The language and series of Hammersley type processes

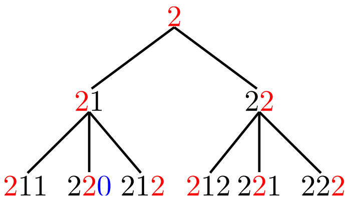
Cosmin Bonchiş, Gabriel Istrate, Vlad Rochian
West University of Timişoara, Romania

and the e-Austria Research Institute
gabrielistrate@acm.org
http://tcs.ieat.ro

Acks

Supported by IDEI Grant PN-III-P4-ID-PCE-2016-0842 ATCO, “Advanced techniques in optimization and computational complexity”

Summary for the technically-minded



- Study the grammatical complexity/formal power series of (generalization of) a model from the theory of interacting particle systems, the Hammersley process
- $k = 1$: $L^1_{HAM} = 1\{0,1\}^*$.
- $k \geq 2$: explicit form for L^k_H: DCFL, nonregular.
- Hammersley interval process: two languages, one equal to L^k_H, other explicit form, non-CFL (via Ogden).
- Algorithm for formal power series $\Rightarrow$ experiments, determining the value of a constant believed to be Φ.

"Story" for the Conceptually-minded

- The (classical) Ulam-Hammersley problem.

"Story" for the Conceptually-minded

- The (classical) Ulam-Hammersley problem.
- Heapability, and the U-H. problem for heapable sequences.

"Story" for the Conceptually-minded

- The (classical) Ulam-Hammersley problem.
- Heapability, and the U-H. problem for heapable sequences.
- The golden ratio conjecture and a "physics-like" argument.

"Story" for the Conceptually-minded

- The (classical) Ulam-Hammersley problem.
- Heapability, and the U-H. problem for heapable sequences.
- The golden ratio conjecture and a "physics-like" argument.
- This paper: Attempt to prove this conjecture via formal power series. Made (baby) first-steps.

"Story" for the Conceptually-minded

- The (classical) Ulam-Hammersley problem.
- Heapability, and the U-H. problem for heapable sequences.
- The golden ratio conjecture and a "physics-like" argument.
- This paper: Attempt to prove this conjecture via formal power series. Made (baby) first-steps.
- This talk: One result, One proof, one algorithm, one experiment.

Starting Point: *Longest Increasing Subsequence*

3 2 5 7 1 6 9

Patience sorting.

Another (greedy, also first-year) algorithm:

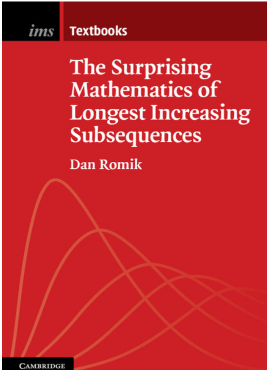
> Start (greedily) building decreasing piles. When not possible, start new pile.

Size of LIS = # of piles in patience sorting.

The Ulam-Hammersley problem (for random permutations)

What is the LIS of a random permutation ?

$$E_{\pi \in S_n}[LIS(\pi)] = 2\sqrt{n} \cdot (1 + o(1)).$$



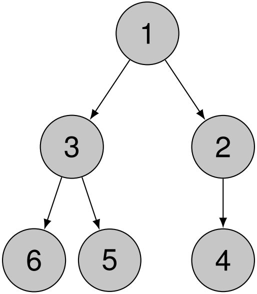
- Logan-Shepp (1977), Veršik-Kerov (1977), Aldous-Diaconis (1995)
- Very rich problem. Connections with nonequilibrium statistical physics and Young tableaux
- Also for intervals: Justicz, Scheinerman, Winkler (AMM 1990): random intervals on [0,1].

From (increasing) sequences to heaps

Byers, Heeringa, Mitzenmacher, Zervas (ANALCO'2011)

Sequence of integers A is heapable if it can be inserted into binary heap-ordered tree (not necessarily complete), always as leaf nodes.

Example: 1 3 2 6 5 4 Counterexample: 5 1 . . .



The Ulam-Hammersley problem for heapable sequences

- Simplest version trivial: $LHS(\pi) = n - o(n)$ (Byers et al.)
- (Dilworth, patience sorting): $LIS(\pi) =$ minimum number of decreasing sequences in a partition of π.

$HEAPS_k =$ minimum number of k-heapable sequences in a partition of π into such seqs.

