A)|$$

And we proved that this was submodular.
However, this is not the rank function of a matroid.
I'm sorry.

Even if you take the deficiency:

$$\delta(A) = |R(A)| - |A|$$

We proved that this was submodular.
This is also submodular.
But this is not the rank function of a matroid either.

However, there is a normalization theorem where you can touch up these ($\mu(A)$ and $\delta(A)$) and make them into rank functions, by an extremely clever trick.
This was discovered by British mathematician Nash-Williams.

So, you get matroids out of any relation.
What good is that? You can apply Hall's Theorem and get fantastic generalizations of Hall's Theorem.

I must say that I will do Möbius functions from a very high level point of view next term in a course called Multilinear Algebra.
Where Möbius functions come out of antipodes of Hopf algebra.

This is the fanciest way I know of studying Möbius functions.

Original example of a matroid

The original example of a matroid is not one of the examples I mentioned last time. It is the following:

Take a rectangular matrix.
Then the set S = set of columns of this matrix.
Then you define a rank function, where the rank of a subset of columns is the rank of that subset of vectors.
The column set is a set of vectors.
That obviously defines a rank function, because the vectors are points in projective space.

That's why Whitney called it a matroid.
Because it's like a matrix. The columns of the matrix are used to determine the rank function.

This is also a very good way of visualizing facts about matroids.

Furthermore, this example of Whitney's is used to state one of the great working areas in the theory of matroids, which is the following:

You are given a matroid.

namely, a finite set with an abstract rank function, with the properties we've discussed.

Then, there is a problem of representation of the matroid.

The problem of representation of a matroid, given a matroid, is to find a matrix such that the rank of the columns coincides with the abstract rank function of the matroid.

This is the problem on which the deepest work on matroid theory has been done, by one of the greatest combinatorialists of all time, namely W.T. Tutte.

W.T. Tutte, at the age of 17, worked at Bletchley Park, during World War II, in the group that was led by Alan Turing that cracked the German code Enigma. The credit for cracking the German code is usually attributed to Turing. That is not true. The credit is Tutte's.
As a matter of fact, if you read any books on the German code, they say a 17 year old boy made the crucial step in cracking the Enigma.

So, at the end of World War II he was going to go home somewhere in England and someone said "Don't go home. You're being awarded a fellowship at Trinity College."

So he went to Trinity College and studied math and wrote his thesis where he reinvented matroids.
He didn't know about Whitney.

He solved some very deep problems on the representation of matroids.
Namely, given an abstract rank function of a matroid, when can you find a matrix whose rank of columns coincide.

Representation Theory is something that is beyond this class.
Professor Stanley, next year, will be teaching 18.315 and he will be developing the theory of hyperplanes. So, in the process, you will probably do a lot of matroid theory.

But, I will mention to you what the most important representation theorems are.

Most Important Representation Theorems

(1) When can you represent a matroid as a matrix whose vectors have components belonging to the field of 2 elements?

This is easy to solve.

(2) When can a matroid be represented by vectors over any field, whatsoever?

The answer was given by Tutte. I will tell you what the answer is in a little while.
This turns out to be the same as the following problem:

When can a matroid be represented by a matrix that is totally unimodular?

Here, again, you have the theory of totally unimodular matrices creeping in. [9.9-10.3]
There is something very important about totally unimodular matrices, which we don't fully understand.

I remind you, as a fact, that totally unimodular matrices are matrices, all of whose minors are equal to $+1, -1,$ or 0. [9.9]

More recently, Professor Seymour of Princeton has proved a very good theorem that says that practically all totally unimodular matrices can be obtained from matrices associated with graphs.

The next result that was proved by Tutte is:

When can a matroid be represented as a matroid of a graph?
{ In the sense that we established last time,
And I'm going to go through that, again, today. }

Lastly, Tutte solved the problem:

when can a matroid be represented as a matroid of a <u>planar</u> graph?

With this, he rediscovered the Theorem of Kuratowski about when a graph is planar.

These are the famous Tutte theorems of matroids.

Now, you may ask where do I come in.
The reason I got interested in matroids is that every matroid gives you a generalization of the <u>problem of coloring a graph.</u>

You can't solve the problem of coloring a graph by taking colored pens and coloring vertices all your life. You have to think through it, in case the problem is a wide enough conjecture or theorem, so that you see what the problem is really about.
That's how mathematical problems get solved.

Remember what the great mathematician George Pólya wrote:
"No mathematical problem is ever solved directly."

In other words, you don't solve a problem by starring at it.
You have to look at the sides.

So, that's how I got interested in matroids in the 1960's.
The generalization of the coloring problem to arbitrary matroids is called <u>The Critical Problem.</u>

We still don't have the answer to this right now.
What's missing is a <u>super homology theorem.</u>

we know, vaguely, what ought to be right.
But I'm just too old.

By the way, an interesting problem for a child coming into combinatorics is not to solve it, but to set up the machinery for the Critical Problem.

I hope, in this course, that we go far enough where I state the Critical Problem, using Möbius functions.

Graphic Matroids (cont'd)

I'd like to develop a little further our intuitive understanding for graphic matroids, as defined last time. The concept evades you and it takes quite a while to get used to it.

We define a matroid as a set S with a rank function r.

$$(S, r)$$

set ↗ ↖ rank function [24.5, 25.1]

The rank function has the properties:

(1) $r(\emptyset)=0$

(2) increasing
if $A \subseteq B$ then $r(A) \leq r(B)$

(3) $r(x) = \begin{cases} 0 \\ 1 \end{cases}, x \in S$

(4) submodular
$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

Then I stated, without proof, the Theorem of Whitney [25.4]:

Theorem (Whitney)

μ (a set function) is a rank function iff:

(1) $\mu(\emptyset) = 0$

(2) $\mu(A \cup x) = \mu(A) + \begin{cases} 0 \\ 1 \end{cases}$, where x is a one element set of S.

(3) μ increasing
if $A \subseteq B$ then $\mu(A) \leq \mu(B)$

(4) μ has the Whitney Property
if $\mu(A \cup x) = \mu(A \cup y) = \mu(A)$
then: $\mu(A \cup x \cup y) = \mu(A)$

These 4 properties imply that the set function μ is submodular and, therefore, a rank function.
The proof that such a μ is submodular is a dull proof. I haven't been able to simplify it. So I will defer it.

Whitney's Theorem is useful to establish that a structure is a matroid.
It's easier, sometimes, to check the conditions of Whitney's Theorem then it is to check that the rank function is submodular.
We saw that in the examples last time (e.g., matroids in projective space [25.6-7]).

- Note that an immediate consequence of Whitney's Theorem is that if the conditions are satisfied and μ is a rank function, then we have:

$$\mu(A) \leq |A| \qquad [24.8, \text{Proposition } 2]$$

$$\mu(I) = |I| \text{, such sets } I \text{ are called } \underline{\text{independent.}} \quad [24.8]$$

- In the case of a matroid being represented by a matrix, where the rank of the columns of the matrix coincide with the rank of the matroid, the set is independent if the columns are actually independent vectors.
And you can find a basis by finding a maximal independent set. [24.10]

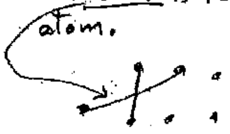
We have begun to study <u>graphic matroids</u> [25.10-14]
Using our Triality Principle [24.2] view, we take the lattice of partitions of T.

Let $\Pi[T]$ = family of all partitions of the set T
S = set of all <u>atoms</u> of the lattice $\Pi[T]$

What's an atom?
An atom covers the zero element.
What's the zero element?
The <u>zero element</u> is the partition where every element belongs alone to one block. Every block has one element.

So an atom means that you have one block with 2 elements and all other blocks have 1 element.

It is customary to represent the set S as the set of all edges on the complete graph T.
An <u>atom</u> is represented by an edge such that the edge is the non-trivial block of the atom.



T

Elements of S are represented by edges of the complete graph on the vertex set T.

Now let's define the matroid.
And let's interpret all concepts pertaining to the matroid in terms of graphs.

It is important to remember that our definition of the matroid depends on partitions.
We are talking about partitions and the <u>graphic representation</u> is due to our <u>human weakness</u>. Not that it should be.
It's really partitions we are talking about.
But because we can't visualize partitions, we like to draw cute graphs instead.

As we saw last time, if we have a subset A, the rank of A is the <u>lattice rank</u> of the sup of A:

$$A \subseteq S$$

$$r(A) = r_\ell(\overset{\text{sup}}{\vee} A)$$

r_ℓ = <u>Lattice rank</u> = n − number of blocks of partition

$$\left\{\begin{array}{ll} r_\ell(\text{top element in lattice}) & = n-1 \\ r_\ell(\text{atom}) & = n-(n-1) = 1 \\ r_\ell(\text{zero element in lattice}) & = n-n = 0 \end{array}\right\}$$

We verified, last time [25.11-12], that $r(A)$ so defined satisfies the conditions of Whitney's Theorem and, thus, is a rank function.
(S, r) defines a matroid.

In particular, you can take a subset of S and restrict the matroid to the subset of S.
This subset of S would be <u>a set of edges on the complete graph.</u>

that's called a graph – plain and simple.

Therefore, every graph defines a matroid, which is the restriction of this "imperial majesty" matroid (S, r) to a subset of S.

Therefore, we only have to study this concept for the lattice of partitions and <u>automatically</u> they're defined for every graph, by restriction.

What's an <u>independent set</u> of this matroid (S, r)?

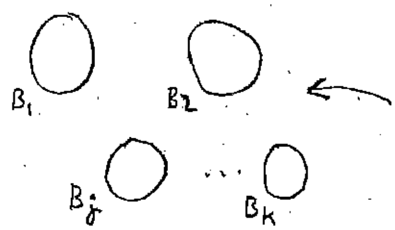
When is $r(A \cup x) = r(A)$? [24.11, Proposition 4]

That's the only good question to ask to understand the nature of matroid, because of Whitney's Theorem.

Well – let's think <u>partitions</u>.

A = a set of atoms.
You're joining them and taking the equivalence relation, which is the sup of the underlying equivalence relations.

So, to ask when is $r(A \cup x) = r(A)$ is the following.
We start with the equivalence relation whose blocks are given by A.

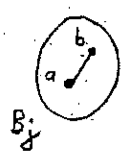


Then, you add x, which is an atom:

$x =$

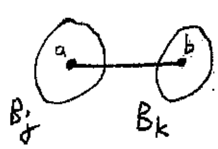
Only 2 things can happen:

<u>Case 1</u>: a and b belong to the <u>same block</u> of A



The rank does not change. $r(A \cup x) = r(A)$

<u>Case 2</u>: a and b belong to <u>different blocks</u> of A



In which case, the blocks are joined, and the number of blocks goes down by 1. This causes the rank to go up by 1.

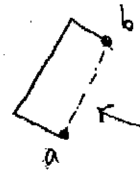
$$r(A \cup x) = r(A) + 1$$

So, only with case 1 do we have $r(A \cup x) = r(A)$,

$r(A \cup x) = r(A)$ iff both endpoints of the edge x belong to the same connected component of the set A of edges,

- So, <u>graph theoretically</u>, the set A is pictured like this:

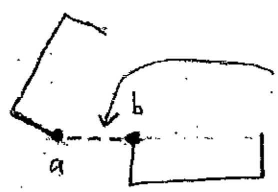
<u>case 1</u>: $r(A \cup x) = r(A)$



same connected component.
adding x forms a <u>cycle</u>.

$A \cup x$

<u>case 2</u>: $r(A \cup x) = r(A) + 1$



two different connected components joined into one after adding x.

$A \cup x$

In this way, we get an intuition of this kind of matroid.
So we can immediately tell now what the independent sets look like.

- The independent sets are the <u>trees</u>.

Why? Consider how you "<u>grow</u>" an independent set.
As you add one element after the other, the rank has to keep going up, each time, by 1. Since $r(I) = |I|$ for an independent set I, you can <u>not</u> afford to lose any rank. You require $|I|$ iterations of case 2 to "grow" independent set I.
This means that you can <u>never</u> close and form a <u>cycle</u> with an independent set.
Any independent set <u>must</u> be in the form of a tree.

path I

T

- Then what's a basis?

A basis is a maximal spanning tree.

↑ in classical graph theoretic terminology, a spanning tree is a tree such that when you add any edge, both endpoints belong to the same connected component.
You can't add any more edges without closing a cycle.
That is, once you have the maximal spanning tree, adding any additional edges satisfies case I, above.
Namely:

$$r(I \cup x) = r(I)$$

With this, we have two non-trival results in our hand, immediately.

- (1) Any two maximal spanning trees of a graph have the same number of edges.

We already proved that any two bases of the same matroid have the same number of elements [24.10, Theorem 4].
So this result follows immediately.
Bases are maximal spanning trees.

So you get, cheapo, this result.

- (2) Exchange Property of Independent Sets

Suppose I have one spanning tree with j elements and one spanning tree with $j+1$ elements.
That means there is one element of the larger spanning tree that can adjoin to the smaller spanning tree and the result is still a spanning tree.
That's the Exchange Property of Independent Sets.

That's it. You get this cheapo.
It's hard to prove this geometrically. I don't know how to prove it that way.

So that's the 3rd intended example of matroids — graphic matroids

Rado's Theorem

This is the analogue of the Marriage Theorem for matroids.
I will state it in terms of a system of independent representatives.

Given a matroid (S, r) and a family of subsets $A_1, A_2, \dots, A_K \subseteq S$, we want:

$$x_i \in A_i \quad \text{s.t.} \quad \{x_1, x_2, \dots, x_K\} \text{ is } \underline{\text{independent.}}$$

in particular, these x_i are distinct

When can we do this?

Such a <u>system of independent representatives</u> exists iff for every subfamily $A_{i_1}, A_{i_2}, \dots, A_{i_j}$, we have:

$$r(A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_j}) \geq j$$

For example, suppose you want to apply this to graphs.
What does this tell us?

Example - graphs

A_i = family of edges of a graph

You are given a family of edges of a graph A_i.
Let me repeat, again, that you need not take the complete graph A. You take any subfamily A_i of A.
You take the restriction of the matroid to form the family of edges.

You take any family of edges of the graph and you want that, as we've just discussed:

$$x_i \in A_i \quad \text{s.t.} \quad \{x_1, \dots, x_K\} \text{ form a } \underline{\text{tree.}}$$

independent representatives

Rado's Theorem tells you <u>when</u> you can <u>form this tree.</u>
Namely, when:

for every subfamily $A_{i_1}, A_{i_2}, \dots, A_{i_j}$ of A_i, we have that:

$$r(A_{i_1}, A_{i_2}, \dots, A_{i_j}) \geq j$$

To prove this (i.e., when a tree can be formed) directly is a mess.
It's easier to prove this general theorem by matroids.

Example - vector space

Given a set of points in a vector space, can you find a subset of these that are independent?

From Rado's Theorem, the answer is iff for every subset of this set, we have:

$$\underbrace{r(A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_j})} \geq j$$

recall that the rank function for a vector space involves dimension. See projective space/vector space discussion. [25.5-8].

That is, whenever the dimension $\geq$ number of elements in the set,

There are, of course, many other applications of Rado's Theorem. Before we prove this, I have to remind you of some concepts:

Restriction

The restriction of a matroid involves taking a matroid.
Then taking a subset.
And you just look at this subset.
You restrict to the subset.

So the restriction of a matroid corresponds to a subset.

Contraction

Contraction is the matroid analogue of a quotient space.

Given a matroid (S, r) and a subset $A \subseteq S$, the contraction by A is the matroid $(S-A, r_A)$, where:

$$r_A(B) = r(A \cup B) - r(A)$$

We have already verified that r_A is a rank function. [24.11-12, Proposition 5]
In particular, I can restrict it to $S-A$.

It may be worthwhile to visualize contractions in the case of the lattice of partitions
For example, for partitions, what do contractions look like?
This is something I should have told you before.

For $\Pi[T]$, what's a contraction?

Let's digress, briefly.
If you are given a partially ordered set, what are the most important data you should know about that partially ordered set, from a combinatorial point of view?

I'll tell you, strictly confidentially, what it is.
Don't tell anyone.

Given a partially ordered set P, an interval (or segment) of P, say $[a, b] = \{ y \in P : a \leq y \leq b \}$,
the most important data to know are:

(1) the structure of every interval.

(2) how every interval factors uniquely into a product of partially ordered sets that are irreducible.

These are the data, in a great many situations, you need to deal with partially ordered sets.

Let's see what happens in the case of the lattice of partitions.
(1) What do the intervals look like?
(2) How do they factor?

Intervals in $\Pi[T]$

(1) $[\pi, \hat{1}] \leftarrow$ all partions above partition π and below $\hat{1}$. Remember $\hat{1}$ = blob

I claim that $[\pi, \hat{1}]$ is isomorphic to $\Pi[\pi]$.
π is a partition. It's a set — a set of blocks.
The blocks don't know where they are.
So, you can take partitions on the set of blocks.
So, I claim that the interval $[\pi, \hat{1}]$ is the same as $\Pi[\pi]$.

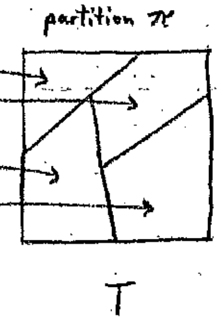
That's intuitively obvious.
Any partition above π shoves together some blocks of π.
So you might as well view the blocks of π as points. Nothing more is going to be done to them.

There's no point in giving a formal proof of this. It's so obvious.

(2) $[\hat{0}, \pi]$ ← all partitions above $\hat{0}$ and below partition π.

You have partition π cut up the set T.
Every partition is defined in terms of blocks.
Therefore, every other partition below π has to cut up some of the blocks of π.
And this cutting up is done independently of each block.

Therefore, to any partition in this interval,
there corresponds one partition of this block,
one partition of this block,
⋮
one partition of this block,
independently.



Therefore:

$$[\hat{0}, \pi] \text{ is } \underline{\text{isomorphic}} \text{ to } \bigotimes_{B \in \pi} \prod [B]$$

↑ product, where B ranges over the blocks of π

(3) arbitrary interval

$[\pi, \sigma]$, $\pi \leq \sigma$

This means that each block of σ is partioned by some block of π.
So you have the product of partition lattices giving you a partition lattice.

$$[\pi, \sigma] \text{ is } \underline{\text{isomorphic}} \text{ to } \bigotimes_{C \in \sigma} \prod \left[\{B : B \subseteq C,\ B \in \pi\}\right]$$

- We will see next time that to every graph, other than the complete graph, there corresponds a generalization of the lattice of partitions, which is obtained by taking the sup's of the edges of that graph only.
That's called the Lattice of Contractions of that graph. (see also [30.12])

And the coloring problem depends crucially on this lattice of contractions of a graph.
That's what it's all about.
Very complicated.

So, let's go back to lattice of contractions of a graph.
The lattice of contractions of a graph is this:

Take subset $A \subseteq S$ and the matroid (S, r)

Then, we want the contraction by A. Namely, the matroid:

$(S-A, r_A)$

↑ What is this?
That's easy. You take the interval from sup of A to $\hat{1}$.
You take the atoms of that and that's your contraction.
Since $[\vee A, \hat{1}]$ is isomorphic to a lattice of partitions, this will form a matroid.
It's a mental exercise to check that:

$[\vee A, \hat{1}]$ is isomorphic to the contraction $(S-A, r_A)$

So let's stop here.
I'm sorry we covered so little material today.
If you find a black notebook identical to this, anywhere, call me immediately.

Next time we'll prove Rado's Theorem.

- Rado's Theorem is an extremely powerful theorem.
More precisely, the powerful theorem is a combination of Rado's Theorem and the Normalization Theorem, which comes next, which I proved in 1966.
Combining the two, you get incredible strengthening of the Hall Marriage Theorem, as you will see.
In fact, any known matching theorem is a combination of these two (Rado's Theorem and the Normalization Theorem).

- The Theory of Matroids (cont'd)

We have seen that a matroid is a finite set S, together with a set function r, which we call a rank function and whose properties you know by now by heart.

$$(S, r)$$

We also stated, so far without proof, Whitney's Theorem, which gives an alternative characterization of a rank function.
One of these conditions is the Whitney Property:

$$\text{if } r(A \cup x) = r(A) \text{ and } r(A \cup y) = r(A)$$
$$\text{then } r(A \cup x \cup y) = r(A)$$

It is a technical feat to show that a function satisfying this and the other conditions implies that the function is submodular. Namely:

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

We've seen, also, some prime examples of matroids.
Matroids, in so far as they apply to subsets of a vector space, where the notions of independent sets and basis correspond exactly to the notions of independent vectors and basis of a subspace.

And matroids as applied to graphs. Graphs used as subsets of the set of atoms in the lattice of partitions. The atoms being viewed as the edges of the graph.
When we do this, then the rank function is:

$$r(x) = n - \text{number of blocks in partition}$$

A set of edges, or atoms, if you wish, is independent iff those edges, when drawn as a graph, form a tree.

In particular, a basis is a maximal spanning tree on the graph.
Our fundamental theorems on matroids immediatly imply some propeties on trees:
(1) Two maximal spanning trees of a graph have the same number of edges.
(2) Given on spanning tree with j elements and another with $j+1$, there is one element of the larger that can be adjoined to the smaller s.t. it is still a spanning tree.

Rado's Theorem

Now, let's see a generalization of matroids to the Marriage Theorem.

Let $A_1, A_2, \ldots, A_k$ = subsets of S, given the matroid (S, r).
We can find a set of independent representatives (i.e., a set $\{x_1, \ldots, x_k\}$, which is independent and $x_i \in A_i$) } x_i are distinct
iff for every subfamily $A_{i_1}, A_{i_2}, \ldots, A_{i_j}$ we have

$$r(A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_j}) \geq j$$

{ In the case where the rank function is the cardinality, we have the Hall's Marriage Theorem.
This is a matroid, of course.
Cardinality defines a trivial matroid, where every set is independent. }

Proof

We imitate the first proof we gave of Hall's Marriage Theorem, with suitable retouchings. [21.1–5]

Case 1: for every proper subfamily, we have $r(A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_j}) > j$ ← strictly

Pick $x_1 \in A_1$, necessarily independent, such that $r(x_1) = 1$
Such an x_1 exists, trivially, by the induction hypothesis on the properties of a rank function and the fact that $r(A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_j}) > j$.

Consider the contraction matroid on $(S - x_1, r_{x_1})$

I claim the Hall condition is still satisfied on this smaller matroid.

Let $B_i = A_i - x_1$ ← the i range over $i_1, i_2, \ldots, i_j$

Then, by the definition of the contraction by $\{x_1\}$ [26.11]:

Recall that: $r(x_1) = 1$

$$r_{x_1}(B_{i_1} \cup B_{i_2} \cup \ldots \cup B_{i_j}) = r(\underbrace{B_{i_1} \cup B_{i_2} \cup \ldots \cup B_{i_j} \cup x_1}) - \overbrace{r(x_1)}^{=1}$$

When you add x_1 back, you get the A_i back. You had subtracted only x_1 in each: $B_i = A_i - x_1$

$$= \underbrace{r(A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_j})}_{> j\text{, by assumption above}} - 1$$

$$\geq j$$

Continue, in a similar way, by induction on the smaller matroid $(S - x_1, r_{x_1})$

case 2 : There exists a proper subfamily, say, without loss of generality, $A_1, A_2, \ldots, A_j$ such that $r(A_1 \cup \cdots \cup A_j) = j$

In this case, we take the <u>restriction</u> to this.

Say $A_1 \cup A_2 \cup \cdots \cup A_j = Q$

The matroid (Q, r) satisfies the Hall condition.
It <u>doesn't know</u> that you're only subsets of Q.
The Hall condition is for all subsets of S.
So, by the Principle of Ignorance, if (S, r) satisfies the Hall condition, then so does (Q, r).

And S is <u>finite</u>.
Therefore, we can apply the induction hypothesis to (S, r).

We need to show, by the induction hypothesis, that (S, r) satisfies the Hall condition and, hence, we can find an independent set of representatives $\{x_1, \ldots, x_j\}$ of $A_1, \ldots, A_j$.

Consider the contraction $(S-Q, r_Q)$.

Now we have to do two things.
First we have to show that the contraction $(S-Q, r_Q)$ satisfies the Hall condition. So we get a set of independent representatives for this contraction.
Then we have to show that <u>this set</u> of independent representatives, <u>together</u> with the ones we have <u>already found</u>, together <u>jointly</u> gives us a set of independent representatives.

