Applications: Polyhedra, here and there, polyhedra everywhere!!

Jesús Antonio De Loera
University of California, Davis

?	?	?	?	220
?	?	?	?	215
?	?	?	?	93
?	?	?	?	64
108	286	71	127	

Combinatorics

- Many discrete structures can be counted this way: e.g. matchings on graphs, Hamiltonian cycles, t-designs, linear extensions of posets,

- Gelf'and-Tsetlin polytopes (1950's)

  $K_{\lambda\mu}$ = number of semi-standard tableaux of shape λ and content μ, *Kostka numbers*.

Two semi−standard tableaux shape (3,2) and content (2,2,1)

1	1	3
2	2	

1	1	2
2	3	

flows on Networks

Let G be a network with n nodes and m arcs, with integer-valued capacity and excess functions $c : arcs(G) \to \mathbf{Z}_{\geq 0}$ and $b : nodes(G) \to \mathbf{Z}$.

A *flow* is a function $f : arcs(G) \to \mathbf{Z}_{\geq 0}$ so that, for any node x, the sum of flow values in outgoing arcs minus the sum of values in incoming arcs equals $b(x)$, and $0 \leq f(i,j) \leq c(i,j)$.

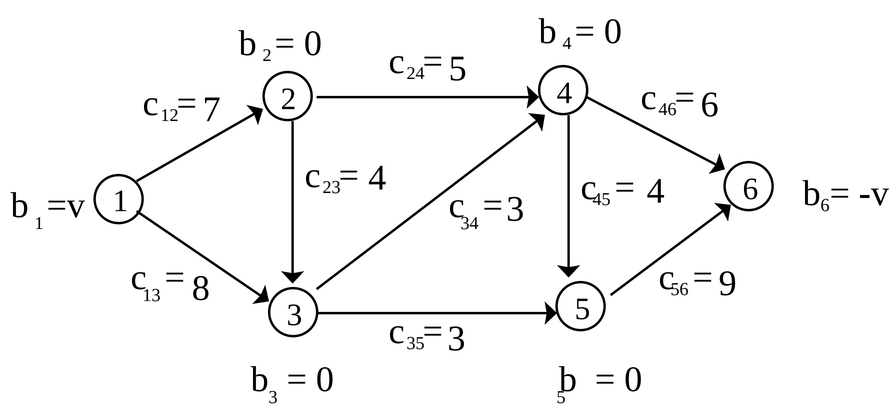


Figure 1: A simple example

How many Max-Flows are there?

From well-known theorems the max-flow value is 11, but how many max-flows are there?