Ulam-Hammersley problem for heapable sequences:

What is the scaling of $E_{\pi \in S_n}[HEAPS_k(\pi)]$, $k \geq 2$?

A beautiful conjecture

For $k \geq 2$ there exists $\lambda_k > 0$ such that

$$\lim_{n \to \infty} \frac{E[HEAPS_k(\pi)]}{ln(n)} = \lambda_k$$

Moreover

$$\lambda_2 = \frac{1 + \sqrt{5}}{2}$$

is the golden ratio.

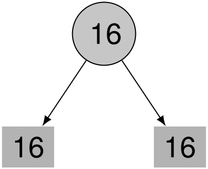
Status of the conjecture

- Some partial results.
- "Physics-like" nonrigorous argument, includes prediction for value of constant λ_k.
- Computations corroborated by experiments, "experimental mathematics" paper in progress.
- Follow-up work: Basdevant et al. (2016, 2017) rigorously establishes logarithmic scaling, but not the value of the constant.

Theorem: The "Patience heaping" algorithm correctly computes the value of parameter $Heaps_k(\pi)$.

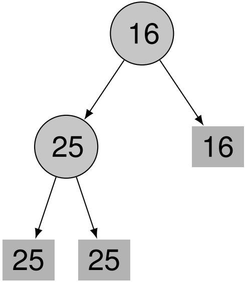
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



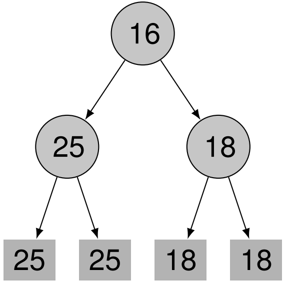
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



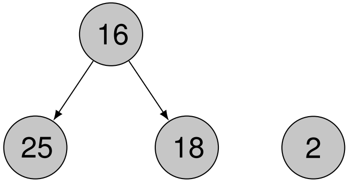
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



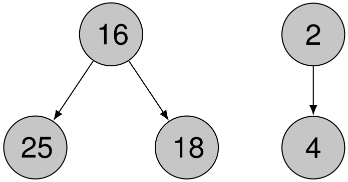
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



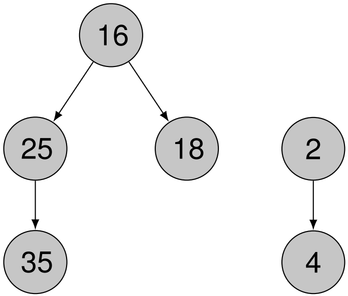
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



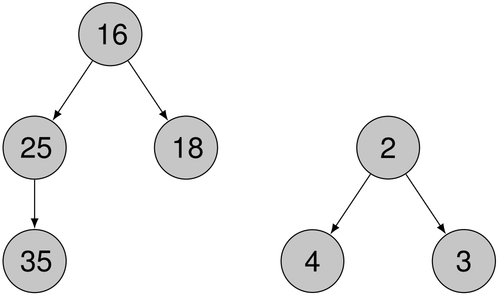
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



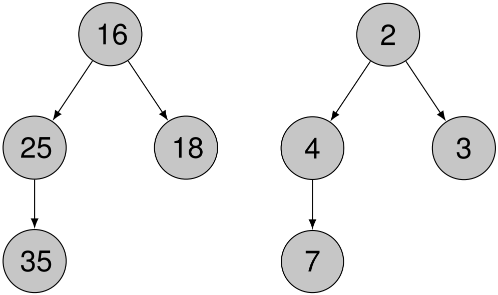
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



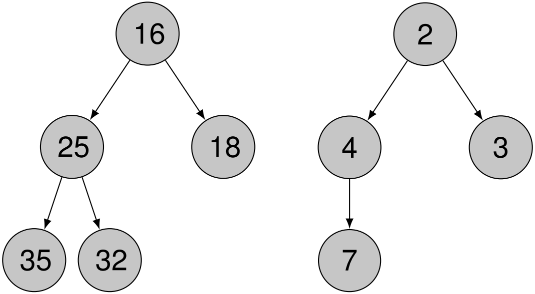
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