(1) Claim: $(S-Q, r_Q)$ satisfies the Hall condition for the sets $B_i = A_i \cap Q^c$, $i = j+1, j+2, \ldots, k$ $\longleftarrow$ $\left\{\begin{array}{l}\text{leave out the} \\ \text{part that is} \\ \text{already } Q\end{array}\right\}$

Let's see how this satisfies the Hall condition. Write out r_Q:

$$r_Q(B_{i_1} \cup \cdots \cup B_{i_\ell}) = r(\underbrace{B_{i_1} \cup \cdots \cup B_{i_\ell}} \cup Q) - \underbrace{r(Q)}$$

when you add Q back, the B_i become A_i.

$r(A_1 \cup \cdots \cup A_j) = j$ given assumption.

$$= r(\underbrace{A_{i_1} \cup \cdots \cup A_{i_\ell}}_{\ell} \cup \underbrace{A_1 \cup \cdots \cup A_j}_{j}) - j$$

$$\geq \ell + j - j$$

$$= \ell$$

So we win. This contraction $(S-Q, r_Q)$ satisfies the Hall condition.

Hence, we can find $\{x_{j+1}, \ldots, x_k\}$ = set of independent representatives of $B_{j+1}, \ldots, B_k$ in $(S-Q, r_Q)$.

(2) Now we need to show that these, together with the first j we have shown, are independent.
What does it mean for $\{x_{j+1}, \ldots, x_k\}$ to be independent?
It means that:

$$r_Q(x_{j+1} \cup \ldots \cup x_k) = k-j$$

That's what being independent means.
The rank of the set is equal to the size of the set.

This means that:

$$r(x_{j+1} \cup \ldots \cup x_k \cup Q) - \underbrace{\cancel{r(Q)}}_{} = k - \cancel{j}$$

$Q = A_1 \cup \ldots \cup A_j$
$r(A_1 \cup \ldots \cup A_j) = j$, by assumption
$r(Q) = j$

$$r(x_{j+1} \cup \ldots \cup x_k \cup Q) \qquad = k$$

Note that $\{x_1, \ldots, x_j\} \subseteq Q$

It is intuitively obvious that if you have a subset that is independent, then it must have the same rank.
But let's show that:

$$r(x_1 \cup \ldots \cup x_k) = k$$

If I add any element $q \in Q$ to $\{x_1 \cup \ldots \cup x_k\}$, I will show that the rank does not change. Namely:

$$r(x_1 \cup x_2 \cup \ldots \cup x_k \cup q),\ q \in Q \quad = r(x_1 \cup \ldots \cup x_k)$$

Intuitively, that's obvious, as $\{x_1, \ldots, x_k\}$ is a basis of $(S-Q, r_Q)$.
Formally, we take the submodularity property of a rank function:

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

Let $A = \{x_1 \cup \ldots \cup x_j \cup q\}$

$B = \{x_1 \cup \ldots \cup x_K\}$

Then:

$$\underbrace{r(x_1 \cup \ldots \cup x_K \cup q)}^{A\cup B} + \cancel{\overbrace{r(x_1 \cup \ldots \cup x_j)}^{A \cap B}} \le \underbrace{\cancel{\overbrace{r(x_1 \cup \ldots \cup x_j \cup q)}^{A}}}_{\text{this rank is the same as } r(x_1 \cup \ldots \cup x_j)\text{, because it is a basis of } Q.} + \overbrace{r(x_1 \cup \ldots \cup x_K)}^{B}$$

$$r(x_1 \cup \ldots \cup x_K \cup q) \le r(x_1 \cup \ldots \cup x_K), \text{ for all } q \in Q$$

from the increasing property of a rank function: $A \subseteq B \Rightarrow r(A) \le r(B)$
this must be equality.

$$r(x_1 \cup \ldots \cup x_K \cup q) = r(x_1 \cup \ldots \cup x_K), \text{ for all } q \in Q$$

And, by the Extended Whitney Property:

$$\underbrace{r(x_1 \cup \ldots \cup x_K \cup Q)}_{\text{we've shown that this rank equals } K.} = r(x_1 \cup \ldots \cup x_K)$$

Therefore:

$$r(x_1 \cup \ldots \cup x_K) = K$$

There are K elements and the rank is K.
Therefore $\{x_1 \cup \ldots \cup x_K\}$ is independent.

Done.

End of the proof.

I already outlined some of the applications.
Vector spaces – if you have any subset of a vector space, you can pick elements of the subsets as having independent sets. You can pick it so that the dimension of any union of j subsets is at least j.

Or you can apply it to trees. You have a family of edges of a graph. And you want to pick one edge from each family of edges, so that you get a tree.
When can you do that? When the rank of the union of edge families is at least the number of families.

- There is better to come.
  Now we build up a new class of matroids and apply Rado's Theorem and get terrific matching theorems.
  Now is the payoff.

- Normalization Theorem

  Given a set function μ on the finite set S, integer valued, with the properties:

  (1) increasing
  $$A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$$

  (2) submodular
  $$\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B)$$

  (3) $\mu(\emptyset) = 0$ ← (we actually may not need this explicitly, but we'll include it for safety.)

  Unfortunately, μ so defined does not define a matroid.
  Why?
  Because μ of a point is NOT 0 or 1.
  On the other hand, it's kind of easy to find these functions μ.

- For example, take the relation R and the set function μ:

  $$R \subseteq S \times T$$
  $$\mu(A) = |R(A)|$$

  We've verified to our hearts content that this is submodular.
  And it's obviously increasing.
  But it doesn't define a matroid $\mu(\text{point}) = |R(\text{point})| = 10$, for example.

- If we now define:

  $$r(A) = \min_{B \subseteq A} \left(\mu(B) + |A - B|\right)$$

  then we obtain a rank function. And we obtain a matroid.

I always forget the proof of this theorem, because I was the one who proved it.
I blank out the effort.
In 1966, before you were born.

Let's see how you prove it. I forgot.
There are many ways to prove it. Almost every way works.

We'll prove it by showing that this r, so defined, satisfies the conditions of Whitney's Theorem. [25.4]
That seems to be the simplest way.

I know you are wondering: "Where does this come from? Where did you get this?"
I sadistically withhold the answer to that question.
First I make you suffer. Then I tell you what's really going on.

Proof
(1) $r(\emptyset) = \min_{B\subseteq\emptyset}(\overset{0}{\cancel{\mu(B)}} + \overset{0}{\cancel{|\emptyset - B|}}) = 0$ ✓
(2) $r(A \cup x) = r(A) + \begin{cases}0\\1\end{cases}$

This is the crucial property, because μ does not satisfy this property.
Let's write out $r(A\cup x)$:

$$r(A\cup x) = \min_{B\subseteq A\cup x}\left(\mu(B) + |A\cup x - B|\right)$$

{There are two kinds of B's contained in $A\cup x$.
There are B's contained in A and B's contained in $A\cup x$.}

$$= \min_{B\subseteq A}\left(\mu(B) + \underbrace{|A\cup x - B|}_{①}, \mu(B\cup x) + \underbrace{|A\cup x - B\cup x|}_{|A-B|}\right)$$

① This is at most 1 greater than the preceeding, because you add x here.

② Recall that: $\min(X,Y) \le \min(X)$

$$\le \min_{B\subseteq A}\left(\mu(B) + |A-B|\right) + 1$$

$$= r(A) + 1$$

And since μ is integer valued, we have:

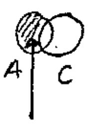
$$r(A\cup x) = r(A) + \begin{cases}0\\1\end{cases} \checkmark$$

(3) <u>increasing property</u>

$$r(A) = \min_{B \subseteq A} \left(\mu(B) + |A-B| \right)$$

This means that the minimum is attained at some subset $C \subseteq A$, because the sets are finite.
Let this minimum be attained for some set C.

$$= \mu(C) + \underbrace{|A-C|}, \text{ for some } C \subseteq A$$



$$A - C = A \cap C^c$$

$$= \mu(C) + |A \cap C^c|$$

① from the increasing property of set function μ, we have:

$$C \supseteq B \cap C \Rightarrow \mu(C) \geq \mu(B \cap C)$$

② $A \supseteq B \Rightarrow |A \cap C^c| \geq |B \cap C^c|$

Combining these gives the following inequality:

$$\geq \mu(B \cap C) + \underbrace{|B \cap C^c|}$$

$$B \cap C^c = \overbrace{(B \cap B^c)}^{\emptyset} \cup (B \cap C^c)$$

$$= B \cap (B^c \cup C^c) \quad \text{by the distributive law}$$

$$= B \cap (B \cap C)^c$$

$$= \mu(\underbrace{B \cap C}) + |B \cap \underbrace{(B \cap C)}^c|$$

Then we compare this instance to the minimum:

$$\geq \min_{D \subseteq B} \left(\mu(D) + |B - D| \right)$$

$$= r(D)$$

$$r(A) \geq r(D)$$

Therefore, we have shown that:

if $\underbrace{A \supseteq D}$ then $r(A) \geq r(D)$ ✓

$A \supseteq B$ and $B \supseteq D \Rightarrow A \supseteq D$

(4) <u>the Whitney Property</u>

μ, as we defined it [27.6], satisfies the Whitney Property.
You may recall that any increasing, submodular set function μ, with the property that $\mu(\emptyset)=0$, satisfies the Whitney Property. That's the case with our μ!
Recall our proof of the Whitney Property. [24.6, Theorem 1]

Suppose that:

$$\left.\begin{aligned} r(A\cup x) &= r(A) \\ r(A\cup y) &= r(A) \end{aligned}\right\} \quad \text{We want to show that it then follows that:} \quad r(A\cup x\cup y) = r(A)$$

Write out $r(A)$:

$$(*) \quad r(A) = \min_{B\subseteq A}\left(\mu(B) + |A-B|\right)$$

$$= r(A\cup x) \quad \longleftarrow r(A\cup x) = r(A) \text{ is given assumption.}$$

$$= \min_{B\subseteq A\cup x}\left(\mu(B) + |A\cup x - B|\right)$$

↑ again, these B's can be of two kind.

$$= \min_{B\subseteq A}\Big(\underbrace{\mu(B)+|A\cup x - B|}_{\text{if } B \text{ contains } x, \text{ you get this.}}, \underbrace{\mu(B\cup x) + \overbrace{|A\cup x - B\cup x|}^{A\cup x - B\cup x \,=\, A-B}}_{\text{if } B \text{ does not contain } x, \text{ you get this.}}\Big)$$

⇑

Note that this <u>first term</u> is exactly 1 greater than $r(A)$:

$$\min_{B\subseteq A}\left(\mu(B) + |A\cup x - B|\right) = \min_{B\subseteq A}\left(\mu(B) + |A-B|\right) + 1$$

$$= r(A) + 1$$

$$\cancel{r(A\cup x) = r(A) + 1}$$

↑ This violates the given assumption that $r(A\cup x) = r(A)$.
So the minimum can <u>not</u> be attained by the first term.

$$= \min_{B\subseteq A}\left(\mu(B\cup x) + |A-B|\right)$$

Therefore, in conjunction with equation (*), we must have:

$$\min_{B\subseteq A}\left(\mu(B) + |A-B|\right) = \min_{B\subseteq A}\left(\mu(B\cup x) + |A-B|\right)$$

Let's say the minimum of equation (*) is attained at C:

$$\mu(C) + \cancel{|A-C|} = \min_{B \subseteq A} \left(\mu(B \cup x) + |A-B| \right)$$
$$= \mu(C \cup x) + \cancel{|A-C|}$$

This implies that:

$$\left.\begin{array}{l} \mu(C \cup x) = \mu(C) \\ \text{Similarily, we can show:} \\ \mu(C \cup y) = \mu(C) \end{array}\right\}$$

μ satisfies the Whitney Property, so this implies that:

$$\mu(C \cup x \cup y) = \mu(C)$$

which, in turn, translates to:

$$r(A \cup x \cup y) = r(A)$$

Therefore, we have:

$$\left.\begin{array}{l} r(A \cup x) = r(A) \\ r(A \cup y) = r(A) \end{array}\right\} \Longrightarrow r(A \cup x \cup y) = r(A) \quad \checkmark$$

So r satisfies the properties of the Whitney Theorem. It is a rank function. We have a <u>matroid</u>.

- In this way, we can create matroids out of nothing. You will see next time.
- When you apply this <u>Normalization Theorem</u> with <u>Rado's Theorem</u>, you get the most marvelous matching theorems – originally proved by crazy methods. These 2 theorems give a total unifying matching theory, from which all known matching theorems come out.

The wonderful world of matroids. Thrilling. Full of surprises.
Actually, it is.
It's a pity there is no time to tell you about the recent developments of matroids.
People have discovered some marvellous connections among matroids, representation theory, geometric probability – all sorts of things.

If you really want to know the algebra behind matroids, you'll have to take my course next term, on multilinear algebra. Then you'll learn matroids.

You'll note that in this course, I stay away from algebraic topics.
This is a pure combinatorics course. On purpose, because the algebra is left for next time.
Sometimes one tends to throw in some algebra, but I resist the temptation.
You get just combinatorics – pure and simple.

Matroids and Matching

Last time, we saw two important theorems on matroids.
Namely, Rado's Theorem and the Normalization Theorem.

Rado's Theorem

Let me state this in a succinct way, because I've already stated it 5 times:

Given matroid (S, r) and a family of subsets $A_1, \ldots, A_K \subseteq S$
we want to find a system of independent representatives $\{x_1, \ldots, x_K\}$ ← set. NOT a multiset. No duplicates.
s.t.

$$x_i \in A_i \quad \text{and} \quad \underbrace{r(\{x_1, \ldots, x_K\}) = k}$$

i.e., the set is independent.
rank equals size of set.

Rado tells us this is possible iff:

for every subfamily $A_{i_1}, \ldots, A_{i_j}$ we have $r(A_{i_1} \cup \ldots \cup A_{i_j}) \geq j$

We saw that the proof is remarkably similar to the original proof we gave of Philip Hall's Marriage Theorem.
If you do restrictions and contractions the right way, then out comes the proof.

This is, in a sense, the ultimate matching theorem – as you will see shortly.
No one has really gone beyond this.
There is sort of a gut feeling that all the known matching theorems fall out by specializing this theorem.

Normalization Theorem

Given a set function μ on S, integer valued, increasing, submodular, and $\mu(\emptyset) = 0$
then
$$r(A) = \min_{B \subseteq A} \left(\mu(B) + |A - B| \right)$$ is a rank function of a matroid.

We verified this last time.

Now, let's squeeze all the juice from these 2 theorems.
Let's start with easy stuff.

Applications

Take $R \subseteq S \times T$
Set $\mu(A) = |R(A)|$ for $A \subseteq S$

We have verified that this is submodular, that it is increasing, and that $\mu(\emptyset) = 0$.
Therefore, the Normalization Theorem tells us that to every relation we can associate a matroid.

What does the matroid look like?

Let r be the rank function associated to this μ, by the Normalization Theorem.

Every relation defines a matroid.
How do we understand this matroid?
Well - get ahold of one of the matroidal concepts and see how it is interpreted.
In this case, let's see what the independent sets look like.
And here we find a pleasant surprise.

What are the independent sets?

By definition:

$I \subseteq S$ is independent iff $\underbrace{\min_{B \subseteq I} \left(\mu(B) + |I - B| \right) = |I|}_{r(I) = |I|}$

By definition of min, this means that:

For every $B \subseteq I$,

$$\mu(B) + |I - B| \geq |I|$$

Remember that $\mu(B) = |R(B)|$, which gives:

$$|R(B)| \geq |B|$$ ← That's the condition of the Marriage Theorem!

Therefore, we have that <u>in the matroid</u> associated with this relation R, a set is <u>independent</u> <u>iff</u> it has a <u>matching</u>.
This is very nice.

Thus, I is independent iff there is a partial matching defined on I.

Now you see what it's all about.
The independent sets are those such that if you restrict the relations to those sets, that there is a partial matching.
And that's what relations are about.

Now you can apply the <u>abstract theory of independent sets</u> to matchings.
And get all sorts of theorems that I didn't state before, because it would have been superfluous.
For example:

- All maximal matchings have the same size. (the basis)
  People used to elaborately prove this before.

- If you have 2 partial matchings, one bigger than the other, then you can take one edge from the larger one and add it to the smaller to get a bigger matching.

And so on and so forth.
That's not the end of the story.
Let's jazz this up.

- We could have a matroid on T already. ← just a set
  A <u>pre-given matroid</u>.
  And we go through this process, but instead of absolute value, we use the rank for μ.
  <u>It still works</u>, because rank is increasing, submodular, etc.
  It satisfies the conditions for the Normalization Theorem to produce <u>another</u> rank function.

More generally:

r' is a pre-given rank function

Given a matroid (T, r'), set $\underbrace{\mu(A) = r'(R(A))}$.

μ is integer valued, increasing, submodular, and $\mu(\emptyset) = 0$.
So we can apply the Normalization Theorem and we get another matroid.

Apply the Normalization Theorem:

$$r(A) = \min_{B \subseteq A} \left(\mu(B) + |A - B| \right)$$

$$\mu(B) = r'(R(B))$$

We get the <u>induced matroid</u> by the relation R.

And the same computation we have just gone through tells you that <u>independent sets</u> of the <u>induced matroid</u> are the sets that have <u>partial sets</u> of <u>independent representatives.</u>
That's it — cheapo.

You can get fantastic theorems.
You can take a relation of a relation.
Mix them up. You can do all sorts of things.
Let's do some more of this.

After the Marriage Theorem, people started to prove generalizations.
So let me state a couple of generalizations that people proved.
Then we see that they are nothing, if you look at them from the point of view of matroids.

<u>Theorem</u> - Partial Matching (excluding k elements)

Given $R \subseteq S \times T$

Philip Hall tells you when there is a matching.
Now we want to know if there is a <u>partial matching</u> containing all but k elements in the matching.
Is there a necessary and sufficient condition for such a matching?

Given $R \subseteq S \times T$, there is a <u>partial matching</u> of R containing $|S| - k$ elements iff
for every $A \subseteq S$, we have $|R(A)| \geq |A| - k$

We have 2 choices.
Either we go through another elaborate proof, à la Philip Hall.
Or else we get it out of the Normalization/Rado Theorems, by making up a matroid.

So here's how you do work it. You must learn the tricks of the trade.
I'm not going to prove this.
I refuse.
So we do it this way instead:

$$\text{Set } \mu(A) = |R(A)| + k$$

This is an increasing, submodular set function.
We can apply the Normalization Theorem to it.

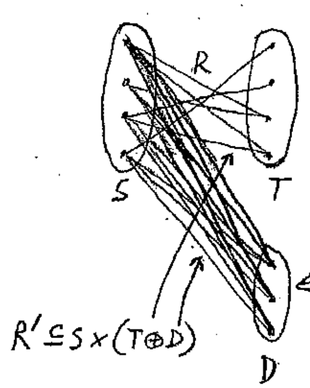
Apply the normalization theorem, then check that the independent elements are the partial matching:

$$r(A) = \min_{B \subseteq A} \left(\mu(B) + |A - B| \right)$$

$$\mu(B) = |R(B)| + k$$

But this still doesn't explain, in full clarity, why this is an independent set or what they are.
Let me tell you yet <u>another trick.</u>

After applying the Normalization Theorem, you get a matroid.
And anyone can see what this matroid looks like.



① Take the relation $R \subseteq S \times T$

② take disjoint set D, $|D| = k$
of k elements

disjoint sum

③ Define a <u>new relation</u> $R' \subseteq S \times (T \oplus D)$
Every element of S is joined to every element of D.

What does it mean for R' to satisfy the conditions of the Marriage Theorem?

$$|R'(A)| \geq |A| \quad \text{iff} \quad |R(A)| \geq |A| - k$$

Because $R'(A)$ has <u>k more</u> matchings than $R(A)$, for any A.
Therefore $R'(A)$ satisfies the Philip Hall condition iff $R(A)$ satisfies the condition with k fewer.

So the <u>partial matching theorem</u> is easy to prove.
It's an immediate consequence of Philip Hall and this construction.
So I did give you a proof, after all.

And this tells you what the matroid defined by this increasing, submodular function $\mu(A) = |R(A)| + k$, after applying the Normalization Theorem, looks like.

It means you are <u>faking</u> the extra elements.
And there's a whole <u>theory</u> that tells you that <u>every submodular set function</u> corresponds to some sort of <u>faking</u> of <u>elements.</u>

Finally, let's look at the <u>independent set</u> of this matroid:

$$r(I) = |I| \Rightarrow \min_{A \subseteq I} \big(\underbrace{\mu(A)}_{\mu(A) = |R(A)| + k} + |I - A| \big) = |I|$$

$$\text{Therefore:} \quad |R(A)| \geq |A| - k \longleftarrow \left\{ \begin{array}{l} \text{See above.} \\ \text{the condition for} \\ \text{a partial matching} \\ \text{(excluding k elements)} \end{array} \right\}$$

So the <u>independent set</u> of this matroid does, in fact, give the partial matching.

Next example,
Here's a theorem that someone, somewhere, proved.
We say - "ok, A relation has a matching iff it satisfies this Hall condition."
What's the <u>next best thing</u> after <u>matching</u>?

The next best thing is this,
You have a partition of S into 2 blocks, such that each block has its own matching.
Or, more generally, the partition of S into k blocks, such that the restriction to each block is a matching.

Let's see if we can get a necessary and sufficient condition for the case with 2 blocks.

<u>Theorem</u>

A necessary and sufficient condition that, given $R \subseteq S \times T$, there exists a partition $\pi = (B_1, B_2)$ such that $R|_{B_i}$ has a matching is that:

$$2|R(A)| \geq |A| \quad , \text{ for all } A \subseteq S$$

$(R|_{B_i} = R \text{ restricted to } B_i)$

Kind of cute. If you have k blocks, then the condition is $k|R(A)| \geq |A|$.
First I'll give you the matroid interpretation.
Then I'll give you the visual interpretation.

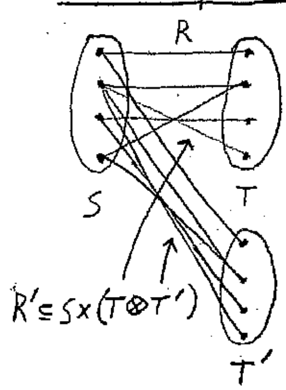
<u>matroid interpretation</u>

If you have a μ that satisfies the hypothesis of the Normalization Theorem, then 2μ does too.
And, therefore, applying the Normalization Theorem with 2μ gives you another matroid.

Apply the Normalization Theorem to $\mu(A) = 2|R(A)|$ and you get a matroid.

What does this matroid look like? Easy.

<u>visual interpretation</u>



① I take T' = a copy of set T and same identical relation R is defined on T'.

② Then create the relation:

$$R' \subseteq S \times (T \otimes T')$$

where:

$$R'|_{S\times T} = R$$

$$R'|_{S\times T'} \simeq R \leftarrow \text{isomorphic}$$

Now apply the Marriage Theorem to R'.

You get the matching immediately.
The matching will be partly between S and T, partly between S and T'.
T' is "virtually" T.
You get the theorem immediately.

- We can jazz up the last 2 theorems.
Instead of taking a relation between S and T, we can put a matroid structure on T in both of these last 2 theorems.
And you immediately get a generalization.

In fact, here is the generalization:

- Generalization

Suppose you have 2 matroids on the same set.

Given (S, r_1) and (S, r_2).

We mix up these two matroids.
How do we unscramble them? Very easy.

Take $\mu = r_1 + r_2$ and apply the Normalization Theorem.

What do they look like?

The independent sets are unions of the r_1 independent sets and the r_2 independent sets.
The same reasoning, as in the previous results, applies here.

Given the matroid obtained by normalizing with $\mu = r_1 + r_2$, the independent sets of this matroid are sets $I = I_1 \cup I_2$, where I_i is r_i independent.

- Q: Do we know how many different partitions π of a given block size there are that satisfy the necessary and sufficient conditions such that each block has a matching?

A: No, How many there are — people have no idea.
That's a dead end.
Counting these matchings is absolutely a dead end.
The theorem [28.7] states only the existence of matchings.

Exercise 28.1

There's another theorem I want to do, but I'll give it as an excercise, because I hope you are catching on to this game.

Remember we talked about the Gale-Ryser Theorem. [13.2, Exercise 13.2] Now I give it to you as an exercise, because it's the Marriage Theorem jazzed up. Before, I wanted you to do this by rolling up your sleeves.

Professor David Gale was Professor of Economics at UC Berkeley.
Professor Herbert Ryser was Professor of Mathematics at Cal Tech.

