A Story about Lattice Paths & Zeros and their Relation to Quicksort

Daniel Krenn

(joint work in progress with
Martin Aumüller, *Martin Dietzfelbinger*,
Clemens Heuberger and *Helmut Prodinger*)

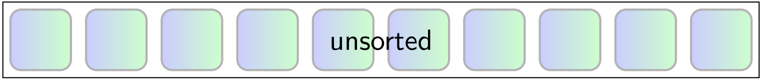
ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

November 3, 2015

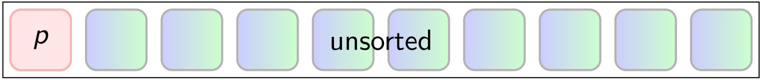
FWF Der Wissenschaftsfonds.

Supported by the
Austrian Science Fund (FWF),
project P24644.

Quicksort

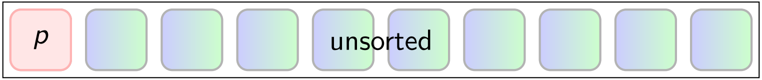


Quicksort

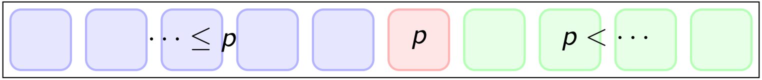


- choose a pivot element p

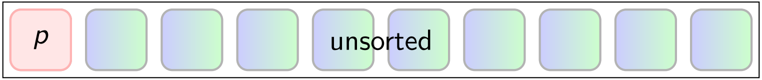
Quicksort



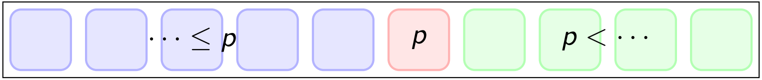
- choose a pivot element p
- partition into
  - small elements
  - large elements



Quicksort

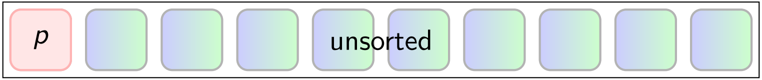


- choose a pivot element p
- partition into
  - small elements
  - large elements

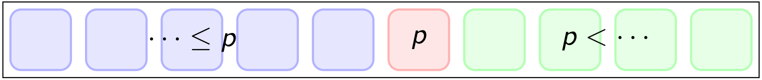


- proceed recursively

Quicksort

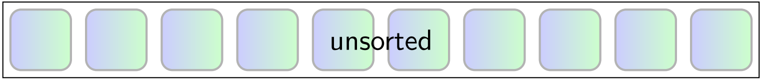


- choose a pivot element p
- partition into
  - small elements
  - large elements

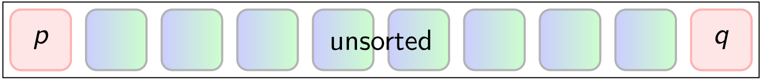


- proceed recursively
- $2n \log n + O(n)$ key comparisons

Dual Pivot Quicksort

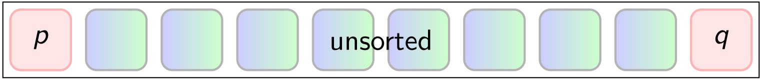


Dual Pivot Quicksort

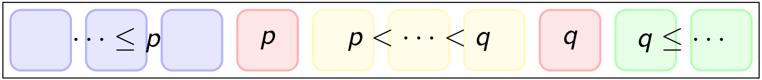


- choose pivot elements p and q

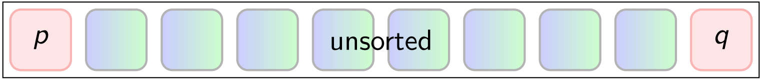
Dual Pivot Quicksort



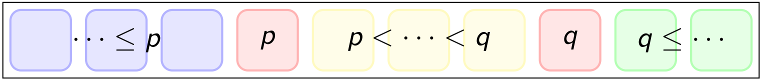
- choose pivot elements p and q
- partition into
  - small elements
  - medium elements
  - large elements



Dual Pivot Quicksort

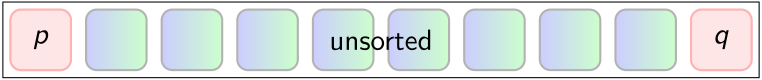


- choose pivot elements p and q
- partition into
  - small elements
  - medium elements
  - large elements