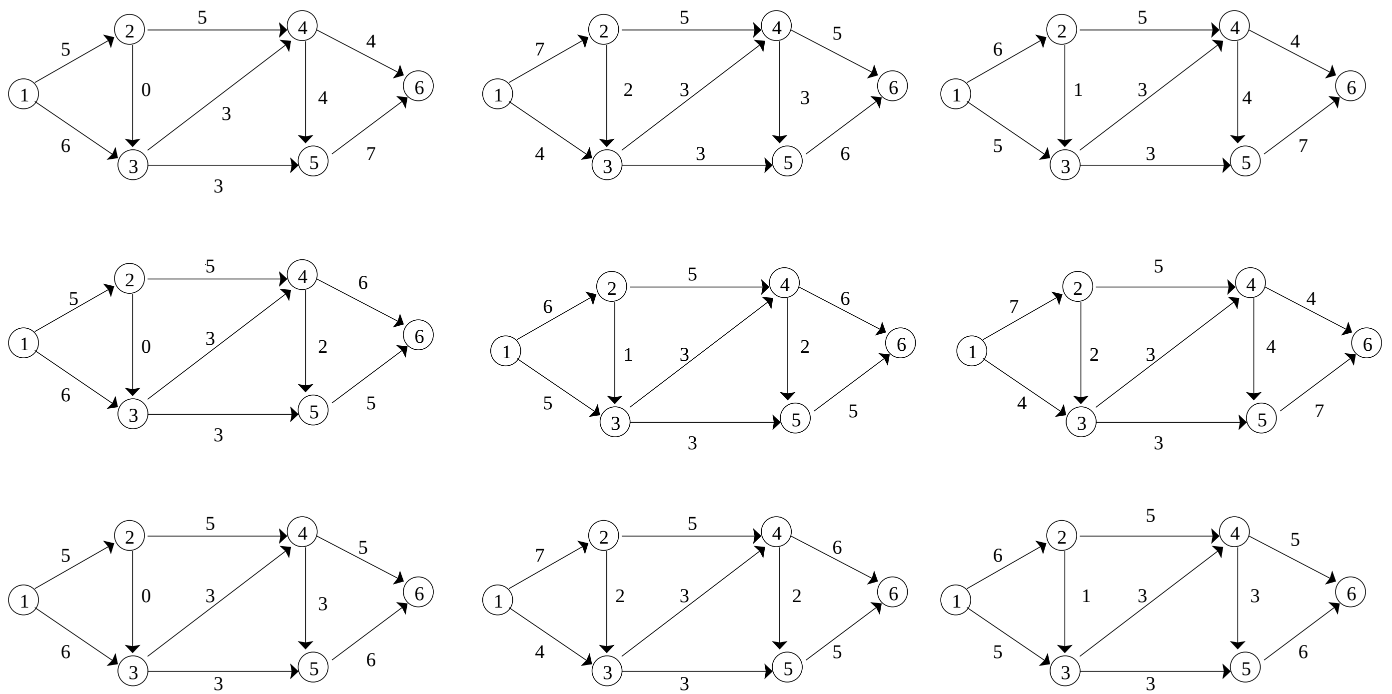
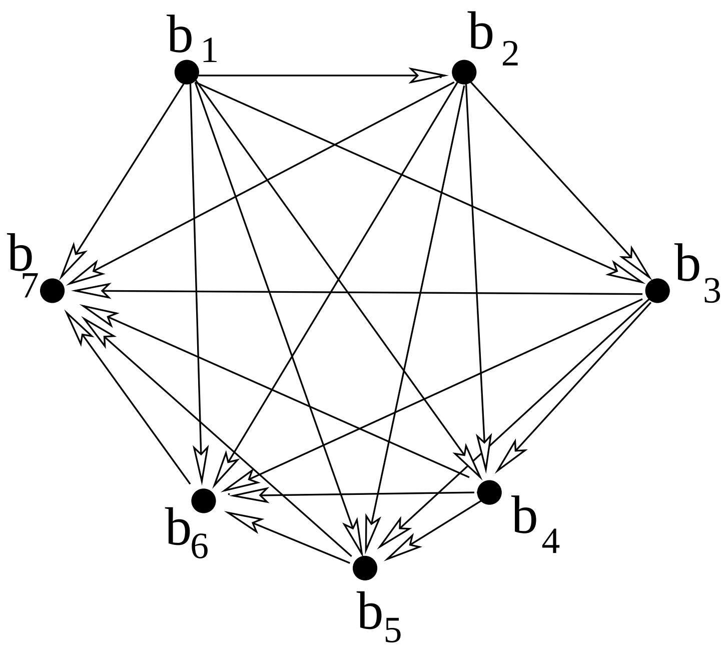


Figure 2: All max flows in the network.

Flows on Graphs

The matrix A is node-arc $\{-1,0,1\}$-incidence matrix of the acyclic complete graph K_n. HOW MANY INTEGRAL FLOWS ON THIS NETWORK??



$b_1 = 21128$
$b_2 = 45716$
$b_3 = 79394$
$b_4 = -76028$
$b_5 = -31176$
$b_6 = 66462$
$b_7 = -105496$

b_i = sum of values coming IN − sum of values going OUT

t-Designs

- Let $v \geq k \geq t$ be fixed positive integers. Given the set $[v] = \{1, \ldots v\}$.

- A $t-(v, k, \lambda)$ **design** D is a finite collection of possibly repeated k-subsets of $[v]$, the **blocks** of D, such that any t-subset of $[v]$ is contained in exactly λ blocks.

- The constant λ is called its **incidence number**. A t-design is **simple** if and only if the blocks are not repeated. When $\lambda = 1$ the t-design is called a **Steiner system.**

- A finite **projective plane** of order q, with the lines as blocks, is an 2-$(q^2 + q + 1, q + 1, 1)$ design, since it has $v = q^2 + q + 1$ points, each line passes through $k = q + 1$ points, and each pair of distinct points $t = 2$ lies on exactly one line.

