An experimental mathematics approach to some combinatorial problems

Anthony Zaleski

Rutgers University

anthony.zaleski@rutgers.edu

March 29, 2018

Overview

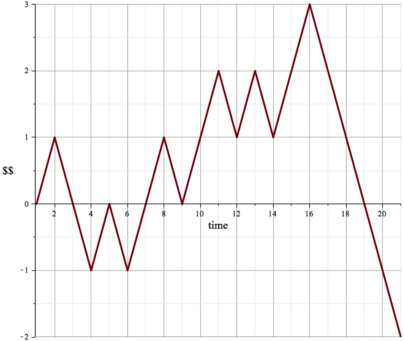
- Introduction
- Random walks
- Simultaneous core partitions
- Inclusion-exclusion
- Boolean functions

Introduction: experimental mathematics

- Computer as an essential tool, not just a fancy calculator
- Symbolic computing power (Maple in our case)
- Central ideas:
  - ansätze ("guess and check")
  - generating function methods
  - dynamical programming
  - OEIS
  - distributions of combinatorial statistics

Random walks

- Gambler starts with \$0.
- Wins or loses \$1 after each round, with equal probability.
- Finitely many steps.

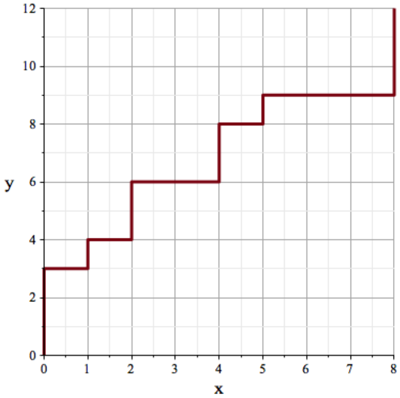


Walk statistics

- Feller [5] defines the following statistics of such a walk w:
  - $l(w)$ = length
  - $a_1(w) = \#$ of losing times (when money < 0)
  - $a_2(w) = \#$ of break-even times (money $= 0$)
  - $a_3(w)$ = time of last break-even
  - $a_4(w) = \#$ of sign changes.
- *Question*: how are the statistics distributed?

Related problem

- Consider walks in plane from $(0,0) \to (a,b)$.
- Each step one unit right (losing) or up (winning)
- Breaking even if $y = x$; losing region is $y < x$.



Moments

Let X be a random variable. Recall the following:

- kth (straight) moment $= \mathbb{E}[X^k]$ ($=\mu$ if $k = 1$)
- kth central moment $= \mathbb{E}[(X-\mu)^k]$ ($=\sigma^2$ if $k = 2$)
- kth standardized moment $= \mathbb{E}[(X-\mu)^k]/\sigma^k$
- Standardized moments = "fingerprint" of distribution
- E.g., $N(0,1)$ has standardized moments $0, 1, 0, 3, 0, 15, 0, 105, \ldots$ (A123023).

Big, important questions

Given an indexed random variable X_n of a certain class of combinatorial objects (e.g., walks),

- Express the moments in terms of n.
- Investigate the asymptotic distribution. Do the standardized moments approach some recognized sequence as $n \to \infty$?

Moments from generating functions

- Suppose X has finite sample space S, all outcomes equally likely.
- Define the g.f.

$$f(t) = \sum_{a \in S} t^{X(a)},$$

a finite polynomial in t.

- kth straight moment is

$$\frac{\left(t \frac{d}{dt}\right)^k f(1)}{|S|}.$$

Our results for Feller's problem

- We utilized Dr. Z.'s existing Feller package to derive results about the moments. Maple's "convert to formal power series" function was useful.
- For example, some information for the statistic "# of visits to $y = x$ of a uniform random walk from $(0,0) \to (n,n)$" are shown on the next slide.