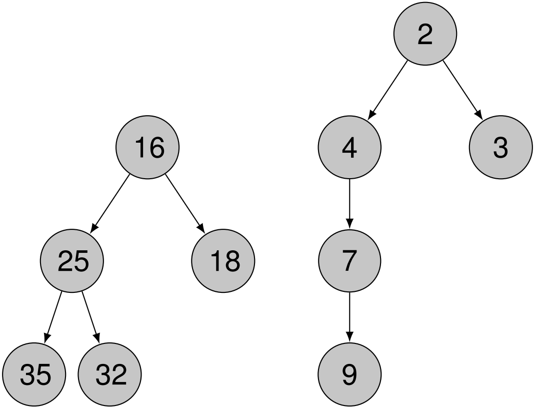
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



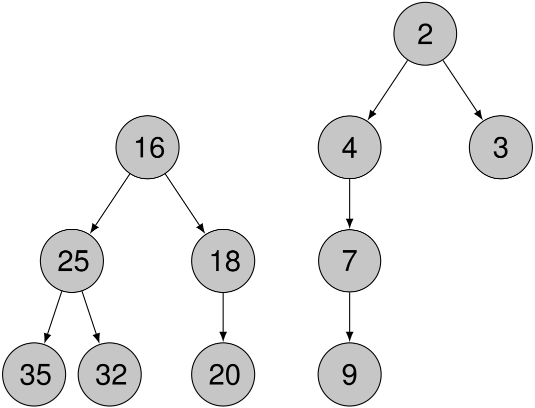
Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



Patience heaping

16, 25, 18, 2, 4, 35, 3, 7, 32, 9, 20



LIS and Hammersley's process

Top of piles in patience sorting = live particles in *Hammersley's process:*

- Particles: random real numbers $X_i \in (0, 1)$.
- Particle X_i kills closest live particle $X_j > X_i$ (if any)

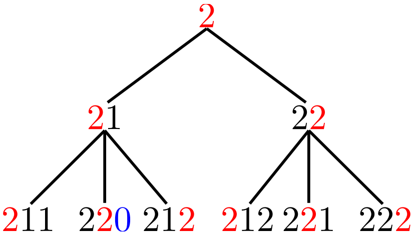
- studied in the area of interacting particle systems
- relative of a more famous process, the so-called Totally Asymmetric Exclusion Process (TASEP)

Aldous-Diaconis: Most illuminating proof of $E[LIS(\pi)] \sim 2\sqrt{n}$, analysis of the so-called hydrodynamic limit of Hammersley's process.

Hammersley's process with k lifelines (HAM_k):

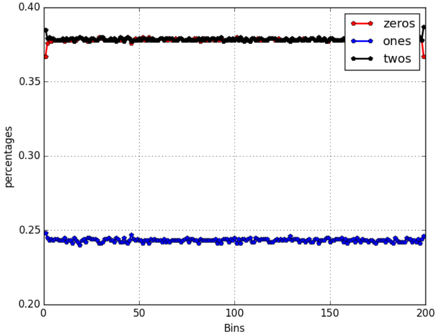
- Particles: slots in patience heaping
- Particles: random $X_i \in (0,1)$, initially k lives.
- X_i removes one lifeline from closest live $X_j > X_i$ (if any)
- Combinatorially, $k = 2$: Words over alphabet $0, 1, 2$.
- Choose a random position. Put there a 2. Remove 1 from the closest nonzero digit to the right (if any).

$E[\Delta(\#$ of heaps$)]$ = 1 + $E[\#$ of trailing zeros of $w]$



A "physicist's explanation" for the dynamics of HAD_k

- *$n \to \infty$: Limit of W_n =* compound Poisson process. $W_n =$ random string of 0,1,2 (densities c_0, c_1, c_2).
- *Assuming well mixing of digits* evolution equations $\to$ prediction on values of c_0, c_1, c_2.



- $c_0 = c_2 \sim \frac{3-\sqrt{5}}{2} \sim 0.381\ldots$, $c_1 \sim \sqrt{5} - 2 \sim 0.236\ldots$
- Distribution of trailing zeros: asymptotically geometric
- From this: $E[\Delta(\#\ heaps.)\ at\ stage\ n] \sim \frac{1+\sqrt{5}}{2} \cdot \frac{1}{(n+1)}$.

How could we (attempt to prove) this ?

- Study the formal power series of HAD_k: $F_k(w)$= multiplicity of word w in the process.
- Obtain probability by dividing by $|w|!$.

Sample Theorem from the paper:

$L_H^k =$ the set of words that satisfy the following condition:

- for all prefixes z of w

$$(*)|z|_k - \sum_{i=0}^{k-2}(k-i-1)\cdot|z|_i > 0.$$

(in particular w starts with a k).