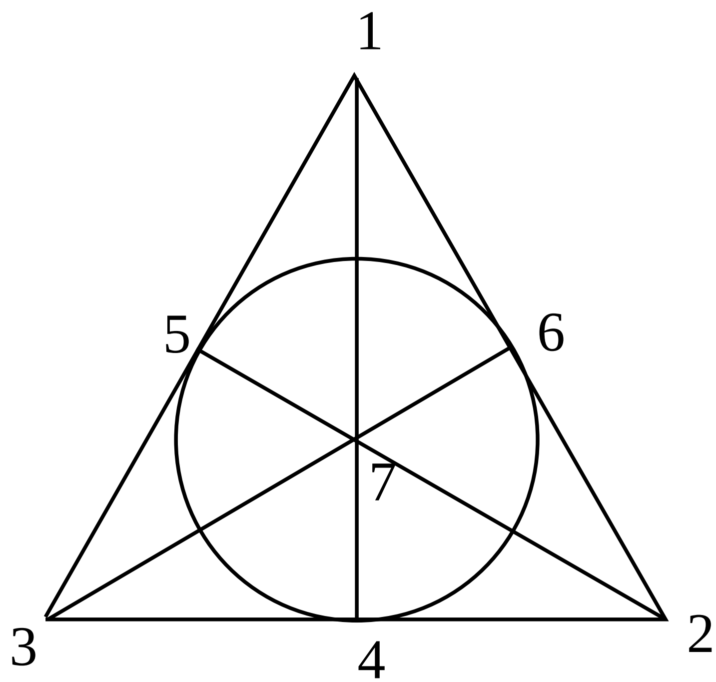


Figure 3: The Fano plane: A $2-(7,3,1)$

- We can use polyhedra to investigate t-designs!!

- Think of $t-(v,k,\lambda)$ designs as the solutions of a system of linear Diophantine equations $M_{t,v,k}x = \lambda(1,\ldots,1)^T$.

- Here $M_{t,v,k} = (m_{i,j})$ is an $\binom{v}{t} \times \binom{v}{k}$ matrix with $m_{i,j} = 1$ exactly when the i-th t-subset is contained in the j-th k-subset and otherwise it is zero.

- When $\lambda = 1$ the integer hull of these polytopes has been studied under the name of **design polytopes** by Lucia Moura.

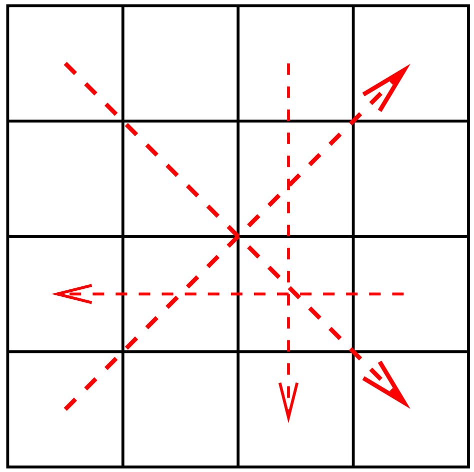
- **EXAMPLE:** (Fano Plane parameters): The polytope has 30 vertices dimension 14, and 155 facets. Each vertex (a design) is contained in 84 facets. Its graph is the complete graph!

- Hard to find t-designs!!! We do not know whether there is a projective plane of order 12 nor whether there is a 2-(46,6,1) design.

- **Theorem:** For fix values $v \geq k \geq t$ the number of distinct $t-(v,k,\lambda)$ designs can be written as a univariate quasi-polynomial in λ of degree $\binom{v}{k} - \binom{v}{t}$ with a positive leading coefficient. Similar result holds true for t-designs with prescribed automorphism group.

- **Corollary** (Wilson's theorem) Given $t \leq k \leq v$, there exists $\lambda_0 > 0$ Such that for any $\lambda \geq \lambda_0$ there exists a $t-(v,k,\lambda)$-design (with possibly repeated blocks).

More on Magic Squares



12	0	5	7
0	12	7	5
7	5	0	12
5	7	12	0

5

QUESTION: **HOW MANY 5×5 magic squares with sum n are there?**

LattE can solve this in minutes!!