Gale-Ryser Theorem

You have the incidence matrix of a relation.
So, you have a matrix of 0's and 1's.

$$\begin{array}{c} \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{array} \begin{array}{c} \begin{array}{cccc} \mu_1 & \mu_2 & \cdots & \mu_n \end{array} \\ \left[\begin{array}{cccc} \; & \; & \; & \; \\ \; & & & \\ \; & & & \\ \; & & & \end{array} \right] \end{array}$$

What you are given are the marginals.
We'll assume, wlog, that:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$
$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$$

The λ_i and μ_i are both partitions of the number, because the number of 1's is the same. So, trivially:

$$\sum \lambda_i = \sum \mu_i$$

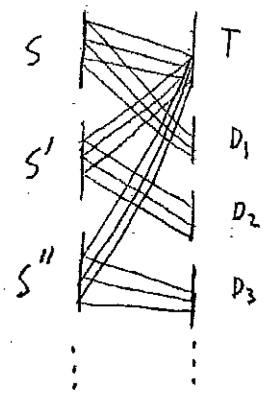
Q: Given the marginals, when does there exist a matrix of 0's and 1's with these marginals?

A: Iff $\lambda \leq \mu^{\perp}$ ← dual partition in the dominance order.

This is just Philip Hall. I'll tell you the trick.
You work it out. I don't want to do it today.
The trick is this:

You take a relation $R \subseteq S \times T$.
And then you sets $D_1, D_2, \ldots, D_K$ (some number of sets) with certain elements that are determined by the μ.
And then you repeat S. You take $S', S'', \ldots, S^{(k)}$ and define, to each, the same relation as $R \subseteq S \times T$.

You have this, roughly speaking:



S is related to elements of T and <u>every</u> element in D_1.

S' is related to elements of T just like in R and <u>every</u> element in D_2.

S'' is related to elements of T just like in R and <u>every</u> element in D_3.

And so on.

You have to set this right, depending on the λ and the μ.
And then apply the Marriage Theorem.
And out comes the Gale-Ryser Theorem.

I don't want to do the gory details.
I leave this as an exercise.

<u>** Exercise 28.2</u>

Nobody has ever looked at what you would get if you have a matroid structure on T.
What kind of generalization of the Gale-Ryser Theorem do you get if T has a matroid structure?

Generalize Gale-Ryser to induced matroids.

<u>** Exercise 28.3</u>

You remember that I gave you this problem: [10.4-5, Exercise 10.2]

If you have 2 matrices of 0's and 1's with the same marginals, then you can get from one to the other by a series of <u>switches.</u>

$$\begin{array}{c c} & \begin{array}{cc} b & d \end{array} \\ \begin{array}{c} a \\ c \end{array} & \begin{pmatrix} \cdots & 1 & 0 & \cdots \\ & 0 & 1 & \end{pmatrix} \end{array} \longrightarrow \begin{array}{c c} & \begin{array}{cc} b & d \end{array} \\ \begin{array}{c} a \\ c \end{array} & \begin{pmatrix} \cdots & 0 & 1 & \cdots \\ & 1 & 0 & \end{pmatrix} \end{array}$$

M $\qquad\qquad\qquad$ M' switch

Get this as a matroid theorem.
Get this as a consequence of the general theorems of matroids.

I think we've done enough matching theory.
You've had your fill.
We still have to do Whitney's Theorem and prove it.
Then there's one last topic we have to do before we leave matroids.
Namely, the definition of the lattices associated with matroids, which I call geometric lattices.

Next time, we will apply the preceeding theory to develop all the main properties of geometric lattices.

I may drag on with matroids. I want to motivate Möbius functions with geometric probability. It's kind of a tour de force.
So I may continue with matroid theory and do a little more with arrangements of hyperplanes and all that stuff.

Geometric Lattice

What's a geometric lattice?

A geomtric lattice is probably the most interesting kind of lattice, after the following.
First, there are distributive lattices, which are very well understood.
Then there are linear lattices, lattices of commuting equivalence relations that satisfy the modular law. Namely, they have a rank function and they're finite and they satisfy the modular law.

NOT submodular, but modular. Equal.

After linear lattices is the next most important class of lattices is the class of geometric lattices.

The gruesome definition is the following: (without motivation. Next time we'll discuss this.)

L = finite lattice,

It has a rank function. Namely, all maximal chains have the same number of elements. So you can count how far away you are from $\hat{0}$.
Note the dual use of the term rank function. This is deliberate.

with a rank function r s.t.

$r(x \vee y) + r(x \wedge y) \leq r(x) + r(y)$ and

(here, submodular means something different than what we've discussed lately.)

for every element $x \in L$, there exists a set of atoms A s.t. $\vee A = x$.
(i.e., every element is the sup of atoms.)

Such a lattice is called a geometric lattice.

We will see that <u>geometric lattices</u> are the <u>same</u> thing as <u>matroids</u>, in disguise.
Yet another disguise of matroids.

Matroids have infinitely many disguises.
No other concept in mathematics I know of has as many cryptomorphic definitions as the concept of matroids.

People try and invent something new and, lo and behold, they prove it's a matroid.
It's a very rigid concept. Very hard to get away from,
which is a good sign.

So, next time, we'll connect <u>matroids</u> to <u>geometric lattices.</u>
And it is through this connection to <u>geometric lattices</u> that you get to <u>coloring</u>.

We have to do one more concept on matroids, orthogonality.
Then we will do the concept of closure.
And then geometric lattices.

Given a matroid on the set S with rank function r:

$$(S, r)$$

the rank function is something very similar to a dimension, as we will see. We have the notions of independence, basis, we have the exchange property for independent sets.

One might think that matroids are similar, in abstraction, to vector spaces. There is, however, one concept of matroids that you would never guess by doing linear algebra.
There are, actually, several. But our time is short, so there is only one that we will do here.
That's the concept of orthogonality.

Orthogonality

Define a set function r^* as follows:

$$r^*(A) = |A| + r(S-A) - r(S), \quad A \subseteq S$$

Theorem

r^* is a rank function and (S, r^*) is called the orthogonal (sometimes dual) matroid to (S, r).

In other words, if you're given a matroid, there is this funny formula that gives another matroid.
First, let's check that r^* is, indeed, a rank function.
Then let's try to understand what it means.

Proof — proof r^* is a rank function using Whitney's Theorem [26.4]:

(1) $r^*(\emptyset) = |\emptyset| + \cancel{r(S-\emptyset)} - \cancel{r(S)}$

$= 0$ ✓

(2) $r^*(A \cup x) = \underbrace{|A \cup x|}_{\text{goes up by 1 from } |A|} + \underbrace{r(S-A-x)}_{\text{stays the same, or goes down by 1, because you've removed an element, from } r(S-A).} - r(S)$

$= r^*(A) + \begin{cases} 0 \\ 1 \end{cases}$ ✓

(3) This implies, immediately, that:

$$A \subseteq B \Rightarrow r^*(A) \leq r^*(B) \quad \checkmark$$

Because you start with A and add x's until you get B.
And each time you add an x, r^* either stays the same or goes up by 1.

(4) Lastly, we want to show that r^* is <u>submodular</u>,
Namely, that:

$$r^*(A \cup B) + r^*(A \cap B) \leq r^*(A) + r^*(B)$$

We've stated many times, so far without proof, that a set function satisfying the <u>Whitney Property</u>, along with properties 1-3 above, <u>implies</u> <u>submodularity</u>.

Let's prove the Whitney Property for r^*:

Suppose $r^*(A \cup x) = r^*(A)$ and $r^*(A \cup y) = r^*(A)$

Need to show that this implies: $r^*(A \cup x \cup y) = r^*(A)$

Thus:

$$\overbrace{\underbrace{|A \cup x|}_{|A \cup x| = |A|+1} + r(S-A-x) - \cancel{r(S)}}^{r^*(A \cup x)} = \overbrace{|A| + r(S-A) - \cancel{r(S)}}^{r^*(A)}$$

This implies that:

$$r(S-A-x) = r(S-A) - 1$$

Similarly:

$$r(S-A-y) = r(S-A) - 1$$

(My professor of logic, Professor Church, when he said <u>similarly</u>, he would repeat the whole argument with y. Because it was not logical to say similarly.)

To get the desired conclusion that $r^*(A \cup x \cup y) = r^*(A)$, we have:

$$|A \cup x \cup y| + \underline{r(S-A-x-y)} - \cancel{r(S)} = |A| + \underline{r(S-A)} - \cancel{r(S)}$$

We want: $r(S-A-x-y) = r(S-A) - 2$

↑ The only way known to man to get this equality is to use the submodularity of r.

Let $A' = S-A-x$
$B' = S-A-y$

Then $A' \cap B' = S-A-x-y$
$A' \cup B' = S-A$

Now, let's apply the submodular inequality of r, using A' and B':

$$r(A' \cup B') + r(A' \cap B') \leq r(A') + r(B')$$

$$\cancel{r(S-A)} + r(S-A-x-y) \leq \underbrace{r(S-A-x)}_{\substack{\text{As we've just shown:} \\ r(S-A-x) = r(S-A)-1}} + \underbrace{r(S-A-y)}_{\substack{\text{similarly:} \\ r(S-A-y) = r(S-A)-1}}$$

$$= \cancel{2}\, r(S-A) - 2$$

$$\underbrace{r(S-A-x-y)}_{} \leq r(S-A) - 2$$

(from the properties of rank function r:

$$r(S-A-x-y) = r(S-A) - \begin{cases} 0 \\ 1 \\ 2 \end{cases}$$

Therefore, the only way this inequality can be satisfied is if there is equality:)

$$r(S-A-x-y) = r(S-A) - 2$$

Thus, the desired conclusion $r^*(A \cup x \cup y) = r^*(A)$ of the Whitney Property holds. ✓

The proof is complete.
r^* is a rank function.

So we have this weirdissimo matroid.
Linear algebra would never give you this.
That's not the dual of a vector space.

So it's my duty to tell you where it comes from.

Exercise 29.1

Remember the matroid of a graph.
What's a graph, from the point of view of matroids?
A graph is a set of edges.
An edge is an atom in the lattice of partitions.
Then you take the restriction of the matroid defined on the atoms, which we discussed.

Let's take the matroid of a planar graph.

This is Kultur.
There is a theorem, know as Fary's Theorem, that states:

If a graph can be drawn in the plane by any curves whatsoever,
then it can also be drawn with straight line segments.

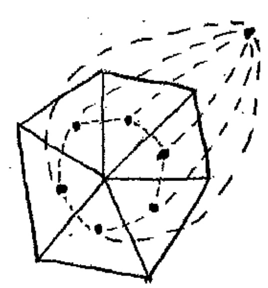
Proof – stretch it. That's the proof, basically.

So a planar graph can always be viewed as consisting of straight line segments.

We said that a set of edges in a matroid is independent if it's a tree.
It's dependent if there is a circuit.
Remember, we discussed this. [26.7 - 9]

What's the orthogonal matroid of a planar graphic matroid?

Exercise – take the dual graph in the classical sense of graph theory.



Put a point in the center of each region, including the outermost region.
Then join two points if their regions are adjacent.

Theorem – The matroid of this graph is the orthogonal matroid.
That's an exercise for you to work on.

Theorem

The orthogonal matroid of a planar graphic matroid is the matroid of the dual graph.

↑ isomorphic to, of course

Prove this.
It doesn't look obvious, but it's kind of easy when you look at it.

What do independent sets look like in the orthogonal matroid?

There is the following theorem.
The basis of the orthogonal matroid is the complement of a basis of the given matroid.

If you take a graph, what's a basis? A maximal spanning tree.
If you take all the edges not in that spanning tree, that's a basis of the orthogonal matroid.
In fact, it's the basis of the dual graph, if you start looking at it and fool around.

Let's set a date.
I'll take you out for Combinatorial Brunch.
There's only one date – Sunday, December 6.
We assemble in my apartment at 1105 Massachusetts Avenue, Apartment 8F, at exactly 11:30.
From there, we walk to the Charles Hotel and we have brunch in the Charles Hotel.

Q: In the morning?
A: Morning? ☺
I plan on getting up early.
I used to get up that late, myself, when I was your age.

It's one of those all you can eat things.
Come hungry.
After all you've heard of combinatorics, you deserve a brunch.

Theorem

If B is a basis of (S, r) then $B^c = S - B$ is a basis of (S, r^*).

Proof

Q: What does it mean to be a basis of (S, r)?
A: When B is a maximal independent set, with $r(B) = |B|$. [24.8, 24.10]

So, let's plod ahead with our definition of r^*:

$$r^*(B^c) = |B^c| + r(\underbrace{S - B^c}_{B^{cc} = B}) - r(S)$$

$$= |B^c| + \cancel{r(B)} \quad \cancel{-r(S)}$$

How do you have B maximal? What makes it a basis?
When:

$$|B| = r(B) = r(S)$$

The maximal independent set has maximal rank. [24.11]

$$r^*(B^c) = |B^c|$$

equality implies that B^c is an independent set of (S, r^*).
One can argue that this independent set is maximal. It's not difficult.
And we are done,
B^c is a basis of (S, r^*).

- Now, don't think this is weird.
Because electrical engineers built all circuit theory from these facts,
Maxwell's equations, Kickhoff's laws are inside this stuff.
I don't have time to go into it, obviously.
But all of circuit theory comes out of this. That's what it's all about.

- Lastly, let me mention the original motivation by Whitney in inventing matroids.
He noticed that if you have a graph that is not planar, there is no dual graph.
However, every graph defines a matroid. This matroid has an orthogonal matroid.
It's not a graph, but it's a matroid,
And it plays the role of the dual graph.

That's how matroids started.

Needless to say, we have considered just the very beginning of the theory of matroids. If you want to learn more, you can read my old booklet with Crapo called "Combinatorial Geometries," which was rewritten in 4 volumes by one of my former students, Neil White.

Volume 1 – Theory of Matroids,
2 – Combinatorial Geometries
3 – Matroid Applications
4 – Oriented Matroids

Our original book, by Crapo and Rota, was called Preliminary Edition, 1970.
The real edition never appeared.
So you have to look at these 4 volumes.
It's a very deep theory that is going on.
I will mention, later on, some of deepest theorems that have been proved recently in matroid theory.
Some of the deepest theorems in combinatorics have to do with matroids.

Now we want to discuss the connection between matroids and lattices.
The lattice theoretic analogue of a matroid is the notion of a <u>geometric lattice</u>.
As a matter of fact, some people like to do the whole theory of matroids just talking about lattices.
As an example:
People who are interested in arrangements of hyperplanes, where instead of points, you take hyperplanes, you find the geometric lattice defined by intersecting these hyperplanes.

So, in order to get from matroids to lattices, we need to discuss one of the most important notions of mathematics—and, in particular, combinatorics.
That's the notion of <u>closure</u>.

I should have done this before, but somehow didn't get around to it.

Sometimes I <u>slip</u> and call this a <u>closure relation</u>.
Lot's of people call these closure relations.
They are <u>not</u> relations, however.
It's a misuse of language.
There are lots of books where you will see closure relations.
They are <u>not</u> relations.

Closure

Closure is a notion invented by the great American mathematician E. H. Moore, who invented many, many things.
For example, he invented finite fields.
When he invented finite fields and published the first paper on finite fields in the American Journal of Mathematics, a famous European mathematician named ? stated: "At last we have a mathematical concept on which we can be sure there will never be any applications, whatsoever."
Little did he know that half of Course 6 (Electrical Engineering and Computer Science) is working on finite fields and coding theory.

E. H. Moore invented many other things, but he was cursed with a very bad habit.
Namely, he wanted his own notation for everything.
And, as a consequence, nobody read anything.

It would be a very nice project for one of you to pick up these books by E.H. Moore (of which there are 3 or 4), which are called "General Analysis" and rewrite them so that people can read them in the language of today.
It would be a genuine help to know what E. H. Moore really had.
Because we are discolored.
He invented the notion of convergence, for example, in topological spaces, which was reinvented by several people.
And heaven knows what else.

Anyway, his notion of closure took.

Given a set S, probably infinite, the <u>closure</u> is a function from $P(S)$ to $P(S)$, written:

$$\overline{} : P(S) \to P(S)$$

bar

Let $A \subseteq S$, then $A \to \overline{A}$ is the closure of A, subject to the following 3 properties:

(1) $A \subseteq \overline{A}$

(2) $\overline{\overline{A}} = \overline{A}$

(3) $A \subseteq B \Rightarrow \overline{A} \subseteq \overline{B}$

Any function from sets to sets with these properties is called a closure.

If $A = \overline{A}$, we say the set A is <u>closed.</u>

Before I give you examples, let's prove the <u>one and only</u> theorem about closures :

Theorem 1

The intersection of any number of closed sets is closed.

If A_i are closed, then $\bigcap_i A_i$ is closed.

Proof

$$\overbrace{\bigcap_i A_i \subseteq A_i}^{\text{from definition of intersection}} \underbrace{= \overline{A_i}}_{\text{since the } A_i \text{ are closed, by assumption}}$$

Then we take closures of both sides. From property 3, we have :

$$\overline{\bigcap_i A_i} \subseteq \overline{A_i} = A_i = \overline{\overline{A_i}}$$

↑ from property 2, $\overline{A_i} = \overline{\overline{A_i}}$

Then one can argue :

$$\overline{\bigcap_i A_i} \subseteq \underbrace{\bigcap_i \overline{A_i}}$$

and since the A_i are closed,

$$\bigcap_i \overline{A_i} = \bigcap_i A_i$$

$$\overline{\bigcap_i A_i} \subseteq \bigcap_i A_i$$

Now let's do it the other way around.
Let's consider the set $\bigcap_i A_i$. By property 1, we have :

$$\bigcap_i A_i \subseteq \overline{\bigcap_i A_i}$$

Combining this with the equation above, we must have the <u>equality</u> :

$$\bigcap_i A_i = \overline{\bigcap_i A_i}$$

Thus $\bigcap_i A_i$ is closed. As desired.

Now what is interesting are the examples of closures.
Like many mathematical concepts, you don't understand them until you see the typical examples.
In the case of the closure, you have completely different examples.

Example 1

Suppose you have a closure where the finite union of closed sets is closed.

Assume, in addition to the 3 properties, that:

$$\overline{\overline{A} \cup \overline{B}} = \overline{A} \cup \overline{B}$$

this doesn't follow from the 3 properties.

Then it's called a topology.

The study of this closure is called topology.
Most closures don't satisfy this additional property.

Example 2

V = vector space

$A \subseteq V$

Set $\overline{A}$ to be the vector space spanned by A.

$$\overline{A} = \text{span}(A)$$

Obviously $A \to \overline{A}$ is a closure.

But note that the property in example 1 is not satisfied:

$$\underbrace{\overline{\overline{A} \cup \overline{B}}}_{\text{span}} \neq \underbrace{\overline{A} \cup \overline{B}}_{\text{union of 2 subspaces}}$$

This closure has the important property that is called, lo and behold, the exchange property.
Not at random.
So let's write the exchange property more properly.

Steinitz Exchange Property

After Steinitz, who was the inventor of fields,
He wrote one big, huge paper of about 200 pages, where the whole field of fields was invented.
Everything. Incredible.

Take this closure in a vector space.
We have the following property:

Span has the following property:
If $y \in \overline{A \cup x}$ but $y \notin \overline{A}$ then $x \in \overline{A \cup y}$

That's the Steinitz Exchange Property.
We'll connect it with the Exchange Property we've seen previously [24.9 Theorem 3] soon.

Let's prove that this closure satisfies the Steinitz Exchange Property.
It's very important to verify this in detail.

$y \in \overline{A \cup x}$ is assumed.

That means that y is in $\mathrm{span}(A \cup x)$.

This means that y is a linear combination of x and elements of A.

$$y = \sum_i \lambda_i a_i + \lambda x \quad , \quad \lambda_i, \lambda \in \text{field}$$

That's just the first assumption.
Now let's use the 2nd assumption.
y is not a linear combination of just elements of A.

$$y \notin \overline{A} \Rightarrow \lambda \neq 0$$

Therefore, we can rewrite the above equation as:

$$x = \underbrace{\lambda^{-1} y - \sum_i \lambda^{-1} \lambda_i a_i}$$

But this is just a way of saying:

$x \in \overline{A \cup y}$

And that's the Steinitz Exchange Property.

We will see that every matroid satisfies this.
In every matroid, you can find a closure that satisfies this exchange property.
In that sense, matroids resemble vectors.

Example 3

Convex closures.

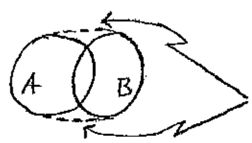
In $\mathbb{R}^n$, $\overline{A}$ = smallest convex closed set containing A

How do you know there is such a set?
The intersection of two convex closed sets is convex closed.
Take all the convex closed sets containing A and intersect them all.
That's a convex closed set containing A.

This is a closure that does not satisfy the Steinitz Exchange Property.
It satisfies certain properties, that I don't want to go into, that more or less characterize it, called antiexchange.

Just to give you an example, this closure does not satisfy the property in example 1:

$$\overline{\overline{A} \cup \overline{B}} \neq \overline{A} \cup \overline{B}$$



if you take the union of two closed sets A + B,
the convex closure is usually bigger.
You have to round it up to get the convex closure.

Example 4

Let P = any partially ordered set

Given $A \subseteq P$, set $\overline{A}$ = smallest order ideal containing A.

This defines a closure.
And this closure does satisfy the property:

$$\overline{\overline{A} \cup \overline{B}} = \overline{A} \cup \overline{B}$$

This is satisfied with lot's to spare.
You can take an arbitrary union.
The union of the closure of an arbitrary union is a closure.

You can say that the order ideals of a partially ordered set form a topological space, but a very special one.
Because the union of any number of closed sets is closed.

A lot of work has been done in characterizing these topological spaces.
But we don't have time to discuss this at length.
It was one of the topics we crossed out. [16.1]

Now, lo and behold, matroids.
I'll just define it now so you think about it until we meet on Friday.

Example 5 — matroids

Given a matroid (S, r), define $\overline{A}$ for $A \subseteq S$ as follows:

$$\overline{A} = \{A \cup x : r(A \cup x) = r(A)\}$$

$\overline{A} =$ A plus all x s.t. $r(A \cup x) = r(A)$

We've seen that: (The Whitney Property) $\left.\begin{array}{l} r(A \cup x) = r(A) \\ r(A \cup y) = r(A) \end{array}\right\} \Rightarrow r(A \cup x \cup y) = r(A)$

You can keep adding and the rank stays the same.
So it's consistent.
You keep adding as much as you can.
It's almost obvious that this is a closure.
We'll prove this in detail next time.

Then we'll prove that this closure <u>satisfies exactly</u> the <u>Steinitz Exchange Property.</u>
Just in the old days of linear algebra.

Closures and Geometric Lattices

Let me begin by reviewing.
Last time we defined the notion of closure,

(improperly called closure relation. But it's not a relation. Yet people often say closure relation. I don't know why.)

A closure is a map from sets to sets.
Namely:

$$A \to \overline{A} = Cl(A), \text{ all } A \subseteq S \text{ (often infinite)}$$

satisfying the properties:

(1) $A \subseteq \overline{A}$

(2) $\overline{\overline{A}} = \overline{A}$

(3) $A \subseteq B \Rightarrow \overline{A} \subseteq \overline{B}$

This is a universal concept of mathematics.
Once you know it, you see it everywhere, like pink elephants.
If you don't know it, you don't see it.
That's why people in biology, for example, don't do mathematics. Because they don't know what to see.
If you don't know what to see, you don't see it.

If you know about closures, you see closures everywhere and you start thinking about things in terms of closures.
And it helps.

And we saw last time that there is essentially only one simple theorem about closures.
There are many complicated theorems, but only one simple theorem.
Namely, that the intersection of closed sets is closed.

And we saw that you must not be misled to confuse the notion of closure with a topological notion of closure.
The topological notion of closure is a very special case of a closure that, for historical reasons, enjoyed an immense amount of attention this century, under the name of topology.

A topological closure also satisfies the property:

$$\overline{A \cup B} = \overline{A} \cup \overline{B}$$

Most closures in this world don't satisfy this.

Topologists don't like to hear that such topological closures are rare.
Because a topological closure defines the closed sets of sets of a topological space.
And a topological space is something very similar to a matroid, in the sense that it enjoys several cryptomorphic definitions.

You can define a topological space using open sets, closed sets, convergence, coverings.
Similarly, you can define a matroid in terms of rank, independent sets, basis, and, as we shall shortly see, closures.

Then we began to define the closure associated with every matroid.

By the way, I decided to teach this course again in the Fall, 1999, in order to cover the topics I couldn't cover here.
You can take the course again in Fall, 1999.
I guarantee that there will be <u>no</u> overlap with this course. Not even the notion of partially ordered sets.
It will be totally disjoint from this course.
We'll start with species, then we'll do totally positive matrices, all sorts of other things.
Actually, we'll start with Möbius functions in the Fall, 1999, because I don't think we'll get to them in this course.

So, we'll finish matroids today and then start with geometric probability.
And we'll cover geometric probability until the end of the term.