- Mean:

$$\mu = \frac{-(2\,n)! + 4^n\,(n!)^2}{(2\,n)!}$$

- Variance:

$$\sigma^2 = -\frac{16^n\,(n!)^4 + 4^n\,(n!)^2\,(2\,n)! - 4\,n\,((2\,n)!)^2 - 2\,((2\,n)!)^2}{((2\,n)!)^2}$$

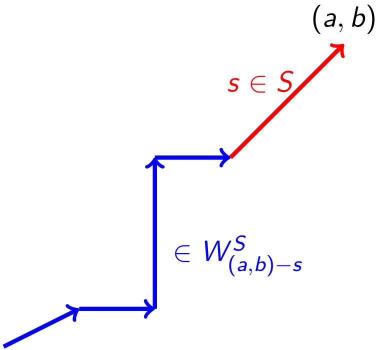
- Limits of 3rd-5th standardized moments:

$$2\,\frac{\sqrt{\pi}\,(\pi - 3)}{(-\pi + 4)^{3/2}}, -\frac{3\,\pi^2 - 32}{\pi^2 - 8\,\pi + 16}, 4\,\frac{\sqrt{\pi}\,\left(\pi^2 + 5\,\pi - 25\right)}{(-\pi + 4)^{5/2}}.$$

Arbitrary step sets

- Fix $S \subset \mathbb{N}^2$, $|S| < \infty$.
- Let $W_{a,b}^S = \{\text{walks from } (0,0) \to (a,b) \text{ with steps in } S\}$.
- How do the statistics of a uniform random $w \in W_{n,n}^S$ behave?
- Can we get asymptotic estimates of the moments as $n \to \infty$?

Dynamic programming scheme



$$W^{S}_{a,b} = \bigcup_{s \in S} W_{(a,b)-s}\{s\}$$

Computing generating functions

- For fixed S, a, b, the g.f.

$$F_{a,b}(t) := \sum_{w \in W^S_{a,b}} t^{a_1(w)}$$

is a *finite* polynomial in t.
- $F_{a,b}(t)$ can efficiently computed using the DP scheme above and `option remember`!
- Generate moment data from $F_{n,n}(t)$ for many values of n, and numerically analyze moment asymptotics.

Sample of the storybook

Steps	1	2	3	4
$\{01, 10\}$	$1.0000n$	$0.3n^2$	0.0000	1.800
$\{01, 20\}$	$0.38n$	$0.1n^2$	0.0	0.900
$\{02, 20\}$	$0.2500n$	$0.043n^2$	0.0000	0.900
$\{01, 02, 10\}$	$0.9n$	$0.27n^2$	0.0	1.8023
$\{01, 02, 20\}$	$0.33n$	$0.07n^2$	-0.02	0.90
$\{01, 10, 11\}$	$0.8n$	$0.2n^2$	0.0	1.8
$\{01, 11, 20\}$	$0.666n$	$0.15n^2$	0.001	1.80
$\{02, 11, 20\}$	$0.5n$	$0.08n^2$	0.0	1.80
$\{01, 02, 10, 11\}$	$0.81n$	$0.22n^2$	$0.$	1.80
$\{01, 02, 10, 20\}$	$0.80n$	$0.21n^2$	-0.01	1.804
$\{01, 02, 11, 20\}$	$0.6n$	$0.1n^2$	-0.01	1.80
$\{01, 10, 11, 20\}$	$0.81n$	$0.22n^2$	$0.$	1.80
$\{01, 02, 10, 11, 20\}$	$0.75n$	$0.19n^2$	-0.004	1.8

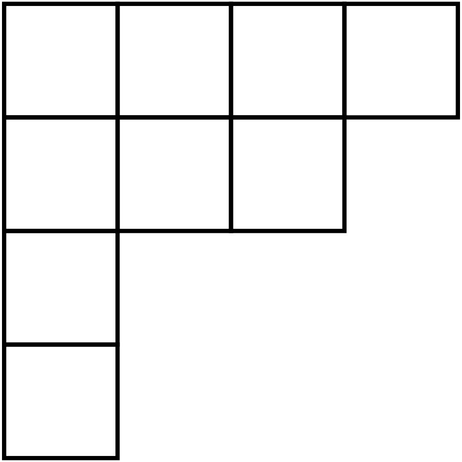
Walks in three dimensions

- Previously: winning ($y > x$), losing ($y < x$), and break-even ($y = x$) regions
- Now, 7 regions: $x < y < z, \ldots, z < y < x$, and "none of the above"
- We implemented analogous generating functions for statistics tracking # of visits to each region.
- A lot slower.