Theorem The generating function number of 5x5 magic squares with magic sum n:

Numerator and Denominator:

$$-(t^{76} + 28t^{75} + 639t^{74} + 11050t^{73} + 136266t^{72} + 1255833t^{71} + 9120009t^{70} + 54389347t^{69} + 274778754t^{68} + 1204206107t^{67} + 4663304831t^{66} + 16193751710t^{65} + 51030919095t^{64} + 147368813970t^{63} + 393197605792t^{62} + 975980866856t^{61} + 2266977091533t^{60} + 4952467350549t^{59} + 10220353765317t^{58} + 20000425620982t^{57} + 37238997469701t^{56} + 66164771134709t^{55} + 112476891429452t^{54} + 183365550921732t^{53} + 287269293973236t^{52} + 433289919534912t^{51} + 630230390692834t^{50} + 885291593024017t^{49} + 1202550133880678t^{48} + 1581424159799051t^{47} + 2015395674628040t^{46} + 2491275358809867t^{45} + 2989255690350053t^{44} + 3483898479782320t^{43} + 3946056312532923t^{42} + 4345559454316341t^{41} + 4654344257066635t^{40} + 4849590327731195t^{39} + 4916398325176454t^{38} + 4849590327731195t^{37} + 4654344257066635t^{36} + 4345559454316341t^{35} + 3946056312532923t^{34} + 3483898479782320t^{33} + 2989255690350053t^{32} + 2491275358809867t^{31} + 2015395674628040t^{30} + 1581424159799051t^{29} + 1202550133880678t^{28} + 885291593024017t^{27} + 630230390692834t^{26} + 433289919534912t^{25} + 287269293973236t^{24} + 183365550921732t^{23} + 112476891429452t^{22} + 66164771134709t^{21} + 37238997469701t^{20} + 20000425620982t^{19} + 10220353765317t^{18} + 4952467350549t^{17} + 2266977091533t^{16} + 975980866856t^{15} + 393197605792t^{14} + 147368813970t^{13} + 51030919095t^{12} + 16193751710t^{11} + 4663304831t^{10} + 1204206107t^{9} + 274778754t^{8} + 54389347t^{7} + 9120009t^{6} + 1255833t^{5} + 136266t^{4} + 11050t^{3} + 639t^{2} + 28t + 1).$$

$$(t^2-1)^6(t^2+t+1)^7(t^6+t^5+t^4+t^3+t^2+t+1)^2(t^6+t^3+1)(t^4+t^3+t^2+t+1)^4(t-1)^9(t+1)^4(t^2+1)^4.$$

Open Question Can you get a formula for the $6x6$ magic matrices?

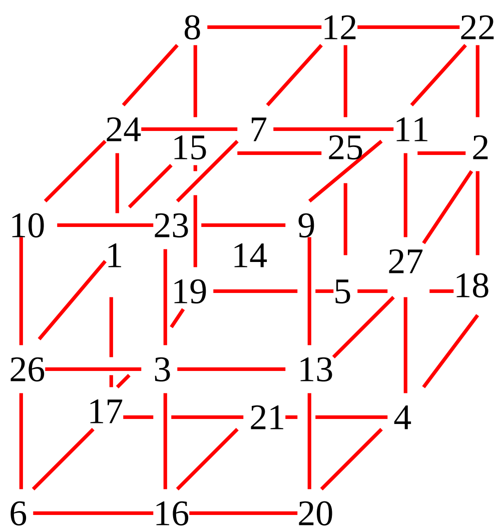
Statistics

Suppose someone takes a poll, asking 3 questions with 3 answers each:

Favorite Color (e.g red, blue, yellow),

Favorite Political party (e.g democrat, republican, independent),

Favorite drink (e.g. lemonade, milk, champagne).



We use 82 descendants of Queen Victoria as our sample.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Jan	1	0	0	0	1	2	0	0	1	0	1	0	6
Feb	1	0	0	1	0	0	0	0	0	1	0	2	5
Mar	1	0	0	0	2	1	0	0	0	0	0	1	5
Apr	3	0	2	0	0	0	1	0	1	3	1	1	12
May	2	1	1	1	1	1	1	1	1	1	1	0	12
Jun	2	0	0	0	1	0	0	0	0	0	0	0	3
Jul	2	0	2	1	0	0	0	0	1	1	1	2	10
Aug	0	0	0	3	0	0	1	0	0	1	0	2	7
Sep	0	0	0	1	1	0	0	0	0	0	1	0	3
Oct	1	1	0	2	0	0	1	0	0	1	1	0	7
Nov	0	1	1	1	2	0	0	2	0	1	1	0	9
Dec	0	1	1	0	0	0	1	0	0	0	0	0	3
Total	13	4	7	10	8	4	5	3	4	9	7	8	82

Question: Is there any connection between birth month and death month?