Geometric Probability is such a neglected subject.
It's full of research problems.
I will mention some of them.

I want to get to the point where at least you see what the fascinating, open research problems in the theory of matroids are.
Then we'll begin with geometric probability.

Closures associated with matroids

Given a matroid (S, r), we define a closure $A \to \overline{A}$, for $A \subseteq S$, as follows:

$$\text{set } \overline{A} = A \cup \{x : r(A \cup x) = r(A)\}$$

Theorem

$A \to \overline{A}$ is a closure and it satisfies the Steinitz Exchange Property:

$$\text{If } y \in \overline{A \cup x} \text{ but } y \notin \overline{A} \text{ then } x \in \overline{A \cup y}$$

means the same as "and"

Exercise 30.1

Every finite set endowed with a closure that satisfies the Steinitz Exchange Property defines a matroid.

Prove this.
How? Like this.
You define an independent set.
And we already know how to go from independent sets to ranks.
From independent sets, we want the rank that is the size of the maximal independent set contained in the set.
So, from this definition, if you can define the notion of independent sets, then you get a matroid.

Hint: Let I be "independent" when, for every $x \in I$, $x \notin \overline{I - x}$.

in quotes, because you don't really know yet

Think of a tree.
You remove any edge x in the tree not in the closure $\overline{I-x}$.

Once you have the definition and you prove that "independent" sets satisfy the exchange property for independent sets, then you're back in business.

This is the way matroids are developed in my old book that became "Theory of Matroids."

So let's prove the theorem. [30.2]
Like all theorems I proved, I have to look it up, because I've blocked out the proof.
First we prove that $A \to \overline{A}$ is a closure, satisfying properties 1-3 of a closure [30.1].
Then we show that it satisfies the Steinitz Exchange Property.

Proof

(1) $A \subseteq \overline{A}$

Obvious, since $\overline{A} = A \cup \underbrace{\{x : r(A \cup x) = r(A)\}}_{\text{you are adding stuff to } A}$

$A \subseteq \overline{A}$ ✓

(3) $A \subseteq B \Rightarrow \overline{A} \subseteq \overline{B}$

Given $A \subseteq B$

Apply the submodular inequality to the matroid (S, r) rank function:

$$r(A' \cup B') + r(A' \cap B') \leq r(A') + r(B')$$

Let $\left.\begin{array}{l} A' = A \cup x \\ B' = B \end{array}\right\} \Rightarrow$ $A' \cup B' = A \cup x \cup B$

since $A \subseteq B$

$$= B \cup x$$

$$A' \cap B' = (A \cup x) \cap B$$

$$= \underbrace{(A \cap B)}_{\text{since } A \subseteq B,\ A \cap B = A} \cup \cancel{(x \cap B)}^{0}$$

$$= A$$

Substituting into the submodular inequality gives:

$$r(B \cup x) + r(A) \leq r(A \cup x) + r(B)$$

Recall that the closure of A is defined here as:

$$\overline{A} = A \cup \underbrace{\{x : r(A \cup x) = r(A)\}}$$

if we have x s.t. $r(A \cup x) = r(A)$, then the submodular inequality above becomes:

$$r(B \cup x) + \cancel{r(A)} \leq \cancel{r(A \cup x)} + r(B)$$

by the increasing property of a rank function,
$r(B) \leq r(B \cup x) \Rightarrow r(B) \leq r(B \cup x)$
Therefore, we must have equality here.

$$r(B \cup x) = r(B)$$

$$x : r(A \cup x) = r(A) \Rightarrow x : r(B \cup x) = r(B)$$

Finally, since $A \subseteq B$:

$$\underbrace{A \cup \{x : r(A \cup x) = r(A)\}}_{\overline{A}} \subseteq \underbrace{B \cup \{x : r(B \cup x) = r(B)\}}_{\overline{B}}$$

$$\overline{A} \subseteq \overline{B} \checkmark$$

(2) $\overline{\overline{A}} = \overline{A}$

Suppose $x \in \overline{\overline{A}}$

recall that:

$$\overline{A} = A \cup \{x : r(A \cup x) = r(A)\}$$

$$\overline{\overline{A}} = \overline{A} \cup \underbrace{\{x : r(\overline{A} \cup x) = r(\overline{A})\}}$$

$$x \in \overline{\overline{A}} \Rightarrow r(\overline{A} \cup x) = r(\overline{A})$$

$$\begin{aligned} r(\overline{A}) &= r(A) && \text{by the definition of } \overline{A} \\ &\le r(A \cup x) && \text{by the increasing property of a rank function} \\ &\le r(\overline{A} \cup x) && \text{again, by the increasing property of a rank function, since: } (A \cup x) \subseteq (\overline{A} \cup x) \\ &= r(\overline{A}) && \text{from implication above: } x \in \overline{\overline{A}} \Rightarrow r(\overline{A} \cup x) = r(\overline{A}) \end{aligned}$$

Therefore, everything gets squeezed and we have the equality:

$$r(A \cup x) = r(A)$$

i.e., $x \in \overline{A}$

Therefore:

$$\overline{\overline{A}} = \overline{A} \quad \checkmark$$

So properties 1–3 [30.1] are satisfied and $\overline{A}$ so defined is, indeed, a closure.
Finally, we prove that $A \to \overline{A}$ satisfies the Steinitz Exchange Property:

(Steinitz Exchange Property)

This is really trivial. You just write everything out and stare at it.

$$y \in \overline{A \cup x} \quad \Rightarrow \quad r(A \cup x \cup y) = r(A \cup x) \quad \text{from the definition of this closure}$$

$$y \notin \overline{A} \quad \Rightarrow \quad r(A \cup y) = r(A) + 1$$

Since y is not in the closure of A, $r(A \cup y) \neq r(A)$. And from definition of rank function, we know that: $r(A \cup y) = r(A) + \{^0_1$ Therefore $r(A \cup y) = r(A) + 1$.

$$\begin{aligned} r(A \cup x \cup y) &= r(A \cup y) + \{^0_1 \\ &= r(A) + 1 + \{^0_1 \\ &= r(A) + \{^1_{\cancel{2}} \end{aligned}$$

But it can not be 2, since $r(A \cup x) = r(A) + \{^0_1$

Thus:

$$r(\overbrace{A \cup x \cup y}^{A \cup y \cup x}) = r(A \cup y) \Leftarrow \text{this is the same as saying } x \in \overline{A \cup y} \quad \checkmark$$

So the Steinitz Exchange Property is just a translation of the definition.
GCR: "Who's buried in Grant's tomb?" as Mr. Guidi says. ☺
JNG: No. It's "Mr. Guidi will find out." I don't know.

So we've proved that this is, indeed, a closure and that the Steinitz Exchange Property is satisfied.
The theorem is proved.

Now something I should have told you before.

- Every closure $A \to \overline{A}$, for $A \subseteq S$, defines a lattice L, as follows:

just properties 1-3 [30.1]

elements of L are all closed sets.

$\overline{A} \wedge \overline{B} = \overline{A} \cap \overline{B}$ The meet of 2 elements is ordinary intersection.

$\overline{A} \vee \overline{B} = \overline{\overline{A} \cup \overline{B}}$ The join of 2 elements is the closure of their union.

- Exercise 30.2

Prove that L, so defined, is a lattice.
Pretty trivial.

- Exercise 30.3

A slightly less trivial exercise.
State precisely and then prove: ⟵ I like to state exercises this way.

Most lattices arise from this construction.

There is a natural condition on lattices.
If I tell you it, it becomes trivial immediately.

- In particular, every closure defined by a matroid defines a lattice.
The closed sets of a matroid are called flats.

The flat of rank 1 = point
2 = line
3 = plane
$n-2$ = coline
$n-1$ = coatom (or copoint) ⟵ Sometimes flats of rank $n-1$ are called hyperplanes.

The lattice of flats of a matroid is called a geometric lattice.

A geometric lattice is the lattice of flats of a matroid.

Is there a specific characterization of geometric lattices?
Let's see.

Let L = geometric lattice

↑ let's look at the atoms.
It's tempting to say that the atoms give you the set S you started with.
That's not true. Almost true.
It ought to be true, as they say in philosophy.

Why is it almost true?
Because in a matroid you can have a closure of the empty set that is not the empty set.
Don't be fooled.
We have the rank of the empty set is zero. $r(\emptyset) = 0$
But there may be points of zero rank. That's not excluded.

So there may be points of zero rank.
There may be two points which are dependent on each other. That's not excluded either.

I never told you before and I'm sorry to inform you, at this point, of this unpleasant fact.
But that's the way it is.
In fact, it's good. We'll give an example where this really happens.

Atoms of geometric lattice L = closed set of rank 1

In many matroids, it's true that the closet sets of rank 1 are exactly the points.
In all the examples we have seen, that was the case.
So this often happens.

And it often happens that the closure of the empty set is the empty set:

$$\overline{\emptyset} = \emptyset$$

We'll consider, later, examples where the above cases do not happen.

Suppose we are given:

$x \in L$ ← abuse of notation
x is an <u>atom</u> of L.

$A = \bar{A} \in L$

Then we have:

$$r(A \cup x) = \begin{cases} r(A) & \text{if } x \leq A \\ r(A)+1 & \text{otherwise} \end{cases}$$

← [if x is an atom, which is the closure of any element which I obtained in A, adding x doesn't increase the rank.]

Furthermore:

Every $\underbrace{A = \bar{A}}_{\text{every closed set}}$ is the sup of a set of atoms,

This is enough to <u>characterize</u> <u>geometric lattices</u>.
Namely:

Every element is the sup of atoms.
If you take an element of the lattice and add an atom (i.e., take the sup of the element and the atom) either the rank stays the same or it goes up by exactly 1.

What about the Steinitz Exchange Property?
How does it translate in terms of lattices?
Very elegantly.
This is the:

<u>Birkhoff Covering Problem of a Geometric Lattice</u>

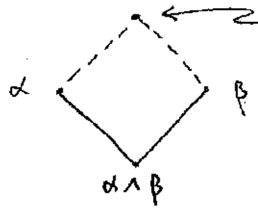
L is a geometric lattice.
I denote elements of L by Greek letters (α, β, etc.).

$\alpha, \beta \in L$ They are closed sets of some sets, but I imagine these closed sets as elements of the Hasse diagram.
Pretend I don't know where they come from.

If both α and β <u>cover</u> $\alpha \wedge \beta$ then $\alpha \vee \beta$ covers both α and β.

cover means immediately above

Every geometric lattice satisfies this covering property.
Let's see what this means.



Whenever you have 2 elements immediately above one, then their sup is immediately above these 2 elements.

This is just the Steinitz Exchange Property restated.
It's "Who's buried in Grant's tomb?" at it's worst.

Exercise 30.4

I'll prove the Steinitz Exchange Property by gestures.
You write it down as an exercise.
It's just too simple.

What does it mean for α to cover $\alpha \wedge \beta$?
It means you get α by taking $\alpha \wedge \beta$ and suping it with some atom.
Similarly, you get β by taking $\alpha \wedge \beta$ and suping it with another atom, since β covers $\alpha \wedge \beta$.
Under these circumstances, the sup of α, β is the sup of $\alpha \wedge \beta$ and <u>two</u> atoms.

Is that hard?
It's obvious.

Conversely, if you have the Birkhoff Covering Property and the sup of atoms, then you have a matroid.
Conversely, every lattice satisfies the Birkhoff Covering Property, where every element in the sup of atoms defines a matroid on the set of atoms.

Conversely, if L is a finite lattice where every element is the sup of a set of atoms and that satisfies the Birkhoff Covering Property, then L is a geometric lattice.

More precisely:

Let S = set of atoms of L

Define $\overline{A}$ = the set of atoms $x \in S$ s.t. $x \leq \sup A$

Then we obtain a matroid. We get a closure with the Steinitz Exchange Property.

In particular, given any matroid (S, r), let:

$$S_1 = \{\bar{x} - \bar{\emptyset},\ x \in S\}$$

The matroid (S, r) defines naturally a matroid on S_1.
We got rid of the crud.
Make every point closed and take away the closure of the empty set.

A matroid is naturally defined on S_1.

If you want a more elegant construction, given any matroid (S, r), take its geometric lattice and let:

S' = set of atoms of this geometric lattice.

Then use the construction above. Namely, define:

$\bar{A}$ = the set of atoms $x \in S'$ s.t. $x \le \sup A$

This construction defines another matroid that, lo and behold, is isomorphic (i.e., has the same geometric lattice).

See, what matters is the <u>geometric lattice</u> of a <u>matroid</u>.

<u>Exercise 30.5</u>

By the way, I haven't proved Whitney's Theorem. [25.4, 26.4]
I leave it to you as an exercise.

Prove Whitney's Theorem.

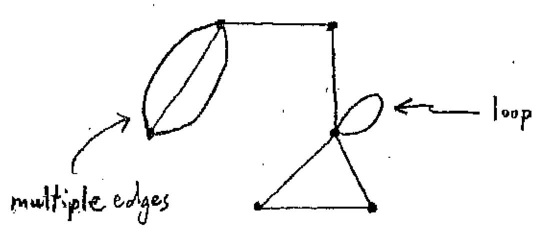
It's purely technical. Just do it by induction. (see also [31.4])

I'm glad I'm teaching this course again next year, because we've covered so little material.
I apologize for going so slow.
On the other hand, there are many undergraduates in the course and people with different backgrounds.
So it's better to go slowly.
So this fundamental stuff – it's better to hammer it in.
Mr. Guidi is over there rubbing it in.

Now you ask, what's an example of a matroid that has all these funny things we discussed earlier? [30.7]
Points dependent on each other, etc.
Let's look at some examples now.

Example 1 – Multigraph

A graph is a set of pairs over the set S, because they are atoms in the lattice of partitions.
But we can imagine two points being connected by different edges – multiple edges.
And, you can imagine an edge having only one edge point. This is called a loop.



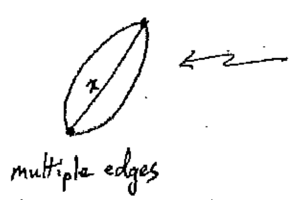
Multigraph

How do we define a matroid in a multigraph?
Well, you look at the various definitions of matroid and pick the one that is most convenient.

In this case, we proceed as follows.
We say a set of edges is independent if it's a tree.
Note, for example, that a loop is not independent.

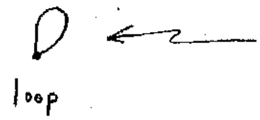
So we define a matroid using independent sets, which are represented as trees.
We define a matroid that way because these are equivalent definitions.

Independent sets = Trees



Now you see that multiple edges connecting the same two endpoints are dependent on each other.

The closure of edge x is the set of 3 edges, because the additional edges do not increase the rank.



The closure of the empty set contains every loop.
A loop has rank equal zero.

So what you do is make all the multiple edges into single edges and discard all the loops.
And you get another matroid, whose geometric lattice is isomorphic to the geometric lattice of this matroid.

This is sometimes called a combinatorial geometry, when every point is closed.
A matroid, where every point is closed and the closure of the empty set is the empty set is called a combinatorial geometry.

Example 2 – Linear Algebra

Instead of taking projective space, let's take a vector space.

The closure of a set of points is the smallest subspace through the origin containing that set of points.

Then 2 points on the same line are dependent on each other.
The closure of a point is the line (through the origin) spanned by that point.
And two points on the same line are dependent.

Remember, we have two operations on matroids – restriction and contraction
Let's see what these mean with geometric lattices.

Restriction

Restriction means you have a matroid and you restrict it to a subset A.
It's still a matroid. It doesn't "know."

In terms of geometric lattices, it's this:

Given S = set of atoms of a geometric lattice L

S defines a matroid, which is defined on the set of atoms.

Take a subset of atoms, $A \subseteq S$.

What will be the geometric lattice corresponding to this set A?
Easy.
You take all the sups of these elements. That's a geometric lattice.
The infs will be different. But sups will be the same.

So here we have a prime example of a situation where sups coincide where infs are completely different, depending on what you take a subset of.

Take all sups of subsets of A and you get another geometric lattice, which is called a restriction.

We did this for graphs. [26.13-14]
We took subsets of the set of atoms of partitions, that's a set of edges, then we took their sups.
That's a geometric lattice.
For a graph, you get what is called the lattice of contractions of a graph.

Contraction

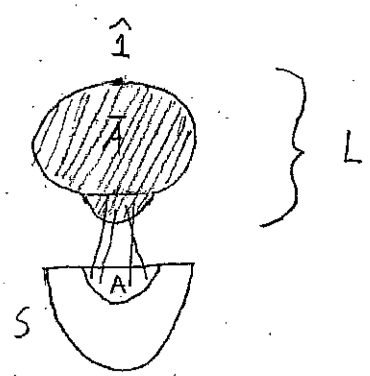
(not to be confused with what I just said, i.e. the lattice of contractions of a graph).

For a matroid, contraction is the following:

$$\text{matroid } (S-A, r_A) \text{ where } r_A(B) = r(A \cup B) - r(A)$$

So, what will be the geometric lattice analogue of contraction?

You take sup of A. That gives you a lattice L.
Then you take the interval of items between A and $\hat{1}$.
And this is a geometric lattice that is isomorphic to the contraction matroid above.



Exercise 30.6

The above is an almost trivial exercise.
State this and prove it.

Upper segments of a geometric lattice correspond to contractions.

Exercise 30.7

Every element of a geometric lattice L is the meet of a set of hyperplanes.

You can look at geometric lattices upside down and define a closure on hyperplanes.
But that closure does not satisfy the Steinitz Exchange Property. — N.B. about points in $\pi(n)$
It can be characterized, but it's artificial, because it's upside down.
By the general theorem of closure, we get another closure of hyperplanes.
You take a meet and get the hyperplane above.

So, next time, we will finally state what the fundamental problems and results of matroids are; Because we now have the language.
And then we start right away on geometric probability.

As I mentioned last time, this course will continue next fall with the subjects that we left out from our list. [16.1]
Next fall the course will start with Möbius functions.

You must not think that the theory of matroids is just what we dealt with in this course.
We have barely touched the surface of the theory of matroids.
The theory of matroids is a very deep and extensive theory.
And we've covered just the bare essentials.

So today I'd like to show you, descriptively, some of the really deep problems in the theory and how they are stated. Of course, I can't give you proofs.

The concept of a matroid is a very stable concept, as I mentioned before.
Historically, it arose in the most unlikely of places.
It arose in the theory of fields, in the theory of <u>transcendental extensions of fields.</u>
That's when Steinitz noted that if you took the lattice of transcendental extensions of a field, this lattice was not modular, but satisfied the Birkhof Covering Property, which is equivalent to the exchange property, which is now called the Steinitz Exchange Property.
That's the first example, historically, where matroids arise. Transcendental extensions of fields.
It's also the least studied example, strangely enough.
In recent years, there has been only 1 paper studying transcendental extensions of fields, from the point of view of matroids.
This paper is by Björner and Lovász, two outstanding combinatorialists.
It would be nice to go back to transcendental extensions of fields and see what one can say from the point of view of matroids.

There will be a Combinatorial Brunch, where we are supposed to discuss only combinatorics.
It will count like a class.

Combinatorial Brunch, Sunday December 6.
I think it's better if we meet at the Charles Hotel rather than my apartment.
Because, otherwise, we have to first assemble and then march over.
We'll meet at the entrance of the dining Hall, inside the hotel.
Everybody who sits in this course is invited.

11:57 AM, because I made a reservation for everybody for noon. I made a reservation for 26 people.
The Charles Hotel is at Harvard Square, near Harvard.
You go to Harvard Square and ask where the hotel is.

Q: How shall we dress for brunch?
A: There is no dress required. Don't be too disheveled.

It's all you can eat. So be hungry.

The discussion will be exclusively on combinatorial topics.

So let's finish up on matroids and geometric lattices.
And I would like to state what the fundamental results and problems in the theory of matroids are.
It will be slightly handwaving, because we haven't developed all the techniques.
But I think you'll get an idea of what's going on.

Geometric Lattices

Given a matroid (S, r), one can obtain from this matroid another matroid, where every point is closed and the closure of the empty set is empty, by simply trimming it.

Given matroid (S, r)

Say:

(1) $\overline{x} = x$, for all $x \in S$

(2) $\overline{\emptyset} = \emptyset$

Sometimes these matroids are called <u>combinatorial geometries.</u>

I tried to give this name, long ago, but it didn't take.

you might assume that every matroid satisfies this. But that's <u>not</u> so.
For example, remember that the orthogonal matroid depends very much on whether points are closed.

If you do not assume this, then you can have several matroids that have the same combinatorial geometry associated with them, but have different orthogonal matroids,
Because a basis of the orthogonal matroid is the combinatorial basis.

So if you have points of rank 0, they count in the orthogonal geometry, because the orthogonal geometry has changed.

However, from the point of view of lattices, a combinatorial geometry is the set of atoms of the <u>geometric lattice of flats</u> (closed sets) of the matroid.

Let L = geometric lattice
S = its atoms

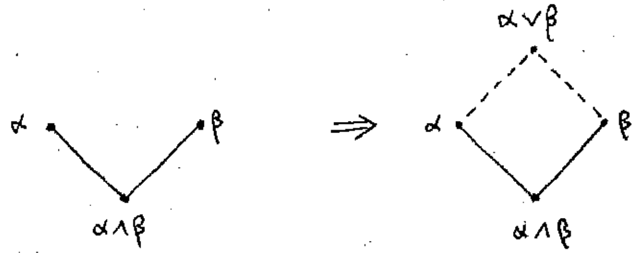
We can characterize intrinsically the <u>geometric lattice</u> by saying every element is the sup of atoms.

Every $\alpha \in L$ is $\alpha = \vee A$, $A \subseteq S$
and it has the <u>Birkhoff Covering Property.</u>

equivalent to properties required for a geometric lattice [30.7-8]

Birkhoff Covering Property

If α and β cover $\alpha \wedge \beta$, then $\alpha \vee \beta$ covers both α and β.



Observe that from the Birkhoff Covering Property, you can immediately infer that all maximal chains between $\hat{0}$ and some arbitrary α have the same length.

Birkhoff Covering Property implies that <u>all maximal chains have the same length.</u>

{ We've proved this 2 or 3 times already, from different points of view. Let's pretend we don't know it and see how it comes out of the Birkhoff Covering Property, alone. }

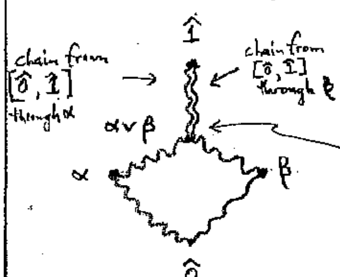
Let's show this.
First, let me observe the following.
If you define a geometric lattice with these properties, then it follows immediately that:

Every interval of a geometric lattice is a geometric lattice.

Hence, if L is a geometric lattice and $\alpha, \beta \in L$, $\alpha \leq \beta$, then $[\alpha, \beta]$ is a geometric lattice.

{ Proof: The sup of atoms property follows from the argument we've seen already. And the Birkhoff Covering Property is independent as to where you are. }

To prove the implication, we can proceed by induction.
We draw an example:



There are 2 maximal chains between $\hat{0}$ and $\hat{1}$.
You need to show they have the same length.
Both of them have to pass through an atom, because they are <u>maximal</u>.

Take atoms $\alpha + \beta$, take their sup $\alpha \vee \beta$. Since $\alpha + \beta$ cover $\hat{0}$, by the Birkhoff Covering Property, $\alpha \vee \beta$ covers both α and β.

Now you apply the induction hypothesis to the interval $[\alpha \vee \beta, \hat{1}]$, which is a geometric lattice.
And to the intervals $[\alpha, \hat{1}]$ and $[\beta, \hat{1}]$, by induction.
You have segments of the chains that overlap.

So the implication that all maximal chains have the same length, given the Birkhoff Covering Property, comes out immediately.
This is the classic argument.

Hence, a <u>rank $r(\alpha)$</u> exists.

which is really the <u>rank</u> of the <u>flat</u> of the matroid

Now, let's pretend we <u>don't know</u> the <u>submodular law</u>.
Let's pretend we don't know about matroids.

(This is actually the idea of the proof, which I left you as an exercise, of Whitney's Theorem [30.10 Exercise 30.5]. Now I give it away.)