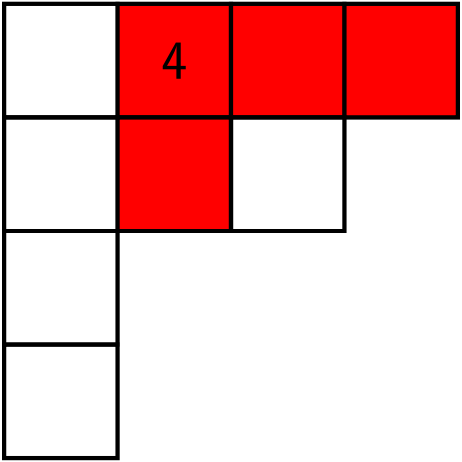
Partitions

- *Partition* of $n \in \mathbb{N}$: a nonincreasing sequence $\lambda = (\lambda_1, \ldots, \lambda_k)$ with $\lambda_i \in \mathbb{N}$ and $\sum \lambda_i = n$.
- $n = |\lambda|$ is the *size*.
- $\lambda_1, \ldots \lambda_k$ are the *parts*.
- E.g., $9 = 4 + 3 + 1 + 1$.
- Appear in representation theory, statistical mechanics, etc.

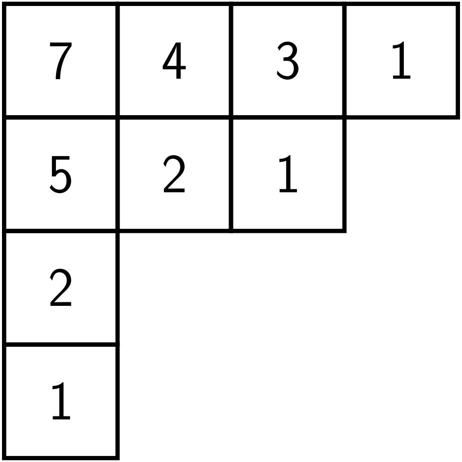
Young diagram of $(4, 3, 1, 1)$



Hook length of a cell



Young diagram of $(4, 3, 1, 1)$ showing hook lengths



Core partitions

- A partition is an *s-core* if it avoids hook length s (some definitions *equivalently* say "*divisible by s*" [9]).
- A (simultaneous) (s, t)-*core* avoids both hook lengths s and t.

Theorem (J. Anderson [1] 2002)

 # of (s, t)*-cores is finite iff* $\gcd(s, t) = 1$*, in which case it is*

$$(s + t - 1)!/(s!t!).$$

- Catalan (A000108) if $t = s + 1$.

Distribution of size

- If $\gcd(s, t) = 1$, consider random variable "size of a u.r. (s, t)-core."

Theorem (Conjectured by Armstrong [2], proved by Johnson [7])

The average size of an (s, t)-core partition is

$$\frac{(s-1)(t-1)(s+t+1)}{24}.$$

- Using Maple, Dr. Z. went up to the 6th moment.

Distinct parts

- Analysis seems harder if we require distinct parts.

Theorem (Straub 2016 [10])

The number of $(s, s+1)$*-cores with distinct parts is* F_{s+1}*.*

- Size distribution?

Our work: size of an $(s, s+1)$-core with distinct parts

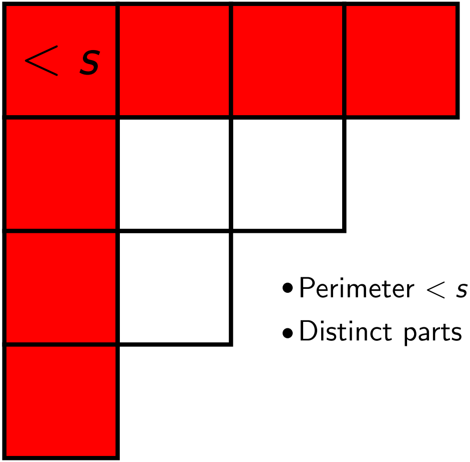
.

- Define $P_s = \{(s, s+1)\text{-cores with distinct parts}\}$
- G.f. of size:

$$G_s(q) := \sum_{p \in P_s} q^{|p|}$$

- Need a fast way to compute $G_s(q)$.

Straub's characterization of $p \in P_s$



- Perimeter $< s$
- Distinct parts

Recursive scheme for g.f.