Birth month vs Death month

- Is this table of data what we would expect assuming that birth month and death month are independent?
  - If 6/82 people were born in January, and 13/82 died in January, then we would expect $\frac{6}{82} \times \frac{13}{82} \times 82 = 0.95$ people were born and died in January.

- We can measure the distance of the sample table from the expected table using the χ-square value.

- If the sample table is **statistically significantly different** from the expected table, we conclude that there is a statistically significant relationship.
  - Statistical significance tells us how confidently we can generalize to a larger population from a sample of that population.

HOW MANY TABLES WITH GIVEN MARGINS?

164424	324745	127239
262784	601074	9369116
149654	7618489	1736281

163445	49395	403568
1151824	767866	8313284
1609500	6331023	1563901

184032	123585	269245
886393	6722333	935582
1854344	302366	9075926

Table 1: Marginals for $3 \times 3 \times 3$ example.

THE MAIN QUESTIONS

(1) TABLE COUNTING PROBLEM:

Given a prescribed collection of marginals,

How many INTEGRAL d-tables are there that share these marginals?

Can one at least estimate the number of such d-tables?

(2) ENTRY SIZE PROBLEM:

Given a collection of marginals coming from an actual (but unknown) integral d-table, an index tuple $(i_1, \ldots, i_d)$, L, M positive integers

Is there an integral d-table x having the given marginals whose entry $x_{i_1,\ldots,i_d}$ is greater than or equal to L and less than or equal to U ?

EASY for two-way tables...

WHY SHOULD STATISTICIANS CARE?

(1) **INDEPENDENCE HYPOTHESIS TESTING:** Tests for the significance of the hypothesis that row and column effects are independent in a $r \times c$ contingency table (work by Diaconis-Efron, Good, others).

(2) **FEASIBILITY PROBLEM:** Given a prescribed collection of marginals that *seem* to describe a d-table of size $(n_1, \ldots, n_d)$,

Does there really exist a table at all with these marginals and can it be effectively determined ?

(3) **LIMITATION OF DISCLOSURE.**

You have a table T (multi-dimensional perhaps) with statistics of private data about individuals. You wish to release marginals of such table without disclosing information about the exact entries of the table.

Gender = Male

	Income Level			
Race	$\leq \$10,000$	$> \$10000$ and $\leq \$25000$	$> \$25000$	Total
White	96	72	161	329
Black	10	7	6	23
Chinese	1	1	2	4
Total	107	80	169	356

Gender = Female

	Income Level			
Race	$\leq \$10,000$	$> \$10000$ and $\leq \$25000$	$> \$25000$	Total
White	186	127	51	364
Black	11	7	3	21
Chinese	0	1	0	1
Total	197	135	54	386

Table 2: Three-way cross-classification of Gender, Race, and Income for a selected U.S. census tract. *Source*: 1990 Census Public Use Microdata Files.

Estimation via Markov moves

Change current table into a new one, still satisfying marginal conditions, by adding or substracting special **Markov Moves** a finite set of integral vectors $M = \{v_1, \ldots, v_k\}$ such that one can "*walk*" from any integral vector in TP to any other by a finite sequence of additions or subtractions of vectors in M.

0	0	0	0	0	0	0	0	0
0	0	−1	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	−1	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

- Apply a Markov move with probability $1/2$, and only if it does not produce negative entries. This gives a connected Markov chain that converges to the uniform distribution (rapid mixing in fixed dimensions Diaconis-Saloff-Coste, Chung-Graham-Yau).
- From random generation one can obtain an approximation to the number of elements can be deduced (Jerrum, Valiant, Vazirani).
- To generate the MARKOV moves: We need Graver bases!!!

Birth month vs Death month (CONTINUED)

- Significantly different is defined in comparison to all other possible contingency tables with the same column and row sums – i.e. the same expected table.
- But, there are too many other tables to list them all.
- So, we sample them using the **Monte Carlo Markov chain** method!!!
- We choose an initial table, moving randomly from one table to another say 50 times, sampling the current table, and repeating this as many times as required.

- We use a set of basic moves called a *Markov basis* to move from one table to another.

- It must be possible to move between any two possible contingency tables using set of basic moves.

- We can apply the basic move to the original table of birth-death months.