Proof sketch

Direct inclusion: count transitions

- $k \to k + (k-1)$.
- $(k-1) \to k + (k-2)$: $a_{k-1} \geq 0$ moves.
- $\ldots$
- $1 \to k + 0$ $a_1 \geq 0$ moves..
- $\lambda \to k$: $a_0 \geq 1$ moves..
- So $|z|_0 = a_1, |z|_1 = a_2 - a_1, \ldots, |z|_k = a_0 + a_1 + \ldots + a_{k-1}$. Compute a_i in terms of $|z|_j$ and use condition $a_0 > 0$.

Opposite inclusions: several lemmas

- All words in L_H^k start with a k.
- L_H^k closed under prefix.

- All words with $(*) = 1$, $(*) > 0$ in L_H^k

Proof sketch

The induction

- $n = 1$: $z = k$, true.
- $n - 1 \Rightarrow n$. Let z be on the r.h.s. with $|z| = n$.
- Define w to be the word obtained from z by deleting rightmost k and increasing by 1 the next letter.
- w's definition correct: Deleted k not the last letter, otherwise some prefix of z would have $(*) = 0$.
- $|w| = n - 1$. All prefixes of w have $(*) > 0$: any decrease (if any) in the number of k's offset by increase in the value of the next letter.
- By induction $w \in L_H^k$. But w yields z in one step.
- Finally, every word z in the r.h.s. prefix of a word, e.g. $z(k-2)(k-2)\ldots$, with $(*) = 1$.

Algorithm for computing F_k

Input: $k \geq 1, w \in \Sigma_k^*$

$S := 0.\ w = w_1 w_2 \ldots w_n$

if $w \notin L_H^k$ **return** 0

if $w == \text{`}k\text{`}$ **return** 1

for i in 1:n-1

if $w_i == k$ and $w_{i+1} \neq k$

let $r = min\{l \geq 1 : w_{i+l} \neq 0 \text{ or } i + l = n + 1\}$

for j in 1:r-1

let $z = w_1 \ldots w_{i-1} w_{i+1} \ldots w_{i+j-1} 1 w_{i+j+1} \ldots w_{i+r} \ldots w_n$

$S := S + ComputeMultiplicity(k, z)$

if $i + r \neq n + 1$ and $w_{i+r} \neq k$

let $z = w_1 \ldots w_{i-1} w_{i+1} \ldots w_{i+r-1} (w_{i+r} + 1) w_{i+r+1} \ldots w_n$

$S := S + ComputeMultiplicity(k, z)$

Algorithm for computing F_k

if $w_n == k$

  let $Z = w_1 \ldots\ldots w_{n-1}$

  $S := S + ComputeMultiplicity(k, z)$

return S

The constant in the golden-ratio conjecture

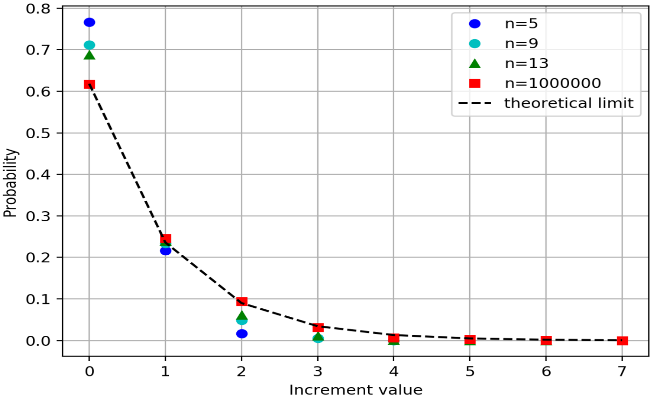


Figure 2: Probability distribution of increments, for $k = 2$, and $n = 5, 9, 13, 1000000$.

Conclusions

Rich problem with many open questions:

- The complexity status of the longest heapable subsequence (Byers et al. 2011)
- The formal power series of the Ham_k process
- The "golden ratio" conjecture (CPM'2015, also manuscript, 2018)
- Heapability of sets/seqs. of random intervals (2018)

$$\lim_{n\to\infty} \frac{E[k\text{-width}(P)]}{n} = \frac{1}{k+1}.$$

- Heapability of random d-dimensional posets (DCFS'2016) (random model: Winkler, Bollobas and Winkler)

$$E[k\text{-width}(P)] = \Theta(\log^{d-1}(n)).$$

Thank you. Questions ?