- Perimeter formulation allows us to express the g.f. as a q-binomial sum:

$$G_s(q) = \sum_{m=0}^{s} q^{\binom{m+1}{2}} \binom{s-m}{m}_q .$$

- Use Dr. Z.'s package `qEKHAD` to find and prove a recursion for efficient computation.
- Also implies that the moments must satisfy the C-finite ansatz!

Sample results

Let X_s= "size of a u.r. $(s, s+1)$-core with distinct parts."

Theorem

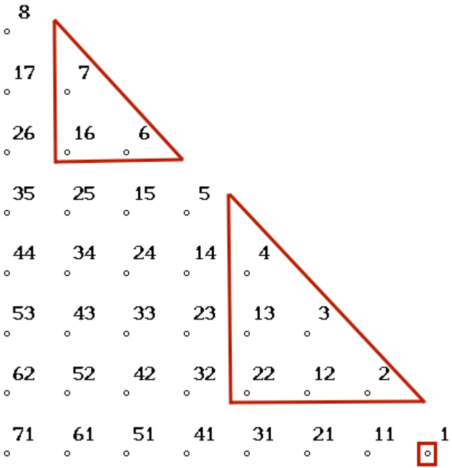
$$\mathbb{E}[X_s] = \frac{1}{50}\,\frac{5\,s^2 F_{s+1} - 6\,sF_s + 7\,sF_{s+1} - 6\,F_s}{F_{s+1}}.$$

- We got up to moment 16.
- Standardized moments approach $0, 1, 0, 3, 0, 15, 0, 105, \ldots$
- Conjecture: X_s is asymptotically normally distributed! (Contrast with previous case)

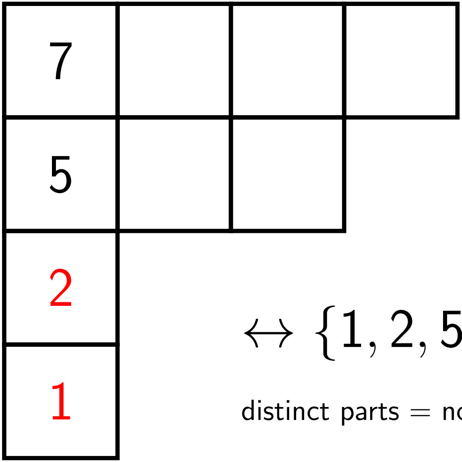
$(2s+1, 2s+3)$-cores with distinct parts

- Here, less lucky
- Use Anderson's bijection
- Define poset $P_{s,t} := \mathbb{N} \backslash (s\mathbb{N} + t\mathbb{N})$, with $c \leq_P d \iff d - c = n_1 s + n_2 t$
- *Order ideal*: subset closed under $\leq_p$
- (s,t)-cores $\leftrightarrow$ order ideals of $P_{s,t}$.

The poset $P_{9,10}$ and an order ideal (Catalan decomposition)



Illustrating the correspondence



$$\leftrightarrow \{1, 2, 5, 7\}$$

distinct parts = no consecutive labels

$P_{2s+1,2s+3}$ for $s = 6$

◦ 11											
◦ 24	◦ 9										
◦ 37	◦ 22	◦ 7									
◦ 50	◦ 35	◦ 20	◦ 5								
◦ 63	◦ 48	◦ 33	◦ 18	◦ 3							
◦ 76	◦ 61	◦ 46	◦ 31	◦ 16	◦ 1						
◦ 89	◦ 74	◦ 59	◦ 44	◦ 29	◦ 14						
◦ 102	◦ 87	◦ 72	◦ 57	◦ 42	◦ 27	◦ 12					
◦ 115	◦ 100	◦ 85	◦ 70	◦ 55	◦ 40	◦ 25	◦ 10				
◦ 128	◦ 113	◦ 98	◦ 83	◦ 68	◦ 53	◦ 38	◦ 23	◦ 8			
◦ 141	◦ 126	◦ 111	◦ 96	◦ 81	◦ 66	◦ 51	◦ 36	◦ 21	◦ 6		
◦ 154	◦ 139	◦ 124	◦ 109	◦ 94	◦ 79	◦ 64	◦ 49	◦ 34	◦ 19	◦ 4	
◦ 167	◦ 152	◦ 137	◦ 122	◦ 107	◦ 92	◦ 77	◦ 62	◦ 47	◦ 32	◦ 17	◦ 2