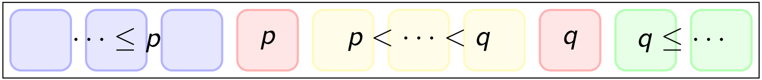


- proceed recursively

Dual Pivot Quicksort

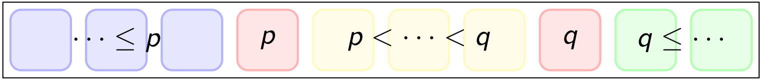


- choose pivot elements p and q
- partition into
  - small elements
  - medium elements
  - large elements



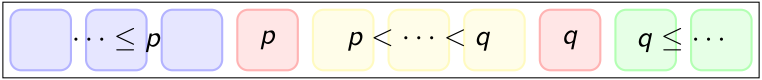
- proceed recursively
- at least $\frac{9}{5} n \log n + O(n)$ key comparisons

Partitioning/Classification Strategies



- average key comparisons
  - “Yaroslavskiy” $\rightsquigarrow$ $1.9 n \log n + O(n)$
  - “Count” $\rightsquigarrow$ $1.8 n \log n + O(n)$
  - “Clairvoyant” $\rightsquigarrow$ $1.8 n \log n + O(n)$
- modelled by a decision tree

Partitioning/Classification Strategies



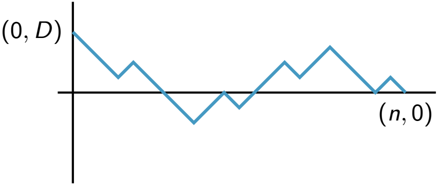
- average key comparisons
  - "Yaroslavskiy" $\rightsquigarrow$ $1.9 n \log n + O(n)$
  - "Count" $\rightsquigarrow$ $1.8 n \log n + O(n)$
  - "Clairvoyant" $\rightsquigarrow$ $1.8 n \log n + O(n)$
- modelled by a decision tree

Our Main Task

What is the precise minimum?

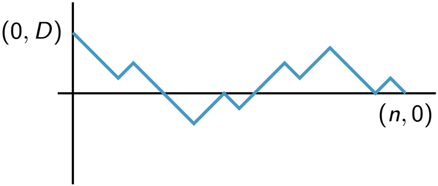
($\rightsquigarrow$ optimal strategy)

Choosing the Right Path: The (Simplified) Model



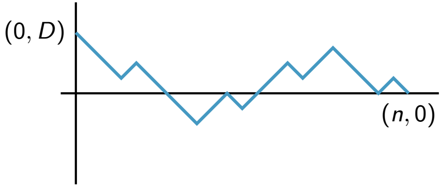
Choosing the Right Path: The (Simplified) Model

3. path from $(0, D)$ to $(n, 0)$ chosen uniformly at random among all possibilites



Choosing the Right Path: The (Simplified) Model

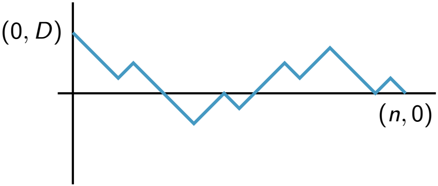
1. fix path length $n \in \mathbb{N}$
2. starting point $(0, D)$
   $D \in \{-n, -n+2, \ldots, n-2, n\}$
   uniformly at random
3. path from $(0, D)$ to $(n, 0)$
   chosen uniformly at random
   among all possibilites



Choosing the Right Path: The (Simplified) Model

1. fix path length $n \in \mathbb{N}$
2. starting point $(0, D)$
   $D \in \{-n, -n+2, \ldots, n-2, n\}$
   uniformly at random
3. path from $(0, D)$ to $(n, 0)$
   chosen uniformly at random
   among all possibilites

$X_n =$ number of zeros of path P_n



How many Zeros?

- length n
- path P_n according to simplified model

Question

- What is the expected number X_n of zeros?
- Asymptotic behavior?

How many Zeros?

- length n
- path P_n according to simplified model

Question

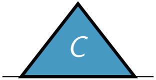
- What is the expected number X_n of zeros?
- Asymptotic behavior?