Consider the inequality:

$$r(\alpha \vee \beta) + r(\alpha \wedge \beta) \leq r(\alpha) + r(\beta)$$

We rearrange terms and rewrite it this way:

$$(*) \quad r(\alpha \vee \beta) \leq r(\alpha) + r(\beta) - r(\alpha \wedge \beta)$$

Observe that the restriction:

$$r_{\alpha \wedge \beta}(\gamma) = r(\gamma \vee (\alpha \wedge \beta)) - r(\alpha \wedge \beta) \text{ is the } \underline{\text{rank}} \text{ of } [\alpha \wedge \beta, \hat{1}]$$

Therefore, inequality $(*)$ can be written as:

$$r_{\alpha \wedge \beta}(\alpha \vee \beta) \leq r_{\alpha \wedge \beta}(\alpha) + r_{\alpha \wedge \beta}(\beta)$$

All we have to prove is that in an arbitrary geometric lattice,
$r(\alpha \vee \beta) \leq r(\alpha) + r(\beta)$.
But this is <u>obvious</u> from the <u>Birkhoff Covering Property</u>, which says, in essence, every time you add an element, you get one back.

[if α and β cover $\alpha \wedge \beta$, then $\alpha \vee \beta$ covers both α + β.]

We also observe that: ← T any subset of atoms

If $T \subseteq S$, then the set of all α's s.t. $\alpha = \vee U$, for some $U \subseteq T$, is a <u>geometric lattice</u>, called the <u>restriction.</u>

A <u>contraction</u> is a geometric lattice in the interval $[\alpha, \hat{1}]$.
A <u>restriction</u> is taking a subset of atoms and taking all the <u>suplets</u>, where the <u>sups correspond</u> with the sups in the <u>big lattice</u>, but the <u>infs</u> do <u>not.</u>

A <u>minor</u> is the <u>restriction</u> of a <u>contraction</u>.

Let's see what happens for graphs.

A <u>graph</u> is a <u>restriction</u> of the <u>lattice of partitions.</u>

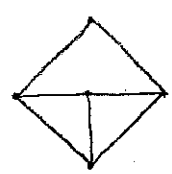
We take the lattice of partitions, take a subset, we call them edges.
Then we take their sups.
We forget they are partitions, we look at the edges.

If we look at the edges and don't want to talk about partitions, what does the lattice look like?
And we get a geometric way of visualizing the <u>lattice of contractions</u> of a graph.

{ the geometric lattice of a graph is called the <u>lattice of contractions</u> of a graph. }

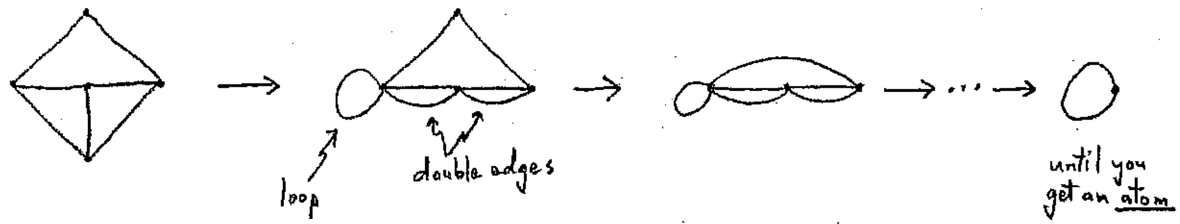
We have an underlying set T.
S = subset of the set of atoms of $\Pi[T]$

We visualize this as a graph, where T are the <u>vertices.</u>



What does it mean to take the <u>geometric lattice</u> generated by this set, where the sups are the same as the sups in the lattice of partitions?
We make elements of T equivalent, according to the edges. Successively.

You keep track, by drawing the loops (which are really unnecessary), of what has been contracted.



This is the classic way of visualizing the lattice of contractions of a graph.
Mathematically, it's just taking joins of partitions.

So what's a minor of the lattice of contractions of a graph?
You take a subset of the edges and you contract only those edges.
That's it.

Big Theorems of Matroid Theory

"A matroid is good iff its geometric lattice does not contain any minor isomorphic to one of the following finite list: $\underbrace{\ldots}$ "

These are the hard theorems.
Some of them proved.
Some of them conjectures.
The problem is we don't understand the mechanism for proving these theorems. We don't have a general machinery for proving these theorems. They're all proven by ad-hoc methods. But there ought to be a general machinery for establishing these.

I've worked most of my life trying to establish some machinery (some super homological machinery), but to this day we don't know how.
So let me tell you what some of these theorems are.

Example

A graph is 4 colorable iff it does not contain a minor isomorphic to the complete 5-graph.
That's equivalent to the 4 color conjecture.
This was proved by Dirac, the son of the physicist Dirac.
And it doesn't involve planarity.
Dirac proved that this is equivalent to the famous 4 Color conjecture about planar graphs.

A graph is 4 colorable iff its geometric lattice of contractions has no minor isomorphic to the lattice of contractions of the complete 5-graph.

Conjecture

Example — Hadwiger's Conjecture

A graph is n-colorable iff its lattice of contractions does not contain a minor isomorphic to the lattice of contractions of the complete $(n+1)$-graph.

A couple of years ago, an extraordinary result was obtained by Professor Seymour of Princeton and Professor Robertson of Ohio State.
They proved that Hadwiger's Conjecture is true, provided that the 4 Color Conjecture is true.
Assuming the 4 Color Theorem is true, then Hadwiger's Conjecture is true.
This was a tremendous tour de force.
Strangely enough, their proof uses the theory of well ordered sets.

So, these are some of the big conjectures.
Now let's see some of the things that are easier to prove.

Dirac's Conjecture excludes minors of the complete 5-graph.
What if, instead, we exclude minors of the complete 4-graph?

How good is a graph if its lattice of contractions does not contain a minor isomorphic to the lattice of contractions of the complete 4-graph?

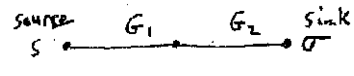
We get something very nice.
Duffin's Theorem. This was proved a long time ago.

Duffin was the teacher of John Nash and Raoul Bott.
He was Professor at Carnegie Mellon.
He was probably the greatest circuit theorist of his time.
It's too bad that we couldn't cover any circuit theory in this course. No time.
It's a beautiful subject that should be covered in a math course.

To explain Duffin's Theorem, we need a new concept.
The concept of a series-parallel network.

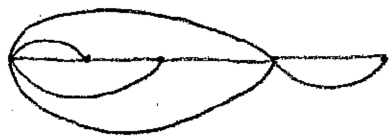
What's a series-parallel network?
It's a multigraph, a graph with loops and multiple edges, which is obtained as follows:

You have an infinite supply of edges.
You can "combine" edges by two operations — a series connection or a parallel connection.
Let G_1 and G_2 be two graphs.

Series connection


parallel connection


A series-parallel network is a multigraph obtained by iterating these two operations.
For example:



A series-parallel network

Since a series-parallel network defines a multigraph, it defines a matroid.
(We've seen that matroids can be defined for multigraphs, as well as for graphs.)

And what do these matroids look like?
Guess what?

Duffin's Theorem

A lattice of contractions of a graph is series-parallel iff the complete 4-graph is an excluded minor.

This is not hard to prove, but it's not trivial either.
This is one of the "easy" theorems of matroids.

Notice that this theorem has an extraordinary consequence.
In this theorem, there is no mention of the source and sink.
So how can the lattice know where the source and sink are.
The answer is — you can take any two vertices and make them source and sink.

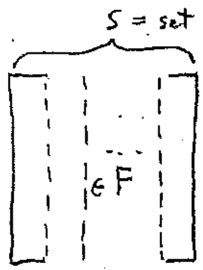
So if a graph is series-parallel for one source and one sink, then pick any two vertices, it will be series-parallel for this new source and new sink.
This is a consequence of Duffin's Theorem.

This is something, philosophically, I have never understood.
Because series and parallel are two operations.
Now we discover that the operations don't matter. You can take any two completely different operations.
You have two arbitrary operations, each of which is commutative and associative, and you combine them in arbitrary ways.
That's a series-parallel graph.

Now Duffin's Theorem tells you that you can get this in a completely different way.

Let's see another "easy" matroid theorem.
Recall that:

A matroid (S, r) is <u>representable</u> over a field F
iff
there is a matrix whose entries are $\in F$ s.t.
if you take S = set of columns and
consider the rank of any subset of columns S, in the linear algebra sense
then the rank of S, in the linear algebra sense, is isomorphic to the matroid (S, r).



This gives you a representation of a matroid over a field.

So the question is: When can a matroid be represented over a given field?
Is there a finite number of excluded minors that guarantees reprentability over a given field?

That's an <u>unsolved</u> question.

This is <u>solved</u> for a <u>field</u> of <u>2 elements.</u>
It's kind of easy.
When can a matroid be representable with a matrix whose entries are 0 or 1?

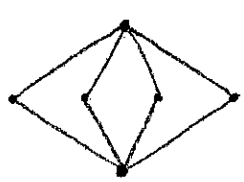
$1+1=0$
$1*1=1$

The answer is the following:

***Exercise 31.1**

A matroid is representable over $\overbrace{GF(2)}^{\text{Galois Field with 2 elements}}$ iff its lattice of contractions does <u>not</u> have the minor:

{Matroids representable over $GF(2)$ are said to be <u>binary matroids</u>.}



This is a necessary and sufficient condition.
It's very elegant. Prove this.

The deepest representation theorems are due to Tutte.
They are all concerned with when is a matroid representable over any field whatsoever.

This is equivalent to asking when can a matroid be represented by a matrix that is totally unimodular.
This can be proved.
Tutte found that there are 3 excluded minors, which I don't have time to describe.
One of them is the minor above in Exercise 31.1 and there are 2 more.

If the matroid does not have any 3 of these minors in its lattice of contractions then the matrix is totally unimodular.

Then, the question is:

When can a matroid be represented as a lattice of contractions of a graph?
The answer is that there are 5 excluded minors. The 3 from the totally unimodular case, plus 2 more.

Then, the question:

When can a matroid be represented as a lattice of contractions of a planar graph?
The answer is 7 excluded minors. The 5, from above, plus 2 more.

These are the big Tutte theorems.
The deepest theorems to date on matroids.

Let me conclude by giving you a very elegant characterization of the lattice of partitions of a set, on the basis of this result.
To do that, we need the notion of a modular element in a geometric lattice.

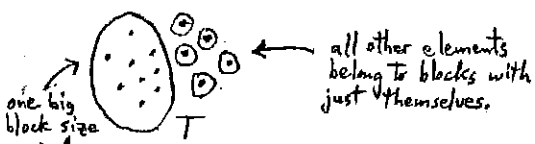
If L is a geometric lattice, $\alpha \in L$ is modular when, for all $\beta \in L$,

$$r(\alpha \vee \beta) + r(\alpha \wedge \beta) = r(\alpha) + r(\beta)$$

Exercise 31.2

So, let's look at the lattice of partitions. And let's get a feel for modular element, by looking at the lattice of partitions.
What's a modular partition? I'll tell you and you check it as an exercise.

In $\Pi[T]$, an element α is modular iff it is a partition with only 1 block of size > 1.



modular element α of $\Pi[T]$

You have to check, as an exercise, that these are the only modular elements of the lattice of partitions.

Then we have the following theorem:

Kung's Theorem

We can characterize the lattice of partitions, as follows.
The lattice of partitions is the only binary matroid, where every element has a modular complement.

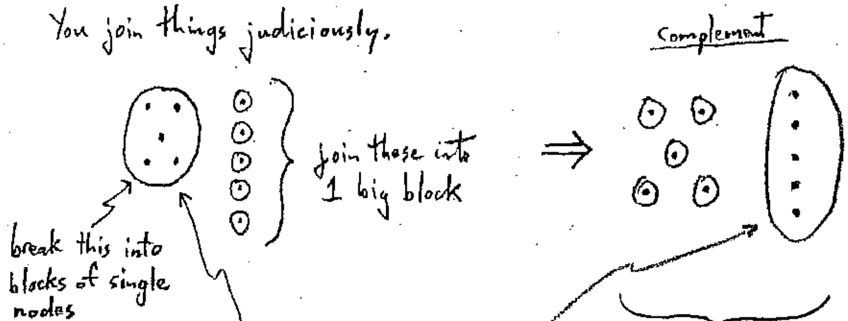
If L is a <u>binary geometric lattice</u>, where every element has a <u>modular complement</u>, then:

$$L = \Pi[T]$$

modular complement = the join is $\hat{1}$,
the meet is $\hat{0}$,
and which happens, also, to be a modular element.

Now you say - why?
In the lattice of partitions, you can only find partitions that are modular where the sup is $\hat{1}$ and the inf is $\hat{0}$.

You join things judiciously.



this gives you a <u>modular partition</u>.

The join of this <u>partition</u> with this <u>partition</u> is $\hat{1}$.
The meet is $\hat{0}$, because they are split.

It would be interesting to extend this to <u>infinite sets</u>.
To characterize the lattice of partitions of an <u>arbitrary</u> set.
There is, currently, no nice characterization.

You see from this that this is just the tip of the iceberg.
There's a lot more.
We didn't talk about the <u>Critical Problem</u>, which is the generalization of the coloring problem on a graph to arbitrary geometric lattices.
What coloring is to the lattice of contractions of the graph, you can apply these theorems to arbitrary geometric lattices. You can ask similar questions. That's a full course.

We'll stop here. Wednesday after Thanksgiving, you'll turn in your problems. And we start on geometric probability.

Reminder: We meet on Sunday (December 6) at 11:57 AM, in the dining room of the Charles Hotel, which is located in the neighborhood of Harvard Square. Walking distance from the MBTA station in Harvard Square.

Also, you have a problem set due on Wednesday, where you do 1/3 of the problems that are assigned. And, if possible, one or two starred problems.
I will give you some problems today to choose from.

Geometric Probability

You are wondering what geometric probability is about.
Let me tell you orally, while I'm erasing the blackboard.
The original problem of geometric probability is the following:

You have, in ordinary n-dimensional space, a certain object. For example, a convex closed set.
Then you have, in your hand, a rigid object, of a very bad shape - all twisted up, but rigid.
Then you drop the rigid bad object, at random, on n-space. For example, on the plane.
Then you ask for the probability that the rigid bad object will meet the good, round object that you have drawn.

In this form, the problem doesn't make sense, because the probability is not defined, since you have a density in space.
So you have to embed the round object into a big cube and compute the conditional probability that the bad object will meet the good, round object, given that it falls within the big cube.
And that makes sense.
And that's the basic problem of geometric probability.

Solve this for any round object and any bad object whatsoever.

The amazing thing is that this problem is not as hard as it sounds.
And the solution depends very little on the shape of the objects.
That's the amazing thing. This is our first motivation.

Our objective is to understand how this problem is solved.
In order to do that, we start on an entirely different tack.
As a matter of fact, we'll look at two completely different motivations that seem totally unrelated.
Then we will see that they are very closely related.

Our second motivation is this.
You take a family of subsets in n-dimensional Euclidean space, which are sufficiently nice so that we are not immeshed in measure theoretic questions.
We want combinatorics, not measure theory.
What is a sufficiently good family of sets? It's what we call a polyconvex set.
Polyconvex sets are finite unions of compact, convex sets.

Geometric Probability

$\mathcal{L}$ = lattice of all polyconvex sets in $\mathbb{R}^n$

↑ finite unions of compact, convex sets

{ These are the sets we will be dealing with. Any reasonable set is a polyconvex set, although it's easy to give examples of sets that are not polyconvex. }

The union of two polyconvex sets is a polyconvex set. That's obvious.

The intersection of two compact, convex sets is a compact, convex set.

Therefore, by the distributive law, the intersection of two polyconvex sets is a polyconvex set.

Therefore, polyconvex sets form a distributive lattice, containing the empty set $\emptyset$, and not containing a $\hat{1}$.

$\mathcal{L}$ is a distributive lattice.

↑ { What are distributive lattices for? They are made to order to define measures. That's what distributive lattices are for. }

Our objective (seemingly different from the objective I stated 5 minutes ago in our first motivation) is to study measures on this distributive lattice.

Measure

A measure on $\mathcal{L}$ is a function μ from $\mathcal{L}$ to the real numbers, not necessarily positive, with the properties:

$$\mu : \mathcal{L} \to \mathbb{R} \quad \text{s.t.}$$

$$(1) \quad \mu(\emptyset) = 0$$

$$(2) \quad \left.\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B), \quad A, B \in \mathcal{L} \right\} \ \mu \text{ is } \underline{\text{additive}}$$

I gave you as an exercise early in the term the fact that every measure satisfies the inclusion-exclusion formula. [8.11 Exercise 8.7]
I hope you've done it. We'll need to use it.

We want to study measures on the lattice of polyconvex sets.
But there are too many of them.
So, we have to impose <u>non-degeneracy assumptions</u> on these measures.

The non-degeneracy assumption that is imposed by analysts is that it should be <u>countably additive</u>.

{That's used in probability. We will <u>NOT</u> assume this.}

We make the following <u>assumptions</u>:

(1) μ is invariant under the <u>group of rigid motions</u>

(for those of you who know group theory, the group of rigid motions is the semidirect product of the orthogonal group to the group of translations.)

If μ were countably additive and invariant under the group of rigid motions, we would immediately know what μ is. A volume.

But if μ is <u>not</u> countably additive, a funny fact is that there are lots of these μ. And the study of these measures is the object of geometric probability.

Now you have to assume the non-degeneracy assumption.
Since we are <u>NOT</u> assuming countable additivity, we have to assume something else that prevents the measures from having funny behavior.

(2) μ is <u>continuous</u>, in the following sense:

C_n = sequence of compact, convex sets

Suppose C_n converges to compact, convex set C:

$$C_n \to C$$

The standard notion of convergence of sets, meaning the maximum distance between points in C_n and C tends to zero.

We say that <u>μ</u> is <u>continuous</u> when:

$$\lim_{n\to\infty} \left(\mu(C_n)\right) = \mu(C)$$

A perfectly reasonable assumption.

Mr. Guidi is here in person. ☺

Our objective will be to study measures on $\mathbb{R}^n$ which are (1) invariant under rigid motions and (2) continuous, in this sense.
And we classify them all.
And we will see that the classification of these measures entails the solution of the geometric probability problem I stated at first.

Observe that lattice $\mathcal{L}$ has an important sublattice.

$$\mathcal{L}_{pol} = \text{lattice of all } \underline{\text{polyhedra}}$$

Q: What's a polyhedron?
A: A polyhedron is the finite union of compact, convex polyhedra.

Q: What's a compact, convex polyhedron?
A: It's the convex analogue of a finite number of points.
Or, the intersection of a finite number of closed hyperplanes, which is convex.
By the Hahn-Banach Theorem, these two definitions are equivalent.

$\mathcal{L}_{pol}$ is also a lattice
The union of a polyhedron with a polyhedron is a polyhedron.
The intersection of a polyhedron with a polyhedron is a polyhedron.
And it's a sublattice of $\mathcal{L}$.

$\mathcal{L}_{pol}$ is a sublattice of $\mathcal{L}$.

{this means that union and intersection in $\mathcal{L}_{pol}$ is the same as union and intersection in $\mathcal{L}$.}

Since this theory is ripe with unsolved problems, let me state right away that, whereas all the continuous invariant measures on $\mathcal{L}$ have been classified, this is <u>not</u> the case for <u>continuous invariant measures on $\mathcal{L}_{pol}$.</u>

no one has classified these.
You want an immediate Ph.D., solve this problem.
Instant Ph.D.
It's probably not hard. You just have to get the right idea.

There <u>seem</u> to be more continuous invariant measures on $\mathcal{L}_{pol}$ than on $\mathcal{L}$.

Instead of giving you yet a <u>third</u> motivation, which turns out to be closely related to the two we have discussed, let's have a humble beginning.

- <u>A Humble Beginning : $\mathbb{R}^1$</u>

Let's see what happens on the ordinary real line.
Let's get a feeling,

What's a <u>polyconvex set in $\mathbb{R}^1$</u> ?

> what's a compact, convex set in $\mathbb{R}^1$?
> It's a <u>closed interval</u>.
> So a polyconvex set is a finite union of closed intervals.

So now the problem is to classify all measures defined on finite unions of closed intervals, which are invariant under translation.
And you say: "Ha, Ha, Ha. I know the answer."
And I say: "No, No. You don't."

$$\mathcal{L} = \text{all finite unions of closed intervals.}$$

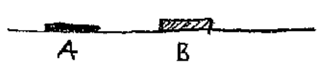
And now I give you two examples of invariant measures.
This is slightly upsetting, because you are expecting just one. Right?

- $\mu_1(A) =$ length of A

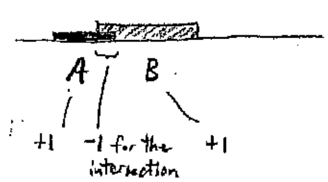
That's obviously an invariant measure.
If you don't see it, I can't help you.

- $\mu_0(A) =$ number of connected components of A

This is a measure.
Let's see. It's best shown by picture.



if disjoint, then μ_0 is clearly additive



if the intervals overlap, $1-1+1=1$, so it checks.

μ_0 is an invariant measure.

Now we prove the following theorem:

Theorem

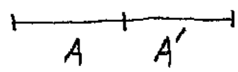
Every continuous invariant measure on $\mathcal{L}$ in $\mathbb{R}^1$ is a linear combination of μ_0 and μ_1.

This is kind of nice.

Proof (remember that we are dealing with compact, convex sets, i.e., our sense of continuous)

Case 1: $\mu(p) = 0$, p = a point

This means that μ of every point is zero, because it's invariant. That means if I take an interval A and I double it with the interval A', there is only one point of intersection:



And from the additive property of a measure:

$$\mu(A \cup A') = \underbrace{\mu(A) + \mu(A')}_{= 2\mu(A)} - \cancelto{0}{\mu(A \cap A')}$$

$A \cap A'$ is a point. And, from the assumption, $\mu(\text{point}) = 0$.

if A' has the same organization as A, that means that doubling the length doubles the measure. Therefore, μ is the length, by a well known argument, which I will not insult you by repeating.

$= 2\mu(A)$

Cauchy's functional equation and all that nonsense.

Therefore:

$$\mu(A) = \text{constant} * \text{length of } A$$

$$\mu(A) = c\mu_1(A) \quad \checkmark$$

case 2: $\mu(p) \neq 0$

μ of a point is NOT zero.

Without loss of generality, assume: $\mu(p) = 1$ } So every point has measure 1.

Consider $\mu' = \underbrace{\mu - \mu_0}_{\text{that's an invariant measure}}$

Then $\mu'(p) = 0$

p = point

And μ' reduces to case 1:

$$\mu'(A) = c\mu_1(A)$$

Hence:

$$\mu'(A) = \mu(A) - \mu_0(A)$$
$$c\mu_1(A) = \mu(A) - \mu_0(A)$$

This gives μ as a linear combination of μ_0 and μ_1:

$$\mu(A) = \mu_0(A) + c\mu_1(A) \quad \checkmark$$

Now we have to do this in n-dimensions.
That's extremely tough.
In fact, the first elementary proof was obtained 2 years ago by Dan Klain at Georgia Tech.
Before that, the only proof known was 122 pages.

So now you see that in 1 dimension, there are two invariant measures – μ_0 and μ_1.
What is μ_0 really?
The Euler characteristic, as we shall see.

By the way, all this material is in my book "Introduction to Geometric Probability" with Klain.
Except I am presenting it differently so as to not cheat you.
A different point of view.
But it is there. The facts are there.

Let's generalize μ_0 and μ_1 to $\mathbb{R}^n$.
Then you will realize that there are more invariant measures in $\mathbb{R}^n$.

In $\mathbb{R}^n$, we take a polyconvex set.

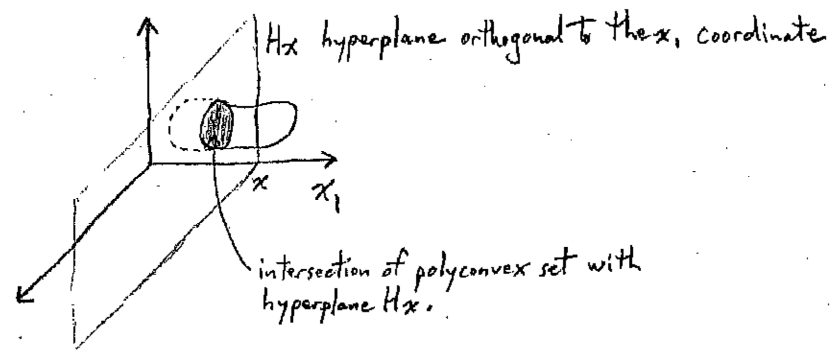
One invariant measure is the <u>volume</u>.
Let me remind you what the <u>volume</u> is, may I?
From course 18.02 (Calculus).
We define:

$$\mu_n(A) = \text{volume of } A$$

every compact, convex set has a volume.
A polyconvex set is a finite union of compact, convex sets.
So the volume is well defined.

How is the volume computed?
You take an orthogonal coordinate system and you do it by integration with multiple integrals.
Let's do that.