Look at the smallest vacant odd label

○ 11		
○ 24	○ 9	
○ 37	○ 22	○ 7

● 3					
○ 16	● 1				
○ 29	○ 14				
○ 42	○ 27	○ 12			
○ 55	○ 40	○ 25	○ 10		
○ 68	○ 53	○ 38	○ 23	○ 8	
○ 81	○ 66	○ 51	○ 36	○ 21	○ 6

Computing the g.f. and moments

- Decomposition gives a (complicated) recursive scheme for the g.f.

Theorem

The average size of a $(2s+1, 2s+3)$*-core partition with distinct parts is*

$$\frac{1}{32}(10\,s^3 + 27\,s^2 + 19\,s).$$

- Expressed moments up to 7th as polynomials in s
- Not asymptotically normal.

Other families of cores

- We can play the poset game some more.
- For each class of cores, we get a class of posets.
- Recursively characterize the order ideals
- Develop a scheme to compute g.f.s and find moments.
- We also got results for $(s, ds-1)$-cores with distinct parts, $(s, s+1)$-cores with parts repeated $\leq k$ times , and $(s, s+1)$-cores with *odd* parts (Johnson responded with an elegant abacus approach [8] and Dr. Z. donated $200 to the OEIS.)

Inclusion-exclusion

Theorem (principle of inclusion-exclusion)

Let $A_1, \ldots, A_N$ be events in a finite probability space. For $I \subset [N]$, define

$$A_I = \bigcap_{j \in I} A_j.$$

Then,

$$\Pr\left[\bigcup_i A_i\right] = \sum_{i=1}^{N} (-1)^{i+1} \sum_{I \subset [N], |I| = i} \Pr[A_I].$$

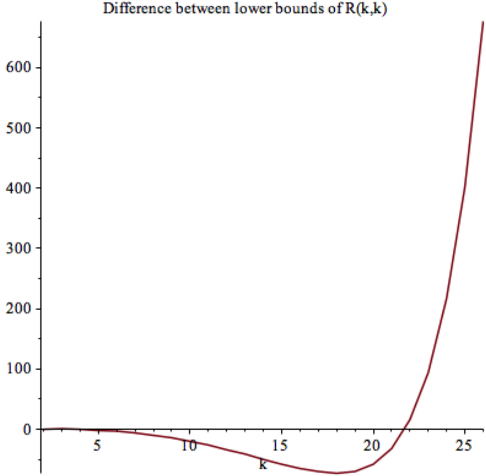
PIE and the probabilistic method

- Truncating after first term gives

$$\Pr\left[\bigcup_i A_i\right] \leq \sum_i \Pr[A_i].$$

- Boole's inequality, often used in probabilistic method.
- Idea: compute more terms in the sum before truncating for better bound (Bonferroni inequalities).

Improving the Erdős lower bound on $R(k, k)$



Boolean satisfiability

- n boolean variables $x_1, \ldots, x_n$
- *Conjunctive normal form (CNF)*: e.g.,

$$\neg x_3 \wedge (x_2 \vee x_3) \wedge (x_1 \vee \neg x_2)$$

  (Dual: *disjunctive normal form (DNF)*)
- SAT: given a CNF, determine whether it is satisfiable
- NP-complete.

SAT and inclusion-exclusion

- Consider dual of SAT: determine whether a DNF $C_1 \vee \cdots \vee C_N$ is a tautology.
- Randomly assign to $x_1, \ldots, x_n$ and let A_i be the event $C_i =$True.
- Tautology iff $\Pr\left[\bigcup_i A_i\right] = 1$
- Use Bonferroni bounds!
- Our solver is not competitive, but maybe theoretically interesting.