Theorem (ADH*K*P 2015+)

$$\mathbb{E}(X_n) = \frac{1}{2}\log n + \frac{1}{2}\log 2 + \frac{1}{2}\gamma - 1 + O\left(\frac{1}{n}\right)$$

Catalan Paths & The Symbolic Method

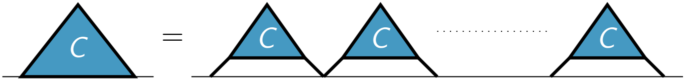
- symbolic equation



$=$

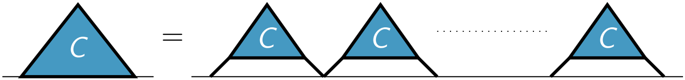
Catalan Paths & The Symbolic Method

- symbolic equation



Catalan Paths & The Symbolic Method

- symbolic equation

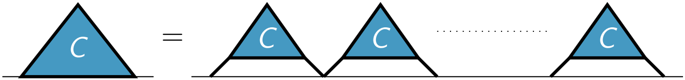


- translates to generating function

$$C(z) = 1 + z^2\, C(z) + \left(z^2\, C(z)\right)^2 + \cdots = \frac{1}{1 - z^2\, C(z)}$$

Catalan Paths & The Symbolic Method

- symbolic equation



- translates to generating function

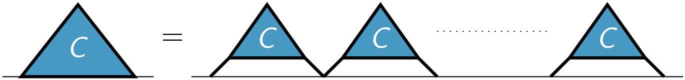
$$C(z) = 1 + z^2\,C(z) + \left(z^2\,C(z)\right)^2 + \cdots = \frac{1}{1 - z^2\,C(z)}$$

- explicit formula

$$C(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z^2}$$

Catalan Paths & The Symbolic Method

- symbolic equation



- translates to generating function

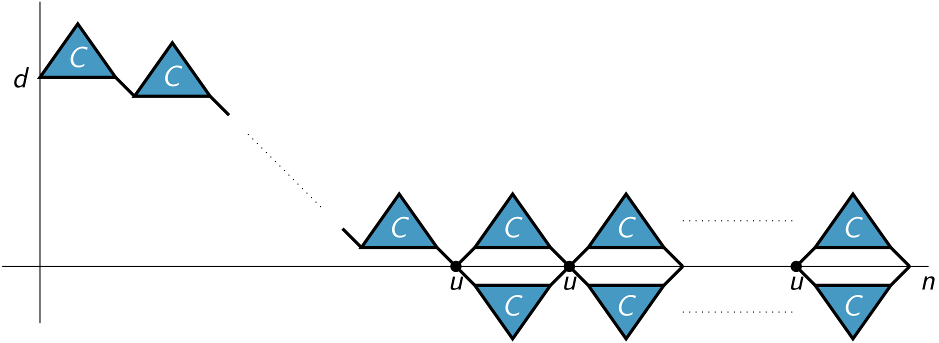
$$C(z) = 1 + z^2\, C(z) + \left(z^2\, C(z)\right)^2 + \cdots = \frac{1}{1 - z^2\, C(z)}$$

- explicit formula

$$C(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z^2} = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} z^{2n}$$

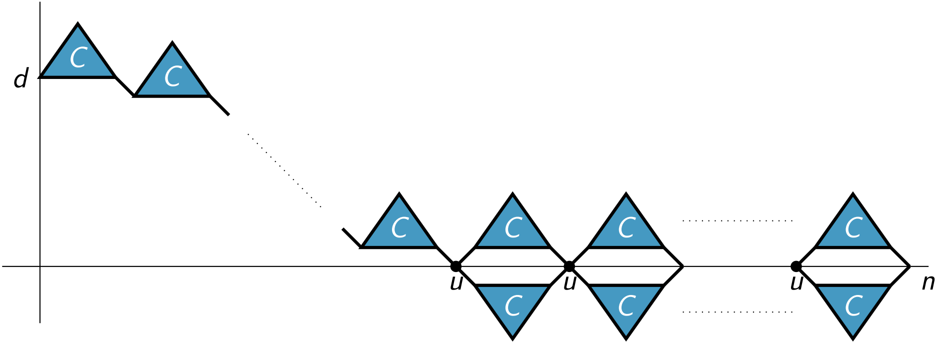
Symbolic Decomposition of Our Lattice Paths

- symbolic equation



Symbolic Decomposition of Our Lattice Paths

- symbolic equation



- translates to generating function

$$Q_d(z,u) = \frac{C(z)^{|d|}\, z^{|d|}}{1 - 2uz^2\, C(z)} = \frac{v^{|d|}(1+v^2)}{1 - v^2(2u-1)} \qquad \text{with } z = \frac{1}{1+v^2}$$