$x_1, \dots, x_n$ = orthogonal coordinates



$$\mu_n(A) = \int \mu_{n-1}(A \cap H_x)\, dx$$

imagine — an integral in a combinatorics class.
Not for long.

That's the way you compute volumes in course 18.02.
Now you say "So what?".

Now we are going to <u>generalize</u> this to define the analog of μ_0 in n-dimensions.

Defining a <u>measure</u> is the same as defining a <u>linear functional</u> on <u>simple functions</u>.
They teach this in course 18.100 (Analysis).

$A \subseteq \mathbb{R}^n$,

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}, \quad \omega \in \mathbb{R}^n$$

A <u>simple function</u> is a <u>finite</u> <u>linear combination</u> of <u>functions I_A</u> (called <u>indicator functions</u>).
Finite linear combinations of indicator functions give you a simple function.

Indicator functions are a vector space, automatically.

So, we have the following theorem.
Back from my functional analysis days, when I was your age.

[A linear functional on the vector space of indicator functions is always integration relative to a measure on polyconvex sets.
Conversely, every measure on polyconvex sets defines a linear functional.]

This is a fundamental fact.
This is the fundamental fact of the theory of integration, stripped of all the convergence crud.

Let's write this down. You seem to be more interested in this than I expected.

Theorem

Let L be a linear functional on the vector space of all simple functions on $\mathcal{L}$.
Then there exists a measure on $\mathcal{L}$ s.t.

if f is a simple function
then $L(f) = \int f \, d\mu$

(We start with a polyconvex set. But we might as well start with convex sets, because of the inclusion-exclusion formula.)

f is a simple function means that f is a finite linear combination of indicator variables:

$$f = \sum_i \alpha_i I_{A_i}$$

By definition:

$$\int f \, d\mu = \sum_i \alpha_i \mu(A_i)$$

This is a non-trivial fact.
Because the same simple function can be written as a linear combination of indicator functions in infinitely many ways.

i.e., $f = \sum_i \alpha_i I_{A_i}$ can be written in infinitely many ways.

You have to prove that this equality holds, regardless of how you write f.

This is non-trivial.
And that's the fundamental non-trivial fact of integration theory.
Don't you ever forget that.
They didn't tell you that in course 18.100. I hope they did, but probably they didn't.

The theorem is that this definition of the integral makes sense.

Exercise 32.1

Prove this theorem as an exercise.
It's not entirely trivial.

To repeat, you have to prove that irrespective of how you express the simple function as a linear combination of indicator functions, you always get the same integral.

No one says the A_i are disjoint.
They may overlap.

Conversely,
if μ is a measure on $\mathcal{L}$ then:

$$L(f) = \sum_i \alpha_i \mu(A_i) \quad , A_i \in \mathcal{L}$$

is a well-defined linear functional on the vector space of simple functions

This also applies when the A_i are compact, convex.

Exercise 32.2

Prove the converse above.

Now you know measure and integration.
This is the gist of the theory of measure and integration.
The rest is just limits.
So, this is fundamental fact theory that is bypassed in analysis courses.
It's something extremely fundamental.
I wish I could tell you how fundamental this is.

Now we go back to our problem of defining μ_n in n-dimensions. Recall we computed volume as: [32.8]

$$\mu_n(A) = \underbrace{\int \mu_{n-1}(A \cap H_x)\, dx}_{\text{multiple integrals}}$$

And guess what?
We're going to imitate this definition of volume in our definition of μ_0 in n-dimensions.
Instead of integrals, we use sums.
Watch:

$$\text{Set } \mu_0(A) = \sum_x \left(\mu_0(A \cap H_x) - \mu_0(A \cap H_{x^+})\right)$$

x^+ is the limit as you approach from $x+\varepsilon$, where $\varepsilon \to 0$, $\varepsilon > 0$.

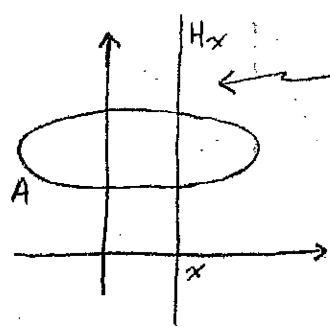
Now you say "Hey, isn't that an infinite sum?"
And I say "No. No."

Suppose A is a compact, convex set.
For how many x's will the term inside the sum be non-zero?

Let's look at this in 2 dimensions:

$$\underbrace{\mu_0(A)}_{\mathbb{R}^2} = \sum_x \left(\underbrace{\mu_0(A \cap H_x)} - \underbrace{\mu_0(A \cap H_{x^+})}\right)$$

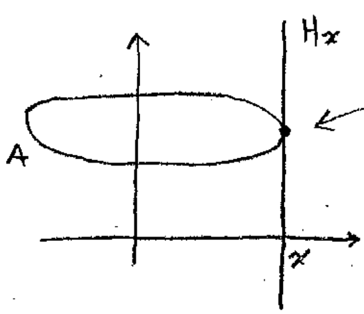
The intersections with the hyperplanes are in $\mathbb{R}^1$.
We've already defined μ_0 in $\mathbb{R}^1$ as the number of connected components of the polyconvex set. [32.5]
The hyperplanes are convex and A is polyconvex, thus the intersections are polyconvex.
The intersections are either an interval, a point, or null.



if H_x is inside, both $\mu_0(A \cap H_x)$ and $\mu_0(A \cap H_{x^+})$ will give 1 as the number of connected components.
So the term is 0.

$$\cancel{\mu_0(A \cap H_x)}^{\,1} - \cancel{\mu_0(A \cap H_{x^+})}^{\,1} = 0$$

Also, if the intersections with the hyperplanes are empty, the term is 0.
So it's only the right tangent point of a compact, convex set that matters.
That's the only non-zero contribution.



$$\underbrace{\mu_0(A \cap H_x)}_{1} - \cancel{\mu_0(\underbrace{A \cap H_{x^+}}_{A \cap H_{x^+} = \emptyset})}^{\,0} = 1$$

Then, by induction:

If A is compact, convex, then $\mu_0(A) = 1$.

But it's obvious that this is a measure.
And if you have a polyconvex set, you have a finite number of unions of compact, convex sets.

Theorem
Hence μ_0 exists in $\mathcal{L}$ on $\mathbb{R}^n$.

μ_0 is a measure on all polyconvex sets, with the property that:

$$\mu_0(A) = 1$$

if A is a non-empty, compact, convex set.

We have just proved one of the fundamental facts of mathematics.
There exists a measure on polyconvex sets that takes the value 1 on compact, convex sets.
If that's obvious, I quit.
That's not obvious at all. Because you can take unions (finite unions of compact, convex sets (themselves unions and intersections of compact, convex sets)) in weirdo ways, and you can have wholes all over the place.
But this theorem says no. The measure is well-defined.

This measure μ_0 is called the Euler characteristic.

Forget about topology.
Topologists go about this for half a term. We did it in half an hour.

Notice the strange parallelism between the sum defining the Euler characteristic μ_0 and the integral defining the volume μ_n:

$$\underbrace{\mu_0(A)}_{\text{Euler characteristic}} = \underbrace{\sum_i \left(\mu_0(A \cap H_x) - \mu_0(A \cap H_{x+})\right)}_{\text{multiple sums}} \qquad \underbrace{\mu_n(A)}_{\text{volume}} = \underbrace{\int \mu_{n-1}(A \cap H_x)\, dx}_{\text{multiple integrals}}$$

This parallelism is tantalizing.
We'd like to understand it better.

Next time, we'll see some applications, when we establish what the other measures are.

Geometric Probability (Cont'd)

Last time we gave the natural construction of the Euler characteristic.
Let's go over it briefly again, because it's an extremely important concept.
It's one of the fundamental concepts of mathematics.

You remember we imitated the definition of volume.
We are in $\mathbb{R}^n$.
We take:

$$\mathcal{L} = \text{lattice of all } \underline{\text{polyconvex}} \text{ sets}$$

↑ finite unions of compact, convex sets

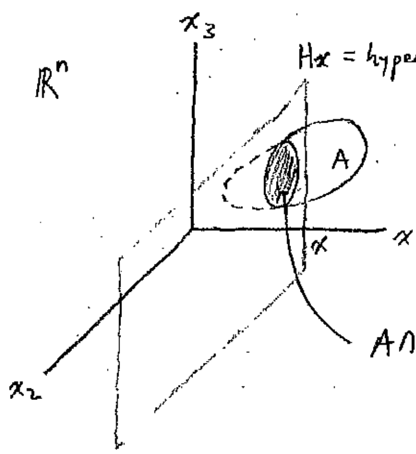
We studied measures on this lattice.
Any reasonable set is a polyconvex set.
In particular, any polyhedron is a convex set.
Once you have polyhedra, you can approximate everything by polyhedra, so this is a general concept.

And:

$$\mu_n(A) = \text{volume of } A \in \mathcal{L}$$

And we have the formula, from elementary Calculus, that:

$$\mu_n(A) = \int \mu_{n-1}(A \cap H_x)\,dx$$



H_x = hyperplane through x, orthogonal to x_1 axis.

$A \cap H_x$ = <u>polyconvex</u> set in <u>n-1 dimensions.</u>

↑ $\left\{\begin{array}{l} H_x \text{ is convex and the intersection of a polyconvex} \\ \text{set with a polyconvex set is polyconvex.} \\ \text{So } A \cap H_x \text{ is polyconvex.} \end{array}\right\}$

You can inductively expand the integral for the volume until you get down to one dimension, where you have the length.
And that's multiple integration.

Now the remarkable fact is that the Euler characteristic is a tantalizingly similar formula.
I have thought, for many years, about how to bring these two definitions under the same root, by one conceptual scheme.
Maybe if you pay me \$10,000 I will do it.
I haven't done it.

$$\underbrace{\mu_0(A)} = \text{Euler characteristic of } A$$

We follow a similar process to computing the volume, but instead of multiple integration, we have multiple summation.

Notice that $\mu_0(A)$ is valid also for lower dimensions.
The Euler characteristic is defined in $\mathbb{R}^n$.
But in $\mathbb{R}^n$, you may have a lower dimensional convex set containing A.
The Euler characteristic will still be fine.

So, we could define:

$$\left.\begin{array}{l}\mu_{0,n} \\ \mu_{0,n-1}\end{array}\right\} \text{ they are the } \underline{\text{same}}$$

The Euler characteristic does not depend on the dimension of the space in which the compact, convex set is immersed.

Strictly speaking, even μ_n should be independent of the dimension.

So, we defined:

$$x^+ = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} x + \varepsilon$$

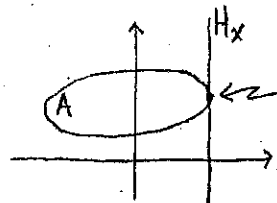
$$\mu_0(A) = \sum_x \left(\mu_0(A \cap H_x) - \mu_0(A \cap H_{x^+})\right)$$

The interesting fact is that this sum is well-defined.
Because there are only a <u>finite</u> number of x's for which the two terms are <u>not</u> equal, if A is a polyconvex set.

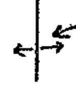
To prove this, you only have to verify this when A is convex.
Because, then by inclusion-exclusion, every polyconvex set can be written in terms of convex sets, by the inclusion-exclusion formula.

If A is compact, convex, we verified last time in 2 dimensions that there is <u>only one case</u> where:

$$\mu_0(A \cap H_x) \neq \mu_0(A \cap H_{x+})$$



Where A touches H_x as it's tangent:
$\mu_0(A \cap H_x) - \mu_0(A \cap H_{x+}) = 1$



if you move H_x just a little to the left or the right:
$\mu_0(A \cap H_x) = \mu_0(A \cap H_{x+})$

Furthermore, it is clear than <u>$\mu_0(A)$</u> is <u>a measure</u>:

$$\underbrace{\mu_0(A)}_{\mu_0 \text{ for } \mathbb{R}^2} = \sum_i \Big(\underbrace{\mu_0(A \cap H_x)}_{\substack{\text{this is a measure,} \\ \text{1 dimension lower } (\mathbb{R}^1).}} - \underbrace{\mu_0(A \cap H_{x+})}_{\substack{\text{and this is a measure,} \\ \text{1 dimension lower } (\mathbb{R}^1).}}\Big)$$

And we already have defined <u>μ_0 for $\mathbb{R}^1$</u>. [32.5]
That's the <u>number of connected components.</u>
So it checks.

And you can write:

$$\mu_0(A) = \underbrace{\sum_i \Big(\mu_0(A \cap H_x) - \mu_0(A \cap H_{x+})\Big)}$$

You could write this as <u>multiple sums</u> over <u>orthogonal coordinates</u>.

So you see this <u>strange parallelism</u> between this sum and this integral.

$$\mu_0(A) = \underbrace{\sum_i \Big(\mu_0(A \cap H_x) - \mu_0(A \cap H_{x+})\Big)}_{\text{multiple sums}} \qquad \mu_n(A) = \underbrace{\int \mu_{n-1}(A \cap H_x)\,dx}_{\text{multiple integrals}}$$

This parallelism is extremely tantalizing and we would like to understand it better. This has to do with commutativity and non-commutativity of variables, in a <u>very deep</u> sense.

In this way, we have defined a new measure.
Why is this measure invariant?

[We defined the measure in a particular coordinate system.
Invariant means the measure is independent of the position of the polyconvex set.
Namely, invariant under the group of rigid motions (i.e., rotations and translations).]

We just proved that:

$\underbrace{\mu_0(A) = 1}$ if A is a non-empty compact, convex set.

And this proves that it's invariant, because it's equal to 1 no matter where you place the compact, convex set.
This immediately proves the measure is invariant.

Let B be a polyconvex set (i.e., a finite union of compact, convex sets)

$$B = A_1 \cup A_2 \cup \ldots \cup A_k \quad , \quad A_i = \text{compact, convex set}$$

We take $\mu_0(B)$, using the classic inclusion-exclusion formula:

$$\mu_0(B) = \sum_i \mu_0(A_i) - \sum_{i<j} \mu_0(A_i \cap A_j) + \sum_{i<j<r} \mu_0(A_i \cap A_j \cap A_r) - \ldots + \ldots$$

$\mu_0(B)$ is always computable and is always an integer.
Here we have another number that you can associate with any body in space.
And it's independent of the position of that body.

If we know all the numbers that we can associate with bodies, which are independent of position, then we would know that any physical properties of these bodies should be expressable in these numbers.
So, it's very important to know what they are.

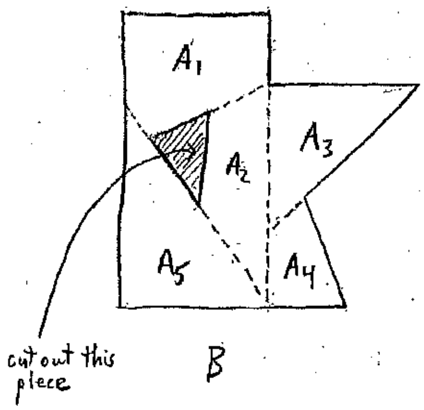
[And we will see the main theorem of geometric probability is that the dimension of the space of these invariant measures, which are continuous in the sense defined last time, is $1n+1$.
So there are $n+1$ basic measures.]

this is an extraordinary result, of fundamental importance and not widely known.
It tells you there are $n+1$ numbers that you associate with any body in space.
And that's all.
And any physical characteristic has to be expressable in terms of these $n+1$ numbers.
This is very important, if you ask me.

Let's fool around now with the Euler characteristic.
And let's connect it with the Euler characteristic, as per topology.
So far I've said that this is the Euler characteristic and you can rightfully ask "why is this the same as the view in topology?".
So we have to connect these.

How do you compute the Euler characteristic of something?
Like this:

Let's take this funny shape. It's a polyhedron. What's it's Euler characteristic?
It's easy.



I cut B up into <u>compact, convex sets.</u>

$B = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$

Then I use inclusion - exclusion : [33.4]

$$\mu_0(B) = \sum_i \mu_0(A_i) \qquad = \mu_0(A_1) + \mu_0(A_2) + \mu_0(A_3) + \mu_0(A_4) + \mu_0(A_5)$$

$$- \sum_{i<j} \mu_0(A_i \cap A_j) \qquad \begin{aligned} &- \mu_0(A_1 \cap A_2) - \mu_0(A_1 \cap A_3) - \mu_0(A_1 \cap A_5) \\ &- \mu_0(A_2 \cap A_3) - \mu_0(A_2 \cap A_4) - \mu_0(A_2 \cap A_5) \\ &- \mu_0(A_3 \cap A_4) - \mu_0(A_4 \cap A_5) \end{aligned}$$

$$+ \sum_{i<j<r} \mu_0(A_i \cap A_j \cap A_r) \qquad \begin{aligned} &+ \mu_0(A_1 \cap A_2 \cap A_3) + \mu_0(A_2 \cap A_3 \cap A_4) \\ &+ \mu_0(A_2 \cap A_4 \cap A_5) \end{aligned}$$

$$- \sum_{i<j<r<s} \mu_0(A_i \cap A_j \cap A_r \cap A_s) \qquad - 0$$

$$+ \sum_{i<j<r<s<t} \mu_0(A_i \cap A_j \cap A_r \cap A_s \cap A_t) \qquad + 0$$

$$= 0$$

$$\mu_0(B) = 0$$

So, if I have any polyhedron whatsoever, then you cut it up into compact, convex polyhedra, and then applying the inclusion-exclusion formula, you get the Euler characteristic.
And the <u>non-trivial fact</u> is that no matter how you cut it up, you get the same number.
↑ that's the theorem

From this fact that no matter how you cut a polyhedron up, you get the same number for the Euler characteristic, you can derive all sorts of theorems of geometry.

Now you say – why is this the Euler characteristic, as per topology?
How do we connect this to topology?
The best way is by the Euler-Schläfli-Poincaré formula that you learned while studying the world of mathematics in high school:

$$\text{Vertices} - \text{Edges} + \text{Faces} - \text{Holes in faces} = 2\,(\text{Components} - \text{Genus})$$

That's the formula we're going to make precise and derive now.
In the simplest possible way.

In order to do that, we have to do a little grammar.
I don't like to talk about this, but I have to.
It's really dull stuff.

Given set S,

$$\mathcal{L} = \text{distributed lattice of subsets}$$

And suppose μ is a finite measure:

$$\mu : \mathcal{L} \to \mathbb{R}$$

case 1: $S \in \mathcal{L}$

That means that $\mu(S)$ is finite:

$$|\mu(S)| < \infty$$

Take the Boolean algebra generated by $\mathcal{L}$.

{take the smallest Boolean algebra containing $\mathcal{L}$ and the complement of any set in $\mathcal{L}$. Take finite unions and intersections.}

Then μ extends uniquely to the Boolean algebra generated by $\mathcal{L}$.
This is called Pettis's Theorem.

Exercise 33.1

Prove Pettis's Theorem.

Unfortunately, in our case, Pettis's Theorem doesn't apply.
Because the set S is $\mathbb{R}^n$. It's infinite.
And we define polyconvex sets on finite unions of compact, convex sets.

So we have to doctor up Pettis's Theorem, so we can have our cake and eat it too.
We have complements in it, but we can't have big complements.
So we do something rather unpleasant, we take relative complements.

case 2: $S \not\subset \mathcal{L}$

Then μ can be extended uniquely to the distributive lattice generated by all sets of the form:

$$A \cap B^c \quad , \text{ for } A, B \in \mathcal{L}$$

you can't have all the complements of a compact, convex set, as that's infinite.
But you can intersect it. And that's okay.

Exercise 33.2

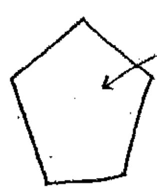
Prove case 2 above.
It's an extremely technical result that is intuitively obvious.

This is called combinatorial measure theory.
I should have given you a couple of lectures on combinatorial measure theory.
But it's too much of a sleeper.

So let's assume these 2 cases.
Then, by case 2, the Euler characteristic can be extended to all sets of the form:

$$A \cap B^c$$

Apply case 2 to the Euler characteristic.
In particular, you have the following.
If you have a compact, convex polyhedron, then the interior of the compact, convex polyhedron is a union of sets of the form $A \cap B^c$.



interior = union of sets of form $A \cap B^c$

this means the Euler characteristic can be extended to the interior of a compact, convex polyhedron.

compact, convex polyhedron.

You should remember that the word interior is ambiguous for a compact, convex polyhedron.
A compact, convex polyhedron has a definite dimension.
Namely, the dimension of the smallest hyperplane that contains it in the whole space.

By interior, I mean interior within the relative interior of the smallest hyperplane that contains it.
For example, if the above is a planar compact, convex polyhedron in 3 dimensional space, its interior is still the set, as indicated.

The Euler characteristic can be <u>extended</u> to the interior of compact, convex polyhedra.

mechanically, because it's all inclusion – exclusion

All this is grammar and now we have the fundamental theorem about the Euler characteristic.
By the way, this was the way I was going to do Möblus functions.
This is the preliminary to Möbius functions.

Fundamental Theorem – Euler Characteristic

If A is a compact, convex polyhedron of dimension n, then:

$$\underbrace{\mu_0(\mathrm{Int}(A))}_{\text{Euler characteristic of the interior of } A.} = (-1)^n$$

{this means that the smallest hyperplane that contains A is a hyperplane of dimension n.}

Proof

By the way, when A is a compact, convex set and x is <u>not</u> a coordinate of the border of A:

$$\mathrm{Int}(A) \cap H_x = \mathrm{Int}(A \cap H_x)$$

Also, the interior of a point is a point.
There's something kinky there, but you can't avoid it.

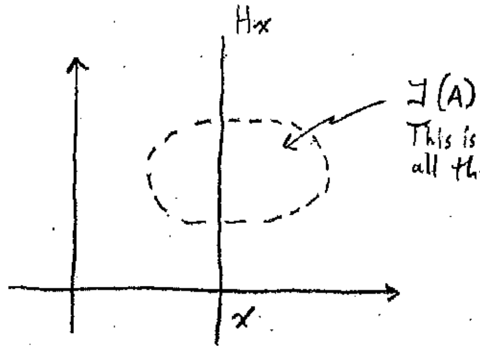
$$\mathrm{Int}(p) = p \quad , p \text{ a point}$$

Let's write out the definition of $\mu_0(\mathrm{Int}(A))$: [33.2]

$$\mu_0(\mathrm{Int}(A)) = \sum_x \left(\mu_0(\mathrm{Int}(A) \cap H_x) - \mu_0(\mathrm{Int}(A) \cap H_{x^+}) \right)$$

Remember that these intersections are one dimension <u>lower</u>.
Therefore, we can proceed by induction.

We establish the <u>base case</u> where the intersections have dimension 1.
Let's see for which x's the term in the sum is non-zero.
As before, let's draw a picture.
Things are a bit different than before. [32.11]



This is the interior of A. We're leaving out all the boundary.

If you take x here, such that the hyperplane H_x intersects inside the interior of A, then:

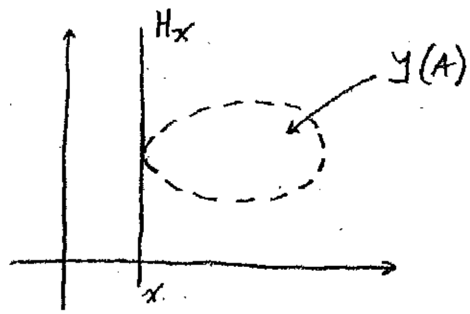
$$\underbrace{\mu_0(\mathrm{Int}(A)\cap H_x)}_{\text{1 connected component}} - \underbrace{\mu_0(\mathrm{Int}(A)\cap H_{x^+})}_{\text{1 connected component}} = 0$$

Now, <u>without</u> the <u>boundary</u>, the intersection with the right "tangent point" is null.
In this case:

$$\mathrm{Int}(A)\cap H_x = \emptyset \qquad \mathrm{Int}(A)\cap H_{x^+} = \emptyset$$

$$\cancelto{0}{\mu_0(\mathrm{Int}(A)\cap H_x)} - \cancelto{0}{\mu_0(\mathrm{Int}(A)\cap H_{x^+})} = 0$$

So the only non-zero contribution is at the left "tangent point":



$$\cancelto{0}{\mu_0(\mathrm{Int}(A)\cap H_x)} - \underbrace{\mu_0(\mathrm{Int}(A)\cap H_{x^+})}_{\text{1 connected component}} = -1$$

$$\mathrm{Int}(A)\cap H_x = \emptyset$$
$$\mu_0(\emptyset) = 0$$

This proves the base case.