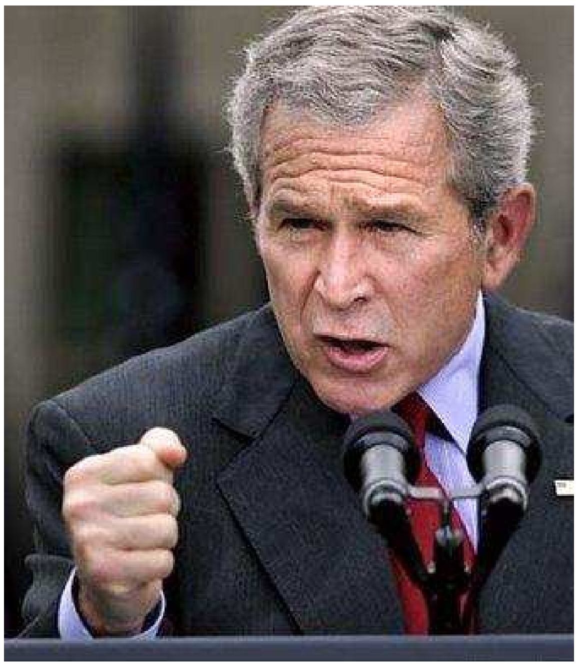
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Jan	**1-1**	0	0	0	**1+1**	2	0	0	1	0	1	0	6
Feb	1	0	0	1	0	0	0	0	0	1	0	2	5
Mar	1	0	0	0	2	1	0	0	0	0	0	1	5
Apr	3	0	2	0	0	0	1	0	1	3	1	1	12
May	**2+1**	1	1	1	**1-1**	1	1	1	1	1	1	0	12
Jun	2	0	0	0	1	0	0	0	0	0	0	0	3
Jul	2	0	2	1	0	0	0	0	1	1	1	2	10
Aug	0	0	0	3	0	0	1	0	0	1	0	2	7
Sep	0	0	0	1	1	0	0	0	0	0	1	0	3
Oct	1	1	0	2	0	0	1	0	0	1	1	0	7
Nov	0	1	1	1	2	0	0	2	0	1	1	0	9
Dec	0	1	1	0	0	0	1	0	0	0	0	0	3
Total	13	4	7	10	8	4	5	3	4	9	7	8	82

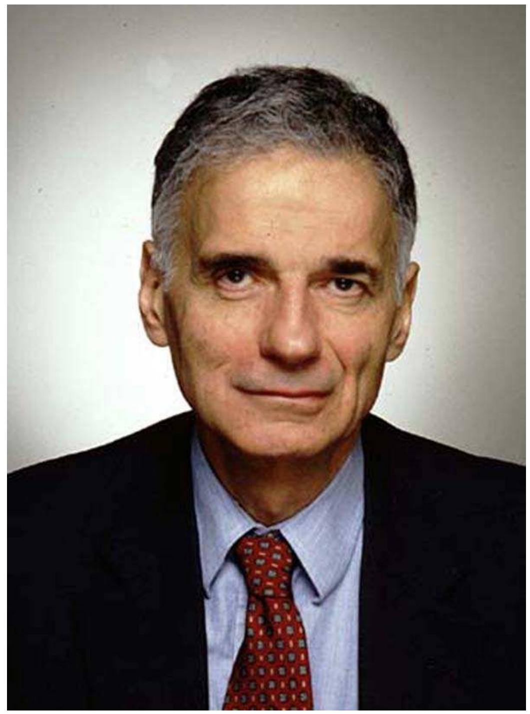
The new table after applying the basic move:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Jan	**0**	0	0	0	**2**	2	0	0	1	0	1	0	6
Feb	1	0	0	1	0	0	0	0	0	1	0	2	5
Mar	1	0	0	0	2	1	0	0	0	0	0	1	5
Apr	3	0	2	0	0	0	1	0	1	3	1	1	12
May	**3**	1	1	1	**0**	1	1	1	1	1	1	0	12
Jun	2	0	0	0	1	0	0	0	0	0	0	0	3
Jul	2	0	2	1	0	0	0	0	1	1	1	2	10
Aug	0	0	0	3	0	0	1	0	0	1	0	2	7
Sep	0	0	0	1	1	0	0	0	0	0	1	0	3
Oct	1	1	0	2	0	0	1	0	0	1	1	0	7
Nov	0	1	1	1	2	0	0	2	0	1	1	0	9
Dec	0	1	1	0	0	0	1	0	0	0	0	0	3
Total	13	4	7	10	8	4	5	3	4	9	7	8	82

we found **no** evidence of a link between birth month and death month!