Getting the Double Sum

- expected number of zeros

$$\mu_{n,d} = \frac{[z^n]\, \frac{\partial}{\partial u} Q_d(z,u)\big|_{u=1}}{[z^n] Q_d(z,1)} = \frac{2}{\binom{n}{\ell}} \sum_{k=0}^{\ell-1} \binom{n}{k}$$

with $\ell = \frac{1}{2}(n - |d|)$

Getting the Double Sum

- expected number of zeros

$$\mu_{n,d} = \frac{[z^n]\ \frac{\partial}{\partial u} Q_d(z,u)\big|_{u=1}}{[z^n] Q_d(z,1)} = \frac{2}{\binom{n}{\ell}} \sum_{k=0}^{\ell-1} \binom{n}{k}$$

  with $\ell = \frac{1}{2}(n - |d|)$
- sum over all d
- expected number of zeros

$$\mathbb{E}(X_n) = \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right)$$

Getting the Double Sum

- expected number of zeros

$$\mu_{n,d} = \frac{[z^n]\, \frac{\partial}{\partial u} Q_d(z,u)\big|_{u=1}}{[z^n] Q_d(z,1)} = \frac{2}{\binom{n}{\ell}} \sum_{k=0}^{\ell-1} \binom{n}{k}$$

with $\ell = \frac{1}{2}(n - |d|)$

- sum over all d
- expected number of zeros

$$\mathbb{E}(X_n) = \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}]\, \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right)$$

What now?

Simplify?—Asymptotic behavior?

Simplification: A Computational Proof

- expected number of zeros

$$\mathbb{E}(X_n) = \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right)$$

- Sigma($\mathbb{E}(X_n)$)
  - $\rightsquigarrow$ returns a single sum
  - $\rightsquigarrow$ proof certificate: yes, but ...

A Recursion . . .

- $\mathbb{E}(X_n) = \dfrac{4}{n+1} \displaystyle\sum_{0 \le k < \ell < \lceil n/2 \rceil} \dfrac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \dfrac{1}{n+1} \left(\dfrac{2^n}{\binom{n}{n/2}} - 1 \right)$

- $\rightsquigarrow \mathbb{E}(X_{2i+1}) - \mathbb{E}(X_{2i}) = ?$ and

  $\rightsquigarrow \mathbb{E}(X_{2i+2}) - \mathbb{E}(X_{2i+1}) = ?$

A Recursion ...

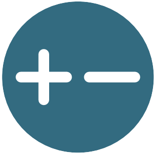
- $\mathbb{E}(X_n) = \dfrac{4}{n+1} \displaystyle\sum_{0 \le k < \ell < \lceil n/2 \rceil} \dfrac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \dfrac{1}{n+1} \left(\dfrac{2^n}{\binom{n}{n/2}} - 1 \right)$
-
  - $F(n, \ell) = \displaystyle\sum_{0 \le k < \ell} \dfrac{\binom{n}{k}}{\binom{n}{\ell}}$
  - $G(n, \ell) = (\ell - 1) + (\ell - 1 - n) F(n, \ell)$

- $\rightsquigarrow \mathbb{E}(X_{2i+1}) - \mathbb{E}(X_{2i}) = ?$ and

  $\rightsquigarrow \mathbb{E}(X_{2i+2}) - \mathbb{E}(X_{2i+1}) = ?$

A Recursion ...

- $\mathbb{E}(X_n) = \frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n \text{ even}] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right)$
-
  - $F(n,\ell) = \sum_{0 \le k < \ell} \frac{\binom{n}{k}}{\binom{n}{\ell}}$
  - $G(n,\ell) = (\ell - 1) + (\ell - 1 - n) F(n,\ell)$



... by Creative Telescoping

$$(n+1)F(n+1,\ell) - (n+2)F(n,\ell) = G(n,\ell+1) - G(n,\ell)$$

- $\rightsquigarrow \mathbb{E}(X_{2i+1}) - \mathbb{E}(X_{2i}) = 0$ and

  $\rightsquigarrow \mathbb{E}(X_{2i+2}) - \mathbb{E}(X_{2i+1}) = \frac{1}{2i+3}$

The Identity

Theorem (ADHKP 2015+)

$$\frac{4}{n+1} \sum_{0 \le k < \ell < \lceil n/2 \rceil} \frac{\binom{n}{k}}{\binom{n}{\ell}} + [n\ even] \frac{1}{n+1} \left(\frac{2^n}{\binom{n}{n/2}} - 1 \right)$$

$$=$$

$$\sum_{i=1}^{\lfloor n/2 \rfloor} \frac{1}{2i+1}$$

$$=$$

$$H_{2\lfloor n/2 \rfloor + 1} - \frac{1}{2} H_{\lfloor n/2 \rfloor} - 1$$

(harmonic numbers H_N)