With the base case in hand, consider only those x's that make a non-zero contribution to the measure.
Namely, those x's that are the coordinates of the hyperplanes in various dimensions that are left tangents to the interior of A.

$$\mu_0(\mathrm{Int}(A)) = \sum_x \Big(\cancelto{0}{\mu_0(\mathrm{Int}(A) \cap H_x)} - \underbrace{\mu_0(\mathrm{Int}(A) \cap H_{x^+})}_{\text{by the induction hypothesis, this is: } (-1)^{n-1}} \Big)$$

$$= -(-1)^{n-1}$$

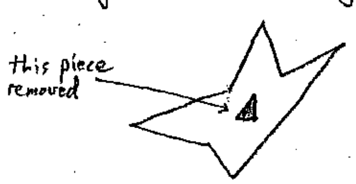
$$= (-1)^n$$

$\mu_0(\mathrm{Int}(A)) = (-1)^n$, Q.E.D.

So what?
Well, there's a collorary that gives the Euler-Schläfli-Poincaré formula.
Not just for compact, convex polyhedra.
For any polyhedra, whatsoever. Compact, convex or not.

What's a polyhedron?
A polyhedron is a finite union of compact, convex polyhedra.
By definition. Piece it together. It's not very hard.

But, if you take a finite union of compact, convex polyhedra, it's not clear what a face is.
If you take something like the following:



What are the faces?
We need faces to get the formula. Therefore I need to define the notion of face.
Let's define a system of faces of a polyhedron.

↑ system is just a set

If P is a polyhedron, a system of faces $\mathbb{F}$ of P is a set of compact, convex polyhedra s.t.

$$(1)\quad A \in \mathbb{F} \Rightarrow A \subseteq P,\ A \neq \emptyset$$

$$(2)\quad \bigcup_{A \in \mathbb{F}} \mathrm{Int}(A) = P$$

$$(3)\quad A, B \in \mathbb{F} \Rightarrow \mathrm{Int}(A) \cap \mathrm{Int}(B) = \emptyset \quad \leftarrow \text{interiors are disjoint}$$

Int means relative interior.

That's a face.
You have to accept it. Because I'm the teacher.

Before we see examples, let's prove the theorem.

Theorem (Euler - Schläfli - Poincaré)

The Euler - Schläfli - Poincaré formula is sometimes called Euler's formula.
The French call it Poincaré's formula.

Schläfli was a Swiss mathematician for whom I have a great admiration, for the following reason.
I bought a collection of papers and I started reading through them.
I saw 3 of my papers that he had done already - in 1857.

Let P be a polyhedron in $\mathbb{R}^n$ and let $\mathbb{F}$ be a system of faces of P.
Let f_i = number of elements of $\mathbb{F}$ of dimension i.

(the system of faces is a compact, convex set, so it has a dimension. Namely, the smallest hyperplane that contains it.
Non-convex sets don't, because the are twisted.)

$$\text{Then } \underbrace{\mu_0(P)}_{\text{Euler characteristic}} = f_0 - f_1 + f_2 - f_3 + \dots - \dots$$

That's the famous formula, made precise

Proof

$$\mu_0(P) = \mu_0\Big(\bigcup_{A\in\mathbb{F}} \mathrm{Int}(A)\Big) \quad \longleftarrow$$ from property 2 of definition of a system of faces.

From property 3 of the definition of a system of faces, any two interiors are disjoint.

Therefore, the inclusion-exclusion formula in this case is just the first sum. The remaining sums involve intersections, but these are disjoint and $\mu_0(\emptyset)=0$.

$$\mu_0\Big(\bigcup_{A\in\mathbb{F}} \mathrm{Int}(A)\Big) = \sum_{A\in\mathbb{F}} \mu_0(\mathrm{Int}(A)) - 0 + 0 \ldots - 0 \ldots + 0 \ldots$$

$$= \sum_{A\in\mathbb{F}} \mu_0(\mathrm{Int}(A))$$

Collect terms of equal dimension.
By the Fundamental Theorem [33.8], if $\mathrm{Int}(A)$ has dimension i, then:

$$\mu_0(\mathrm{Int}(A)) = (-1)^i$$

And there are f_i elements of dimension i.

$$= f_0 - f_1 + f_2 - f_3 + \ldots - \ldots$$

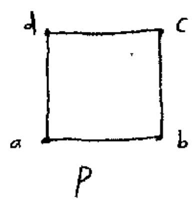
$\left\{ \begin{array}{l} f_0 = \text{vertices.} \\ \text{interior of a point is a point.} \end{array} \right\}$

That's it.
That's the formula, as desired.

So let's see how this works.
Enough with all this topology stuff.
You do this with main combinatorics stuff.
You can teach this to high school students.

Example

The Euler characteristic of a square $= 1$, because a square is a convex set.

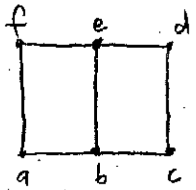


Let's find the system of faces.

$f_0 = 4$ vertices
$f_1 = 4$ sides
$f_2 = 1$ plane

$$\mu_0(P) = f_0 - f_1 + f_2 = 4 - 4 + 1 = 1 \checkmark$$

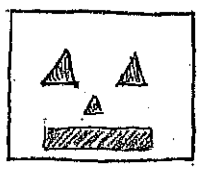
Example



P

$$\mu_0(P) = f_0 - f_1 + f_2$$
$$= 6 - 7 + 2$$
$$= 1 \checkmark$$

To compute the Euler characteristic of a complex shape in n dimensions, cut it up into compact, convex polyhedra and then take the system of faces.



P

$$\mu_0(P) = f_0 - f_1 + f_2 - f_3 + \dots - \dots$$

Time is almost up.
Let me do a little theorem.

Klee's Theorem

$A_i =$ compact, convex set

Then $B = \bigcup_i A_i$ is also a compact, convex set

If $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \neq \emptyset$ for all $i_1 < i_2 < \dots < i_k$
then $A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_{k+1}} \neq \emptyset$ for some $j_1 < j_2 < \dots < j_{k+1}$.

The theorem says if any k of these A_i have non-empty intersections, then some $k+1$ of these A_i have non-empty intersections.

This sounds highfalutin, but it's trivial.

Proof (this is in my book "Introduction to Geometric Probability," by the way)

$$1-1 = 0$$

$$(1-1)^n = 0$$

$$\underbrace{\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0}$$

the point being that any partial sum of these is non-zero:

$$\sum_{i=0}^{k} (-1)^i \binom{n}{i} \neq 0 \quad , \quad k < n$$

Handwaving, but you get the idea.
Take $k \leq \frac{n}{2}$.
Then the binomial coefficients are strictly increasing.
Thus the alternating sum can not be zero.
The argument for $k > \frac{n}{2}$ is equally straightforward.

Therefore, we have that:

$$(*) \quad \binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \dots + \dots \pm \binom{n}{k} \neq 1, \quad k < n$$

Since B is a compact, convex set, the Euler characteristic of B is:

$$\mu_0(B) = 1$$

$$= \underbrace{\mu_0\left(\bigcup_i A_i\right)}_{\text{expand by inclusion-exclusion}}$$

$$= \sum_i \mu_0(A_i) - \sum_{i<j} \mu_0(A_i \cap A_j) + \sum_{i<j<r} \mu_0(A_i \cap A_j \cap A_r) - \dots + \dots$$

If any k of the A_i have non-empty intersections, then all intersections of fewer than k of any A_i have non-empty intersections.

Since the A_i are compact, convex sets, any non-empty intersection of A_i's is a compact, convex set.
And, as we have shown [32.12], the Euler characteristic μ_0 of a compact, convex set is 1.

Thus each summation above in which all intersections of A_i's are non-empty simply involves counting the number of intersections.
Each such summation equals the corresponding binomial coefficient.
For example:

$$\sum_i \mu_0(A_i) = \binom{n}{1}$$

$$\sum_{i<j} \mu_0(A_i \cap A_j) = \binom{n}{2}$$

$$\vdots$$

$$\sum_{i<j<\dots<k} \mu_0(\underbrace{A_i \cap A_j \cap \dots \cap A_k}_{k}) = \binom{n}{k}$$

Given our assumption that all intersections of any k A_i's are non-empty.

We are given that all intersections of any k A_i's are non-empty.

Let's assume that the conclusion does not hold.
Namely, that all intersections of any $k+1$ A_i's are empty.
This, of course, immediately implies that all intersections of any $k+1$ or more A_i's are empty.

$$= \overbrace{\underbrace{\binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \dots + \dots \pm \binom{n}{k}}_{\text{first } k \text{ summations of the inclusion-exclusion expansion.}} \underbrace{\mp 0 \pm 0 \mp \dots \pm 0}_{\text{All intersections of any } k+1 \text{ or more } A_i\text{'s are empty. And, } \mu_0(\emptyset)=0.}}^{n \text{ terms}}$$

$$= \underbrace{\binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \dots + \dots \pm \binom{n}{k}}$$

And we just showed, equation (*), that:

$$\binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \dots + \dots \pm \binom{n}{k} \neq 1 \quad , \quad k < n$$

But $\mu_0(B) = 1$, so we have our contradiction.
There has to be at least an extra term that is non-zero.

So there must be at least one intersection of some $k+1$ A_i's that is non-empty.

So the proof is completely trivial.
The original proof was a big mess.

Next time, we will see what the other invariant measures are.

John Guidi
guidi@math.mit.edu

We continue today on geometric probability.
You are wondering what this has to do with probability.
We've seen that it has to do with measures, which is often a way to do probability.

We have seen that in the ordinary Euclidean space of n dimensions, that there are invariant measures, which are equally remarkable.
Namely, the volume and the Euler characteristic.

These can be considered as physical properties of Euclidean objects, if you wish, because they are invariant under rigid motions.

So if we can determine all of these invariant measures, we can rightly have said that we know how to express any physical property of these objects.
Any physical property should be expressable in terms of the object's invariant measures.

We will state, today, the main theorem of geometric probability to be the fact that the space of all invariant measures has dimension $n+1$, for an object in n dimensions.

We have seen in 1 dimension that the space of invariant measures has dimension 2. [32.5-7]
Because this space is spanned by the Euler characteristic μ_0, which in $\mathbb{R}^1$ is the number of connected components of a closed set, and the length μ_1.
Recall that in $\mathbb{R}^1$, we showed that:

$$\mu(A) = \mu_0(A) + c\,\mu_1(A)$$

A fundamental result will be that in n dimensions, the space of all invariant measures has dimension $n+1$.
Secondly, that there is a distinguished basis of these invariant measures that is physically meaningful.

So, how are we going to do this?
We need a little more grammar.
We need some more combinatorial measure theory.

Combinatorial Measure Theory (Cont'd)

From a strictly combinatorial viewpoint, we have:
Given set S,

$$\mathcal{L} = \text{distributive lattice of subsets}$$

We have a measure μ a function from $\mathcal{L}$ to the real numbers, not necessarily positive, satisfying the following properties:

$$\mu : \mathcal{L} \to \mathbb{R} \text{ s.t. } (1)\ \mu(\emptyset) = 0$$

$$(2)\ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

Why do we take a distributive lattice of subsets and not a Boolean algebra, such as they do in course 18.100 (Analysis)?
Because, in general, the measure μ might be infinite on the complement of the set.
For example, if you take the volume, the volume of a compact, convex set is finite.
But the volume of the complement is infinite.

We want our measures to be finite.
That's why, in general, complements are not included.
I explained last time, but did not prove, that such a measure can be extended to the relative Boolean algebra.

where you take relative complements, in the sense we discussed last time [33.7]

So, at least partly, complement can be included.
Provided you take the complement within a set in the distributive lattice.

Philosophically, measures are the set theory analogue of linear functionals.
And there is a complete crytomorphism between between the language of linear functionals on the vector space of functions from the set S to the reals and the language of measures.
That's an important lesson to learn, which I've already outlined before.

Combinatorially, of course, the only functions we allow, if you don't allow limits, are simple functions.
Namely:

$$\underbrace{f = \sum_i \alpha_i I_{A_i}}_{\text{all the linear combinations, with scalar coefficients, of indicator functions.}} \quad , \quad A_i \in \mathcal{L}$$

The important fact about the simple functions to remember is that a simple function can be written as a linear combination of indicator functions in infinitely many ways, in general.
There is no unique way of writing a simple function.

And that's what makes the fundamental fact about integration theory remarkable.
And that's why we do it again.
That even though the simple functions can be written infinitely many ways, nonetheless, there is an invariant called the integral (for the last 300 years), which means we define the integral:

$$\underbrace{\int f \, d\mu}_{\text{this expression is well-defined.}} = \sum_i \alpha_i \mu(A_i)$$

this expression is well-defined. Watch, because I'm going to pull a fast one on you.
Namely, it's the same irrespective of how you express the linear functional.

This is the fundamental fact about integration my friends.
The rest is limits.

Conversely,
Given a linear functional L on the space of simple functions,

Set: $\mu(A) = L(I_A)$

And you get a measure on $\mathcal{L}$.

And, lo and behold, the integral, relative to this measure, is the function that you started with:

$$L(f) = \int f d\mu$$

This is the whole story of measure integration.
Don't you ever forget it.
This is very fundamental.
Unfortunately, not taught this way in course 18.100 (Analysis).

Now, you want to take the next step in combinatorializing measure theory.
Namely, product measures.
Because we need that.
{You'd never think that in a course on combinatorics that you'd learn about measure integration.
But this stuff is very fundamental.}

Product Measures

We have 2 measures:

Given $\mu, \mathcal{L}, S$ and $\mu', \mathcal{L}', S'$

Then we take $S \times S'$ and you want to define a measure on $S \times S'$.

this is more delicate than it seems, at first.

You all know it, but I want to summarize a combinatorial crisis.
You have to define a distributive lattice of subsets of $S \times S'$.
But you can not take a product of an element of $\mathcal{L}$ and an element of $\mathcal{L}'$, because they do not form a lattice — you have to have unions and intersections of those.

If $A \in \mathcal{L}$, $A' \in \mathcal{L}'$, then $A \times A'$ = a rectangle.

But, the lattice you need is the lattice of all unions and intersections of rectangles.
This doesn't come out by just taking products.
What we take is the tensor product of the two lattices,

- Tensor Product

  $\mathcal{L} \otimes \mathcal{L}'$ is the lattice generated by finite unions and intersections of rectangles.
  ↑ tensor product

- Then, on this lattice $\mathcal{L} \otimes \mathcal{L}'$, you define a measure:

  Product measure μ'' is defined on $S \times S'$ and $\mathcal{L} \otimes \mathcal{L}'$ by setting:

  $$\mu''(A \times A') = \mu(A)\mu(A')$$

  for a rectangle.

- Exercise 34.1

  Prove that product measure μ'' has a unique extension to $\mathcal{L} \otimes \mathcal{L}'$.
  This is what product measures are about.

This is all very nice, but we'll be seeing in a minute that this is insufficient. We can not escape limits.
Let's see what happens.

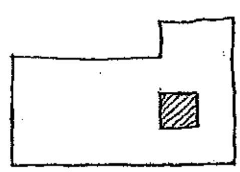
- By the way, why don't you use $\mu'' = \mu \times \mu'$ in defining the Euler characteristic?

  μ_0 = Euler characteristic on $\mathbb{R}^n$

  Why don't we take:

  $\mu_0 \times \mu_0 \times \cdots \times \mu_0$ on $\mathbb{R}^n$?

  ↑ that gives us a measure on $\mathbb{R}^n$.
  It's 1 on compact rectangles. That's very nice.
  Except it's only defined on sets that are finite unions of rectangles (parallelotopes).
  So it's only defined on sets that look like:



Not on all polyconvex sets.

So, if you define the Euler characteristic as the product $\mu_0 \times \mu_0 \times \cdots \times \mu_0$, then you are confronted with the problem of extending it.
This is not nice,
Whereas, we define it in another way, bypassing this crisis.

- Similarly, the volume could have been defined as:

$$\mu_1 \times \mu_1 \times \dots \times \mu_1$$

Then you get the volume of all parallelotopes.
And then we have to <u>extend</u>, <u>via a limiting process</u>, to <u>all polyconvex sets</u>.

- Nonetheless, this idea of taking product measures will guide us to discover what the other measures are.
We showed that all the invariant measures on $\mathbb{R}^1$ are linear combinations of μ_0 and μ_1: [32.5-7]

$$\mu(A) = \mu_0(A) + c\mu_1(A)$$

μ_0 = Euler characteristic μ_1 = length

Let's take:

$$(\mu_0 + t\mu_1) * (\mu_0 + t\mu_1) * \dots * (\mu_0 + t\mu_1) = \mu_t \text{ on } \mathbb{R}^n$$

{ t is a <u>parameter</u>.
What's a parameter?
A <u>parameter</u> is a <u>variable constant</u>. }

{ That's a measure.
It's a measure of parallelotopes and all their unions and intersections. }

μ_t is defined only on the lattice generated by all parallelotypes (unions, intersections).
All the sets have sides square, but they can have holes.
And they need not be convex.

What does $\mu_t(A_1 \times A_2 \times \dots \times A_n)$ look like, where A_i = closed interval in $\mathbb{R}^n$?

By definition, it's the product:

$$\mu_t(A_1 \times A_2 \times \dots \times A_n) = (\mu_0 + t\mu_1)(A_1) * (\mu_0 + t\mu_1)(A_2) * \dots * (\mu_0 + t\mu_1)(A_n)$$

One can work this out and obtain:

$$= \mu_0(A_1 \times A_2 \times \dots \times A_n)$$
$$+ t\sum_i \mu_1(A_i)$$
$$+ t^2 \sum_{i<j} \mu_1^2(A_i \times A_j)$$
$$+ \dots$$
$$+ t^n \mu_1^n (A_1 \times A_2 \times \dots \times A_n)$$

How do we interpret this result?
We have a polynomial in t for the measure $\mu_t(A_1 \times A_2 \times \dots \times A_n)$.
And each coefficient will be a measure in it's own right.
So, we rewrite $\mu_t(A_1 \times A_2 \times \dots \times A_n)$ as:

$$\begin{aligned}\mu_t(A_1 \times A_2 \times \dots \times A_n) = {} & \mu_0(A_1 \times A_2 \times \dots \times A_n) \\ & + t\mu_1(A_1 \times A_2 \times \dots \times A_n) \\ & + t^2\mu_2(A_1 \times A_2 \times \dots \times A_n) \\ & + \dots \\ & + t^n\mu_n(A_1 \times A_2 \times \dots \times A_n)\end{aligned}$$

We see that μ_0 is indeed the Euler characteristic.
And we see that μ_n is indeed the volume.
And, in between, we get these funny measures.
What are they?

Suppose that the sides of $A_1 \times A_2 \times \dots \times A_n$ equal $x_1, x_2, \dots, x_n$.
(i.e., $\mu(A_i) = x_i$, for psychological reasons)

Then:

$$\mu_1(A_1 \times A_2 \times \dots \times A_n) = \sum_i x_i \quad ,$$

$$\mu_2(A_1 \times A_2 \times \dots \times A_n) = \sum_{i<j} x_i x_j$$

Isn't this something familiar?
You get the elementary symmetric functions:

$$\mu_i(A_1 \times A_2 \times \dots \times A_n) = \underbrace{\sum_{j_1<j_2<\dots<j_r} x_{j_1} x_{j_2} \cdots x_{j_r}}_{\text{this is called: } e_i(x_1, x_2, \dots, x_n)}$$

So we see that the intermediate measures $(\mu_1, \mu_2, \dots, \mu_{n-1})$ evaluate on parallelograms from rectangles as the elementary symmetric functions.

And we have proved the following result:

Defining $\mu_i(A_1 \times A_2 \times \dots \times A_n) = \overbrace{e_i(x_1, x_2, \dots, x_n)}^{\text{elementary symmetric function}}$ gives a measure on the lattice generated by all "rectangles."

We obtain a well-defined measure on all of the lattice generated by all the rectangles.

Notice that the distance depends on the choice of particular coordinates of an orthogonal coordinate system.

In this way, we obtain $n+1$ measures:

$$\mu_0, \mu_1, \dots, \mu_n$$

The main theorem of geometric probability is that these measures can be uniquely extended to all polyconvex sets and they are the bases for all invariant measures.

Main Theorem of Geometric Probability

These measures have a unique extension to the lattice $\mathcal{L}$ of all polyconvex sets on $\mathbb{R}^n$, and every continuous invariant measure is a linear combination of them.

(remember how we defined continuous as limits on convex sets.)

The measures $\mu_0, \mu_1, \dots, \mu_n$ are called the intrinsic volumes.

This is one major result of mathematics.
There are exactly $n+1$ intrinsic volumes.
There are exactly $n+1$ numbers that you can associate to any body in space.
That's the only thing you can do.

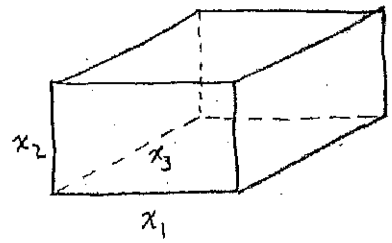
Any other number (i.e., measure) that you associate to a body in space is a linear combination of the $n+1$ intrinsic volumes.

I'll tell you in a minute about the extension.
The main point is that the extension can not be carried out by limiting procedures.
You can't use calculus, take limits. No. No. It doesn't work.
You need a diabolical trick to carry out the extension.

The limiting procedure works only for the volume.
That's course 18.02 (Calculus).
For these other intrinsic volumes, to go from these parallelograms and their unions to polyconvex sets — ah, you can't do it by limits.
Nobody has been able to do it, even for the Euler characteristic μ_0.
So, another trick was invented, of a completely different nature.

Example $\mathbb{R}^3$

Let's see what happens in 3 dimensions. Let's take parallelogram P:



P

Euler characteristic: $\mu_0(P) = 1$, if P non-empty

$\mu_1(P) = x_1 + x_2 + x_3$ ← [except for a normalization factor, it's the <u>perimeter</u>.]

$\mu_2(P) = x_1 x_2 + x_1 x_3 + x_2 x_3$ ← [except for a constant number, it's the <u>area</u>. If you multiply this by 2, it's the area. Except for a normalization factor, it's the area.]

Volume : $\mu_3(P) = x_1 x_2 x_3$

The main theorem tells you that you can <u>extend</u> these intrinsic volumes <u>uniquely</u> to <u>every</u> <u>polyconvex</u> set.

For example, take a convex set.
Take a potato.

A potato has a <u>volume</u>.
It has an <u>Euler characteristic</u>, which is 1.
It has an <u>area</u>.
And it has a <u>length</u> (no one would ever know that every potato has a length).

{ This is an important fact, my friends.
Potatoes have a length.
And as soon as physicists discovered this, they would not fail to have attached laws of physics to them.
Except they don't know this measure exists, because we didn't tell them.
Why didn't we tell them?
Because we are stupid.
Mathematicians are stupid. }

The measure μ_1, when extended, is called the mean width.

{a completely new thing we have no feeling for, since we've never seen it before.}

For objects in 3 dimensions, there are 4 basic invariant numbers that you can associate.
The Euler characteristic, the volume, the area, and the mean width.

This is a basic fact of life.

Now, roll up your sleeves my friends.
Because now we have to prove the extension.
How do we extend these measures to all polyconvex sets?
Forget about limits.
We have to approach this from a completely disparate point of view.

Now I say "I like lattices."
And we've talked a lot about the lattices of subsets of a finite set, the Boolean algebra.

What's the next best lattice?
Let's take the lattice of all subspaces (through the origin) of a vector space over the real numbers.

Let $L(V)$ = lattice of all subspaces (through the origin) of a vector space V over $\mathbb{R}$.

This is not a distributive lattice, as we've seen many times.

However, if you take an orthonormal basis, it has a certain orthonormal property.
Namely, if you take a subspace W:

$$W \in L(V)$$

You can associate the orthogonal complement: $W^{\perp}$ [16,7]

Then you have:

$$\text{if } W' \subseteq W \text{ then } W = W' \vee (W \wedge W'^{\perp})$$

This is the closest you come to the distributive law.

Exercise 34.2

Prove that the above implication is true.

The point being that we would like to do computations on $L(V)$ like we do with subsets.
Permutations, chains, Dilworth's Theorem – stuff like that.
Except that $L(V)$ is <u>continuous</u>.
So we have to use <u>measures</u>.

No problem.
Let's use measures.
Remember the computation we did with $P(S)$; the lattice of subsets of a finite set, Boolean algebra?
We counted the complete chains. [22.4]
How do you count the complete chains.

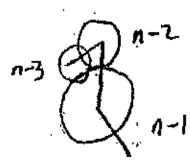
You take a point in S, which, if S has n elements, you take n ways.
Then you are one step up. You're in a Boolean algebra of subsets of $n-1$; so you can pick any one of $n-1$.
Then you go another step up.
Etc.
This gives you the number of complete chains:

$$n(n-1)(n-2)\cdots 1 = n!$$

number of complete chains

Now we can do the same for <u>$L(V)$</u>.
Start with 0.
Pick a line. How many ways can you pick a line?
You take the measure on the surface of the $n-1$ spheres and that gives you the number of lines.
Then you are in a subspace of $n-1$ dimensions.
Pick a line in $n-2$ spheres. Take the measure on the surface of the $n-2$ spheres and that gives you the number of lines.