Covering systems (Erdős 1950 [4])

- *Covering system*: finite set of congruences

$$\{a_i \pmod{m_i} : 1 \leq i \leq N\}$$

whose union is $\mathbb{N}$

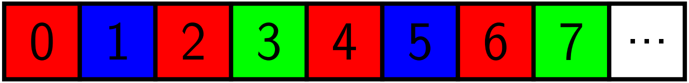
- *Exact* covering: disjoint congruences. E.g.,

$$\{0 \pmod 2, 1 \pmod 2\}$$

- *Distinct*: distinct moduli. E.g.,

$$\{0\,(2), 0\,(3), 1\,(4), 5\,(6), 7\,(12)\}.$$

Another exact covering



$$\{0 \pmod 2, 1 \pmod 4, 3 \pmod 4\}$$

Covering system facts

- A C.S. cannot be both exact and distinct. (Mirsky-Newman/Davenport-Rado)
- Erdős asked whether the smallest modulus m_1 of a distinct C.S. can be arbitrarily large. In 2015, Hough [6] proved $m_1 < 10^6$.
- Via the Chinese remainder theorem, Berger and Felzenbaum [3] described C.S. as a covering of a *finite* $p_1 \times \dots p_k$ box with sub-boxes.
- Exact = disjoint sub-boxes
- Distinct = non-parallel sub-boxes.

Our work: Boolean analogs

- We can view a DNF tautology $C_1 \vee \cdots \vee C_k$ in $x_1, \ldots, x_n$ as a covering of the Boolean n-cube by sub-cubes corresponding to the clauses ("congruences").
- Define the *support* ("modulus") of a clause C_i as the participating variables; e.g. $x_1 \wedge \neg x_2$ has support $\{x_1, x_2\}$.
- *Exact*: disjoint sub-cubes (any two clauses conflict)
- *Distinct*: clauses supported on distinct sets.

Minimum clause size in a distinct DNF tautology

- Unlike the Erdős case, minimum clause size k is unbounded (easy $k = n/2$ construction). For each n, how large can we get?
- Density argument:

$$\sum_{i=k}^{n} \binom{n}{i} \frac{1}{2^i} \geq 1$$

  gives rise to upper bound $k \leq A_n = 1, 1, 1, 2, 3, 4, 5, 5, \ldots$
- Using computer search methods, we constructed optimal DNF tautologies for $n \leq 14$ except $n = 10$ (1008/1024 vertices covered).

Constructing uniform distinct DNF tautologies

- What if we instead require all clauses to have same size, k?
- Density argument:

$$\binom{n}{k}\frac{1}{2^k} \geq 1,$$

  gives gives rise to upper bound
  $k \leq A_n = 1, 1, 1, 2, 3, 4, 5, 5, \ldots$
- Using computer search methods, we constructed optimal DNF tautologies for $n \leq 14$ except $n = 3, 5, 9, 13$.

Thank you!

References I

[1] J. Anderson.
Partitions which are simultaneously t_1 and t_2-core.
Discrete Math., 248:237–243, 2002.

[2] D. Armstrong, C. Hanusa, and B. Jones.
Results and conjectures on simultaneous core partitions.
European J. Combin., 41:205–220, 2014.

[3] M. A. Berger, A. Felzenbaum, and A. Fraenkel.
A nonanalytic proof of the Newman-Znam result for disjoint covering systems.
Combinatorica, 6:235–243, 1986.

References II

[4] P. Erdős.
On integers of the form $2^k + p$ and some related problems.
Summa Brasil. Math., 2:113–123, 1950.

[5] W. Feller.
An introduction to probability theory and its application.
John Wiley and Sons, 3rd edition, 1968.

[6] B. Hough.
Solution of the minimum modulus problem for covering systems.
Ann. Math., 181:361–382, 2015.

References III

[7] P. Johnson.
Lattice points and simultaneous core partitions.
https://arxiv.org/abs/1502.07934, 2015.

[8] P. Johnson.
Simultaneous cores with restrictions and a question of Zaleski and Zeilberger.
https://arxiv.org/abs/1802.09621, 2018.

[9] R. P. Stanley and F. Zanello.
The Catalan case of Armstrong's conjecture on simultaneous core partitions.
SIAM J. Discrete Math., 29:658–666, 2015.

References IV

[10] A. Straub.
Core partitions into distinct parts and an analog of Euler's theorem.
European J. Combin., 57:40–49, 2016.

[11] D. Zeilberger.
Fully AUTOMATED computerized redux of Feller's (v.1) Ch. III (and much more!).
http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/feller.html, 2006.

The End