A (More) Probabilistic Approach

- length n, path P_n
- consider point $(n-m, k)$
- k with $|k| \leq m$ and $k \equiv m \mod 2$

Key Property

$$\mathbb{P}((n-m, k) \in P_n) = \frac{1}{m+1}$$

A (More) Probabilistic Approach

- length n, path P_n
- consider point $(n-m,k)$
- k with $|k| \le m$ and $k \equiv m \mod 2$

Key Property

$$\mathbb{P}((n-m,k) \in P_n) = \frac{1}{m+1}$$

Theorem (ADH*K*P 2015+)

expected number of zeros

$$\mathbb{E}(X_n) = \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{1}{2i+1} = H_{2\lfloor n/2 \rfloor+1} - \frac{1}{2} H_{\lfloor n/2 \rfloor} - 1$$

(harmonic numbers H_N)

Everything is Easy Now—The Asymptotic Behavior

The expected number of zeros is

$$\mathbb{E}(X_n) = \frac{1}{2}\log n + \frac{1}{2}\log 2 + \frac{1}{2}\gamma - 1 + O\left(\frac{1}{n}\right)$$

Everything is Easy Now—The Asymptotic Behavior

The expected number of zeros is

$$\mathbb{E}(X_n) = \frac{1}{2}\log n + \frac{1}{2}\log 2 + \frac{1}{2}\gamma - 1 + O\left(\frac{1}{n}\right)$$

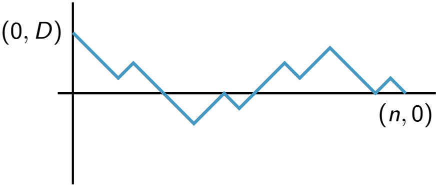
Theorem (ADHKP 2015+)

- $0 < \varepsilon \leq \frac{1}{2}$, $r = O\left(n^{1/2-\varepsilon}\right)$
- *distribution*

$$\mathbb{P}(X_n = r) = \frac{1}{(r+2)(r+1)}\left(1 + O\left(1/n^{2\varepsilon}\right)\right)$$

Choosing the Right Path: Now the Full Model

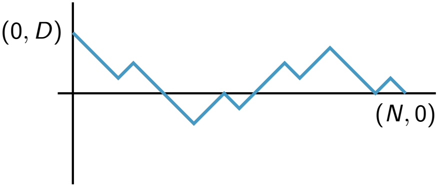
3. fix path length $n \in \mathbb{N}$
4. starting point $(0, D)$
   $D \in \{-n, -n+2, \dots, n-2, n\}$
   uniformly at random
5. path from $(0, D)$ to $(n, 0)$
   chosen uniformly at random
   among all possibilites



X_n = number of zeros of path P_n

Choosing the Right Path: Now the Full Model

1. fix $n_0 \in \mathbb{N}$
2. choose $\{P, Q\} \subseteq \{1, 2, \ldots, n_0\}$ uniformly among all $\binom{n_0}{2}$ possibilities
3. calculate path length $N = n_0 - 1 - (Q - P)$
4. starting point $(0, D)$ $D \in \{-N, -N+2, \ldots, N-2, N\}$ uniformly at random
5. path from $(0, D)$ to $(N, 0)$ chosen uniformly at random among all possibilites



X_N = number of zeros of path P_N

Back to Dual Pivot Quicksort

- path P_N according to full model

Corollary (ADHKP 2015+)

$$\mathbb{E}(X_N) = \frac{1}{2}\log n_0 + \left(\frac{1}{2}\log 2 + \frac{1}{2}\gamma - \frac{5}{4}\right) + O\left(\frac{1}{n_0}\right)$$

Back to Dual Pivot Quicksort

- path P_N according to full model

Corollary (ADH*K*P 2015+)

$$\mathbb{E}(X_N) = \frac{1}{2}\log n_0 + \left(\frac{1}{2}\log 2 + \frac{1}{2}\gamma - \frac{5}{4}\right) + O\left(\frac{1}{n_0}\right)$$

Theorem (ADH*K*P 2015+)

average number of key comparisons
in dual pivot quicksort
with **optimal** *partitioning strategy is*

$$\frac{9}{5}(n_0+1)H_{n_0+1} - \frac{2928+60\ln 2}{800}n_0 + \frac{H_{n_0}}{8} - \frac{1453-40\ln 2}{800} + O\left(\frac{1}{n_0}\right)$$

(harmonic numbers H_N)