So you have all these measures on the surfaces of all these spheres.
And you multiply these measures and that gives you a measure of the set of all chains in $L(V)$.



That's the <u>continuous analogue</u> of <u>$n!$</u>.
That we'll leave for next time.

In this way, we get a measure on the set of all chains that is invariant on the orthogonal group, of course.
And having obtained the measure on the set of all chains, we take the set of all subspaces of dimension k (that's called a <u>Grassmannian</u>) and the <u>measure</u> on the <u>set of all chains</u> we immediately view as an <u>invariant measure</u> on the <u>Grassmannian.</u>

If you have a subspace of dimension k, you take all the complete chains going through that set.

These people in differential geometry - Ooh.
There's a simple combinatorial way to do it.

And having done that, we have a measure on the set of all subspaces.
And using this, we will get the extension of the measure to all polyconvex sets.

We only have one more lecture, unfortunately.
We'll have to compress everything into one lecture.

Geometric Probability : the Kinematic Formula

We saw last time that:

In $\mathbb{R}^n$, we have $\mathcal{L}$ = lattice of polyconvex sets.

Lattice of polyhedra $\mathcal{L}_p \subseteq \mathcal{L}$

$\mathcal{L}_0 \subseteq \mathcal{L}_p$

$\mathcal{L}_0 \subseteq \mathcal{L}_p \subseteq \mathcal{L}$

Lattice of rectangles, then you generalize rectangles when you take an orthogonal coordinate system and you take boxes and all possible unions and intersections of boxes.
Let's call this "lattice of boxes."
There's no really standard word for this.

$\mathcal{L}_0$ = lattice generated by boxes relative to an orthogonal coordinate system $x_1, x_2, \dots, x_n$.

On $\mathcal{L}_0$, one can define n+1 measures, called the intrinsic volumes, by setting:

$$\mu_k(P) = e_k(x_1, x_2, \dots, x_n) \quad , \quad 0 \le k \le n$$

P = box with sides equal to $x_1, x_2, \dots, x_n$.

e_k = elementary symmetric function
Remember that the elementary symmetric function of order k=0 is 1, i.e., $e_0 = 1$.

$$e_k(x_1, x_2, \dots, x_n) = \sum_{1 \le i_1 < i_2 < \dots \le n}^{n} x_{i_1} x_{i_2} \cdots x_{i_k} \quad , \quad 1 \le k \le n$$

So, in n dimensions, you have n+1 measures, in this way.

And then the main theorem has two major observations:

(1) these measures can be extended, not only to the lattice of polyhedra $\mathcal{L}_p$, but to the lattice of polyconvex sets $\mathcal{L}$.

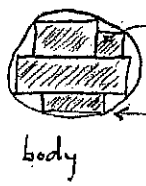
(2) these are invariant measures on the lattice of polyconvex sets, continuous in the sense previously determined.
And, furthermore, every continuous invariant measure is a linear combination of these measures.

Since I have only 1 hour left, I have to do this slightly by handwaving.
This is all in books.
I have to give you the ideas.
For details, you have to read the books.

- Let's talk about the extension.

I told you last time that the extension can not be carried out by ordinary limiting processes.

A limiting process is something like this.
You have a body and you inscribe in it a rectangle.
Then you add smaller rectangles, as you take the limit.



You are forced to add rectangles in a general direction.

And you get all these angles, which are kinky.
It doesn't work.

Even for the Euler characteristic it doesn't work.

So, for limiting processes of the ordinary course 18.02 (Calculus) kind – forget it.
The limiting process that works is much fancier.
Here, I have to start handwaving.

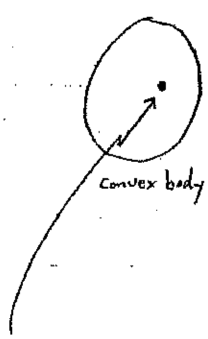
Let's take 3 dimensions. I'll tell you what the limiting properties will be.

- Example – $\mathbb{R}^3$

First, we extend the measures (intrinsic volumes) to convex bodies.
Then, by a technical trick, we extend to polyconvex sets.

↑ finite unions of compact, convex sets.

After you have defined a convex body, by use of inclusion-exclusion, you can define a polyconvex set.
So a big part of this extension is the convex body.

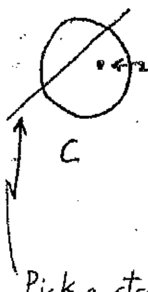


In $\mathbb{R}^3$, it is possible to choose a point at random.
Strictly speaking, this does not make sense.
Because probability is not defined in $\mathbb{R}^3$.
So, by an abuse of speech, I will talk about probability while I mean conditional probability.

You must condition over a big containing ball.

But, by abuse of speech, we'll talk about probability.

So, the probability that I pick a point belonging to a convex body is obviously equal to the volume of the convex body.



Pick a point at random in convex body C.
The probability is:

$$\mu_3(C) \leftarrow \text{volume}$$

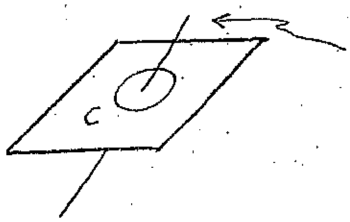
Pick a straight line in convex body C at random.
It isn't very clear that it is possible to pick a straight line at random.
If you want to make this precise, it's very deep.
Because it means there is an invariant measure on the set of straight lines in space.
It's the same thing.
Being able to pick a straight line at random in space means that there is an invariant measure of the set of straight lines — when you write this in correct, grammatical terms.

So, let's assume we can pick a straight line at random.
Then compute the probability that the line meets C.

I say that that probability is the area of C:

$$2\mu_2(C)$$

Why?
Let's take C like this:

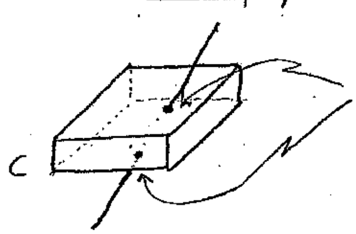


The probability that you pick a straight line at random that meets C, when C is this, is proportional to this area whenever C is flat.
Why?
By Cauchy, because a line meets a flat rectangle or even a flat set in the rectangle either in one point, or not at all.

And therefore, by Cauchy's functional equation, you get this probability proportional to the area.

Therefore, the measure of the set of all lines meeting a given 2 dimensional surface is proportional to the area of the 2 dimensional surface.

But for a convex polyhedron in 3 dimensions, a line meets it twice.



line meets C twice.

So, mirror to a limiting process, when you make C round, you get twice the meetings. So the probability is:

$$2\mu_2(C)$$

I have to cut corners to cover the material today,
All this, written down, is called the integral of invariant measures.

Now, pick a plane at random.
The probability is:

$\mu_1(C) \leftarrow$ mean width

You assume you can't pick planes at random in this space.
This is intuitively clear, but you have to compute the invariant measure of the set of all planes (on the Grassmannian of planes).

- In this way, you prove that if you have a box C:

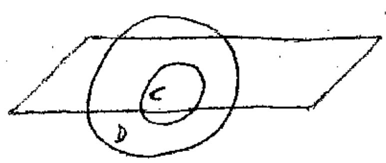
the measure of the set of all points into the box $= \mu_3(C)$
" " " lines " " $= 2\mu_2(C)$
" " " planes " " $= \mu_1(C)$

Therefore, since it agrees with all boxes, then automatically this construction extends it to all convex sets.

Therefore, you redefine the intrinsic volumes as the measures of sets of all points, lines, planes into convex sets.

- So, in this way we have a precise, intuitive interpretation of the mean width.
Take two compact, convex sets — one inside the other:

$C \subseteq D$, C and D both compact, convex



The probability that a random plane meets C, given that it meets D, is the ratio:

$$\underbrace{\frac{\mu_1(C)}{\mu_1(D)}}_{\text{you can find this number, experimentally.}} \quad \begin{array}{l} \leftarrow \text{mean width of } C \\ \leftarrow \text{mean width of } D \end{array}$$

You see that the normalization factors cancel, anyway.

This is an extraordinary result.

Where did this come from?
It's the Buffon Needle Problem.
This result is equivalent to the Buffon Needle Problem.

What's the Buffon Needle Problem?
You drop a needle on a plane with parallel straight lines.
You generalize this and these sets arise.

So, in this way, you extend the intrinsic volumes to all convex sets.
Then, for polyconvex sets, you use inclusion – exclusion to extend from convex sets to polyconvex sets.

{ you really have to prove this.
I'll show you how to do it later. }

In this way, you get $n+1$ invariant measures on $\mathcal{L}$.

Then, the main theorem is that there are no others. ⟸ This is a very deep result.

{ This theorem depends on a lemma, whose proof is extremely difficult.
It sounds obvious, but when you try to prove it, it's a mess.
I'll tell you what it is so you can try proving it.
I hope you do try to prove it, because I'm sure there is a simple proof.
I can't get it. I'm just too old.
There is a simple proof. }

Let me tell you what the crucial lemma is:

Remember what it means for a measure on polyconvex sets to be continuous. [32.3]
Let's write this down, because it's very important to have the right definition.

A measure μ on $\mathcal{L}$ is continuous when $\mu(C_n) \longrightarrow \mu(C)$ whenever C_n is a sequence of compact, convex sets converging to C.

Crucial Lemma

Let μ be a continuous, invariant measure on polyconvex sets.

This is not a countably additive measure. It's a finitely additive measure.
Our measures are not countably additive, otherwise we'd have the volume.

Assume that this measure vanishes on lower dimensional polyconvex sets.
In other words, if you have a polyconvex set that is contained in a lower dimensional hyperplane, the μ of that set is 0.
If the set is "thin" then $\mu = 0$.

Lemma

Assume $\mu(C)=0$ if C is contained in a proper hyperplane.
Then $\underbrace{\mu = c\cdot\mu_n}$. ↑ which means it's a lower dimension

μ is a constant times the volume μ_n

This is what people have not been able to prove easily.
If you prove it, I will send your paper to the Proceedings of the National Academy of Science and it will be published.
I am sure there is a simple proof.
The first proof has 137 pages.
The second proof has 32 pages. It's an improvement of the first proof by Dan Klain.
Some day, someone will get a 2 page proof.
It is written.
I don't see how to do it.
Please. What are you doing now?
Come on. Help me out. Help me cut out 30 pages out of my book.

**** Exercise 35.1**

Find a simple proof of the Crucial Lemma.

This is a good research problem for the vacation.
When you come back, you say how you spent your vacation.
Prove this.
It's a nice puzzle.

Remember, this is not countably additive, so you can't cut it into infinitely many pieces.

This Crucial Lemma is the one that gives uniqueness of the intrinsic volumes.

**** Exercise 35.2**

So now, you have the intrinsic volumes defined for all polyconvex sets.
In particular, you can take the analogue of the tetrahedron in n dimensions – the n-simplex.
Take n+1 points and take the convex hull.
Then you can ask: What are the intrinsic volumes of an n-simplex?
The answer is not known.
This is an open problem:

Compute the intrinsic volumes of ~~the~~ an n-simplex.

There must be formulas for area, perimeter, etc.
But they aren't known.
This is a backward field. An undeveloped subject.

I don't think this is particularly hard. It's just that nobody has done it.

We know very little about angles in n-dimensions.
It's an undeveloped field.
These formulas for the intrinsic volumes depend on our understanding of angles in n dimensions.
We don't know.

The analogue of trigonometry in n dimensions — nobody has worked it out.

** Exercise 35.3

Here's another open problem.
We have that the lattice of polyhedra $\mathcal{L}_p$ is a subset of the lattice of polyconvex sets $\mathcal{L}$.

$$\mathcal{L}_p \subseteq \mathcal{L}$$

n+1 invariant measures?

On $\mathcal{L}$, the uniqueness theorem tells us there are exactly n+1 invariant measures. The space of invariant measures is n+1.

In particular, the intrinsic volumes are defined on $\mathcal{L}_p$.
We extended measures to the lattice of polyhedra $\mathcal{L}_p$ and then to the lattice of polyconvex sets $\mathcal{L}$.

But no one has proved that the n+1 intrinsic volumes are unique on the lattice of polyhedra $\mathcal{L}_p$.
There may be more on $\mathcal{L}_p$.

Uniqueness? ← Prove whether these n+1 intrinsic volumes are unique on $\mathcal{L}_p$.

It is possible that there may be some extra invariant measures on polyhedra that are not extendable to polyconvex sets.

Perhaps there are weird points, like Steiner points, for which this is the case.

*** Exercise 35.4

Instead of taking $\mathbb{R}^n$, we take the surface of a sphere.
You can define compact, convex sets on a sphere.
So you can define polyconvex sets on spheres.
And you can define measures, invariant under rotations of the spheres.
And you can ask how many there are. ← Nobody knows.

Open problem. This is a Ph.D. thesis.

Work out the intrinsic volumes on spheres.

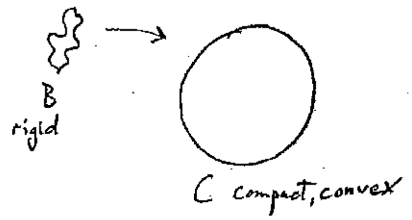
This is solved only for the 2 dimensional sphere.
It's also been solved for the 3 dimensional sphere.
For the 3 dimensional sphere, you can take the boundary of a 4 dimensional ball.

This is a backward field.
Sorry.

Kinematic Formula

Again, I have to do some handwaving, because I don't have time.

I take a compact, convex set C.
Then I take a "bad" object B, which is rigid, of dimension n-k.



I drop B on C at random.
What's the probability that B meets C?

it looks hard, but it isn't.
Why?

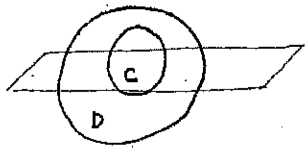
Because the probability that B meets C is an invariant measure.
It's an invariant measure that depends only on B and C.

therefore, it's a linear combination of intrinsic volumes.
Ha. Ha.
And, therefore, what you need are the coefficients of this linear combination. Which you get by varying C while keeping B fixed.
That's how you solve this.
That's the kinematic formula.

So, the uniqueness of the intrinsic volumes allows you to immediately infer that if you drop any B, of any shape whatsoever, on a compact, compact set C, the probability is a linear combination of intrinsic volumes.

That's how all these geometric probability problems are solved.

Assume $C \subseteq D$.



$$\frac{\mu_k(C)}{\mu_k(D)} = \text{probability that an } (n-k) \text{ dimensional flat meets } C, \text{ \underline{given} that it meets } D.$$

that's a genuine probability.
Again, generalizing the Buffon Needle Problem.
It involves linear combinations of intrinsic volumes and all that stuff.

And that's all.
That's all we know about geometric probability.
That's it.

It would take 3-4 lectures to write down all the details.
You can read it in my book.

Who is taking 18.315?
Roll call. 22 people.

I'm really sorry I covered so little material this term.
I really apologize.
I hope you're not disappointed.
I hope next year to cover a little more material.
I promise next year will be completely disjoint from this year.
Nothing will be common.
It will look like another world.
The only common thing is that it will be given by the same person.
So the style is the same.
The same wishy washy style.

So, I hope you solve some of the problems I stated this term.
It would please me a lot if some of you solved any two star or three star problems.

None of them are hard.
If I were given one million dollars, I would solve all of them.

We still have time.
Let's do a little more.

Why is it that we can pick a line at random?

There 3 ways of doing this:

(1) One of them is the way differential geometers look at this.
The space of lines is a homogeneous basis of a Lie group. Geometers condition A, B, and C as a unique invariant measure. That's it.
That's approach number 1.

(2) Approach number 2 is the most naive, which leads to yet another unsolved problem.
You consider the space of lines as a big space, where a point is a line.
That's called a Grassmannian and you have these algebraic varieties that satisfy certain algebraic equations, which we have seen.

We are talking about lines in 3 space to fix our ideals,
So that works for k dimensional subspaces in n dimensional space.

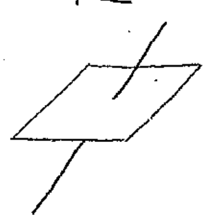
Let's talk about any lines, not necessarily through the origin.
We want lines in space.
How do we define a measure?

First you define a measure on easy sets.
Then you take the unions and intersections of easy sets to be the hard sets.
Then you extend the measure to the hard sets.
That's the way all measures are defined.

Since you always define a measure on the space of lines in 3 dimensional space, you cleverly choose the easy sets. How do you choose the easy sets.
Like this:

In $\mathbb{R}^3$, consider the Grassmannian G_1^3. $\longleftarrow$ the set of all lines in 3 space.

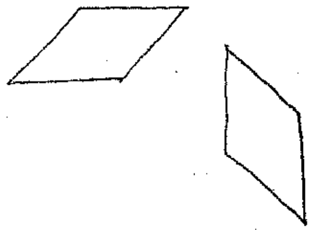
The easy sets are the set of all lines that meet a given 2 dimensional surface.



Because, by the argument I have already outlined,
the set of all lines that meets a given 2 dimensional surface
is proportional to the area of that surface.
And the line meets the surface at one point, or not at all.

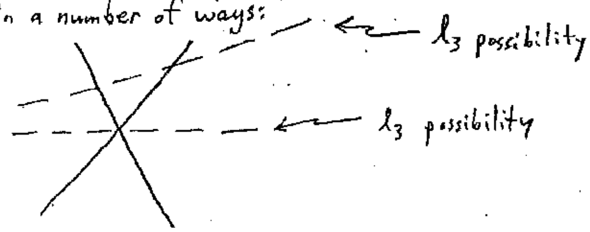
Therefore, you can immediately tell the measures of certain sets of lines,
Namely, the set of all lines that meet a given surface.
Those are the easy sets.
Then you have to extend.

But, here the extension is not so easy, because if you have two of these surfaces:



the set of all lines that meets both these surfaces is not obtained by inclusion-exclusion.

For example, in the plane, if you have 2 lines as below, a third line l_3 can meet both lines in a number of ways:



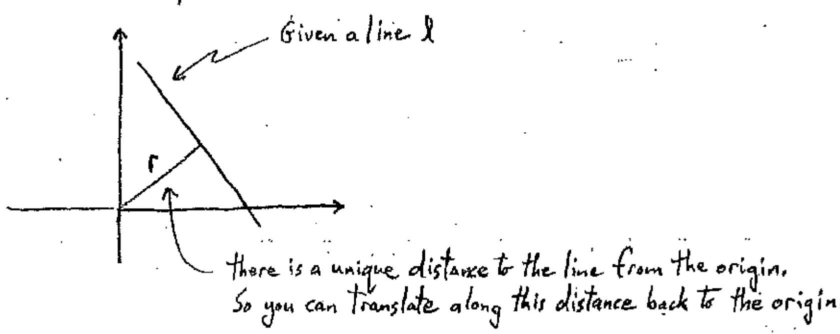
It's not clear how to get the inclusion-exclusion working, because we have the geometric condition working.

So, the extension can be carried out.
But what we do not know, i.e., the open problem, is the analogue of inclusion-exclusion of these easy sets.
We do not know.
What are the algebraic relations holding with all the indicator functions of these sets of lines.

What you do is that you do the integral instead, when the integral can be written.
And then, of course, you specialize.

(3) The third approach is the one I outlined last time. [34.10-11]
You split the problem into two.

You want to give a measure to a set of 3 dimensional subspaces of that space - the set of lines in 3 dimensional space.



So you can get any line ℓ by taking any line through the origin, and then moving it.
That means the product of invariant measure is:

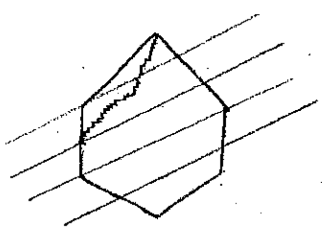
distance × invariant measure of the line through the origin

[So the problem reduces to computing the invariant measures of the set of lines through the origin.
This is semi direct products.
Because the Euclidean group is the semidirect product of the orthogonal group and the translation group.]

So now, how do we find the invariant measures of the set of lines through the origin?
For this, we use the method we used last time.

Take $L(V)$ = lattice of all subspaces (through the origin) of a vector space V over $\mathbb{R}$.

You visualize this lattice.
It's a set of lines of dimension 1 if you have a plane.
It's a measure of all elements at level 1.



So what you do is take the measure of the set of all complete chains.
And you get the measure on the set of lines (through the origin) by taking the measure of all complete chains passing through this set, divide by the number of all complete chains going up and divide by the number of all complete chains going down:

$$\frac{n!}{k!\,(n-k)!}$$

In the non continuous case, that's called the binomial coefficient. [22.4]

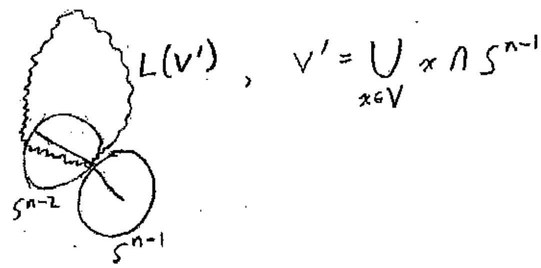
Now we do the same for the continuous case.
The measure on the Grassmannian is like the binomial coefficient.

All you need is a measure on the set of complete chains.
And this measure on the set of complete chains delivers the desired measure on the set of lines through the origin.

How do you get the measure of the set of complete chains?

You pick a direction on the sphere S^{n-1}.
And divide by 2 because the same line has two directions.

This leaves the lattice of subspaces of the vectorspace V of dimension n-1,
Pick a line on the sphere S^{n-2}.



And the measure of the set of complete chains is the product of all the dimensions.

It's all really trivial.
It's in my book.

The key thing is to reduce the problem of invariant measures on Grassmanians to invariant measures on sets of lines through the origin.
And then, to imitate the combinatorial way of defining binomial coefficients.

As you see in my book, we try to get continuous analogues of the facts about binomial coefficients, using these continuous binomial coefficients.
One thing we couldn't get. I won't let you down, but it's an open problem.
Namely:

** Exercise 35.5

What's the continuous analogue, using continuous binomial coefficients, of the binomial theorem?

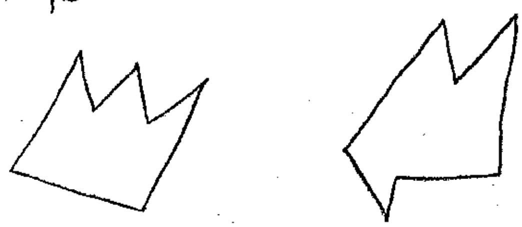
You have to read about the flag coefficients and so on.
In this way, we get continuous analogues of continuous binomial coefficients.
But these are products of volumes of spheres of various dimensions.
And you these volumes are defined in terms of the Euler gamma function.

**** Exercise 35.6

A really hard problem is this:

Given 2 polyhedra in n dimensions.

Example:



P_1 P_2

{The countably case was solved by Tarski.}

When can you cut up the first polyhedron P_1 into a finite number of polyhedra, which can then be used to construct the second polyhedron P_2?

In 2 dimensions, this was solved by Hilbert.
He proved that when you have 2 polygons with the same area, then you can cut up the first polygon into a finite number of triangles and recompose the second polygon.
This is the famous theorem of Hilbert. This is Hilbert's Third problem, which was then proved 2 years later.

But in more than 2 dimensions, nobody knows the necessary and sufficient conditions.
The conjecture is that it should be related to certain things about intrinsic volumes.
It's not enough for the intrinsic volumes to be the same.

(I'm sorry to say.
Two bodies may have exactly the same intrinsic volumes,
but you may not be able to cut up one and construct the other.)

This problem was solved about 15 years ago by Sah, if you allow only translations.
In other words, if you cut up pieces and you can not rotate them, but you can translate them, when you recompose them.
If you allow only translations, then this was solved after tremendous effort.
And there are generalizations with intrinsic volumes.

This is what makes you suspect that there are other invariant measures involved under the group.
This is a Field's Medal problem.

This is a field that is very rich.
What happened is that classical geometry had been neglected in this century,
Now we go back to classical geometry because of the needs of computer graphics
Because of computer graphics, we are asking all these problems.
And we discover that we don't know anything.
We know everything about abstract algebraic structures and varieties, but we don't know any combinatorial geometry.
So I hope you work on this stuff.

Now,

— That's the End —

Note: within the body of the text, pagination is of the form [*lecture.page*]. For example, page [3.5] refers to the fifth page of the third lecture, which was given on September 14, 1998.

Index

Typos: 4, 160, 82, 256