VOTING THEORY!





- There are three (presidential) candidates a, b and c....

- Let the preference orders of the $n = \sum_{i=1}^{6} n_i$ voters be

$$abc\,(n_1),\ acb\,(n_2)\ bac\,(n_3),\ bca\,(n_4),\ cab\,(n_5),\ cba\,(n_6)$$

- Here, there are n_1 voters who rank candidate a as first, b second, and c third, n_2 voters who rank b first, a second, c third, etc.

- Let us assume all 6 rankings or orderings are equally likely!

- Under **simple plurality voting**, the candidate with the most vote wins.

- In a **plurality runoff system**, if no candidate wins more than 50% of the vote, the two candidates with the highest vote count advance to a second voting round.

- **Fact:** Different voting systems give different winners!!!
- Question What is the probability that the simple plurality and plurality runoff systems give different winners
- POLYTOPES is the answer:

  Set up a system of equations that describes the situation say

  item a wins by plurality but, using plurality runoff b obtains higher score than c and a majority of voters then prefer b to a.

$$\begin{aligned} 0 < \quad & n_1 + n_2 - n_3 - n_4 \\ 0 < \quad & n_3 + n_4 - n_5 - n_6 \\ \tfrac{-1}{2} < \quad & -n_1 - n_2 - n_5 \\ 1 = \quad & n_1 + n_2 + n_3 + n_4 + n_5 + n_6 \\ 0 \leq \quad & n_i, i = 1, \ldots, 6 \end{aligned}$$

- The probability of this event equals

$$\frac{\textbf{Number of lattice points on polytope above}}{\textbf{number lattice pts in}\{\,(n_1,\ldots,n_6):n_1+n_2+\ldots+n_6=1,\ n_i\geq 0\,\}}$$

- But all 6 possible voting rankings are possible. Multiply by 6 above number ,because the plurality winner may be a, b or c and the second position could be c not just b.

- **NOTE:** Asymptotically, the leading coefficients of these two quasipolynomials is all that matter!!!

- the volume of polytope is $\frac{71}{414720}$, and when multiplied by 6 one divides by $1/120$ (volumes of simplex) gives the probability these two voting systems give different winners for a large population to be 12%.

HOW about four candidates??

- The ordering outcomes are:

  $abcd(n_1), abdc(n_2), acbd(n_3), acdb(n_4), adbc(n_5), adcb(n_6)$

  $bacd(n_7), badc(n_8), bcad(n_9), bcda(n_{10}), bdac(n_{11}), bdca(n_{12})$

  $cabd(n_{13}), cadb(n_{14}), cbad(n_{15}), cbda(n_{16}), cdab(n_{17}), cdba(n_{18})$

  $dabc(n_{19}), dacb(n_{20}), dbac(n_{21}), dbca(n_{22}), dcab(n_{23}), dcba(n_{24})$

- Again, the sum of all variables n_i must be equal to the total number of voters.

- We have four inequalities expressing that when a is the plurality winner, b obtained a score higher than c and c obtained a score higher than d, thus

$$n_1 + n_2 + n_3 + n_4 + n_5 + n_6 > n_7 + n_8 + n_9 + n_{10} + n_{11} + n_{12}$$

$$n_7 + n_8 + n_9 + n_{10} + n_{11} + n_{12} > n_{13} + n_{14} + n_{15} + n_{16} + n_{17} + n_{18}$$

$$n_{13} + n_{14} + n_{15} + n_{16} + n_{17} + n_{18} > n_{19} + n_{20} + n_{21} + n_{22} + n_{23} + n_{24}$$

- A majority of voters prefer b over a and that a did not achieve more than 50 percent of the vote thus ($n_1 + n_2 + n_3 + n_4 + n_5 + n_6 < n/2$).

- These inequalities assume that the order was $a > b > c > d$ but the answer we get should be multiplied by $4! = 24$ to take into account other possible orders.

- The probability equals the volume of the above polyhedron times $4!$ divided by the volume of the simplex $\{(n_1, n_2, \ldots, n_{24}) : \sum_i n_i = 1,\ n_i \geq 0\}$. The answer is again roughly 12%.