The Ulam-Hammesley problem for permutations and partial orders

(~~Heapability of partial orders~~)

Gabriel Istrate
West University of Timişoara, Romania
gabrielistrate@acm.org
http://tcs.ieat.ro

This work was supported by a grant of Ministery of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2016-0842, within PNCDI III

Disclaimers

- This presentation: Combines results from multiple papers (CPM'2015, DCFS'2016), plus manuscript in progress.
- manuscript: János Balogh (Szeged), Cosmin Bonchiş , Diana Diniş, G.I. (Timişoara), and Ioan Todincã (Orleans).
- Results and ideas, not proofs !
- ... tried hard to make things simpler than they actually are.

Outline

- The (classical) Ulam-Hammersley problem.

Outline

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

Outline

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

Outline

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

Starting Point: *Longest Increasing Subsequence*

3 2 5 7 1 6 9

Patience sorting.

Another (greedy, also first-year) algorithm:

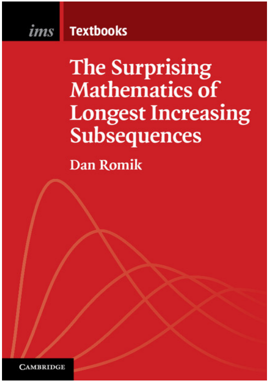
> Start (greedily) building decreasing piles. When not possible, start new pile.

Size of LIS = # of piles in patience sorting.

The Ulam-Hammersley problem (for random permutations)

What is the LIS of a random permutation ?

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



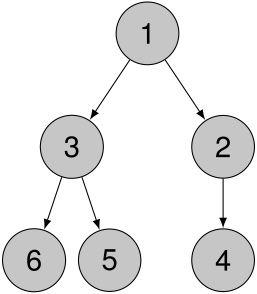
- Logan-Shepp (1977), Veršik-Kerov (1977), Aldous-Diaconis (1995)
- Very rich problem. Connections with nonequilibrium statistical physics and theory of Young tableaux

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.

Sample rigorous result: "Patience heaping"

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

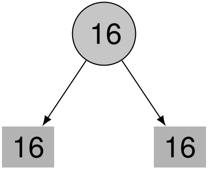
Proof idea:

By induction: The multiset of slots created by patience heaping is smaller ("dominates") the multiset of slots created by any other algorithm.

Consequently, when PatienceHeaping creates a new heap, so does any other algorithm.

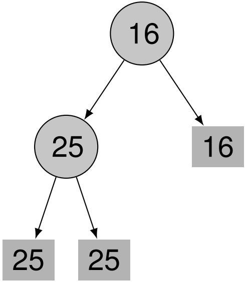
Patience heaping

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



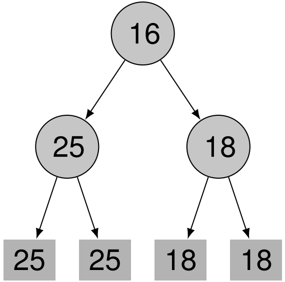
Patience heaping

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



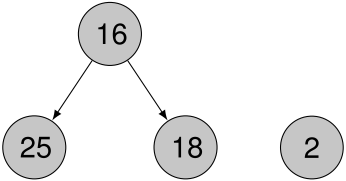
Patience heaping

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



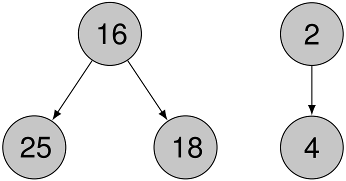
Patience heaping

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



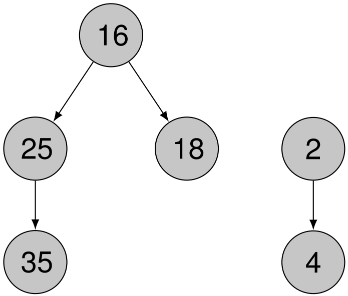
Patience heaping

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



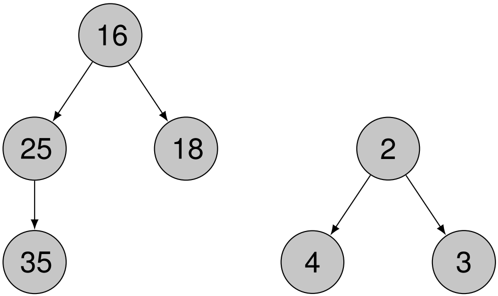
Patience heaping

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



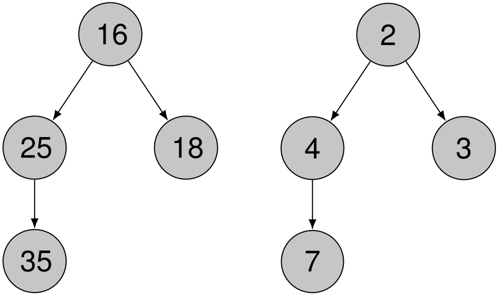
Patience heaping

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



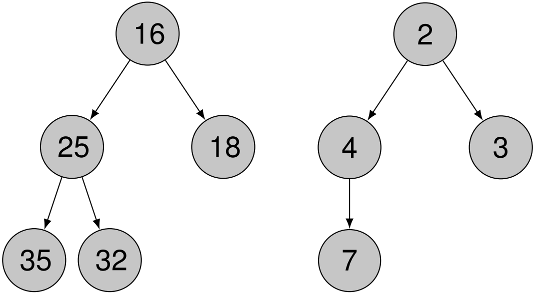
Patience heaping

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



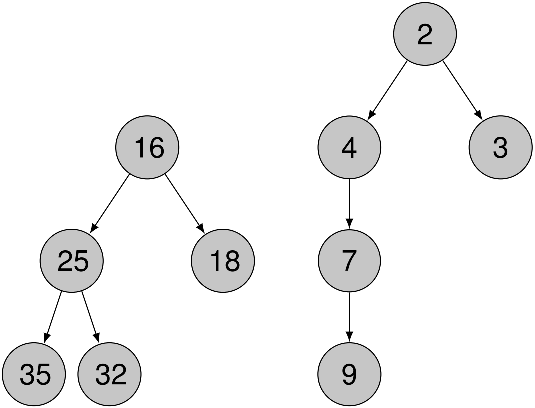
Patience heaping

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



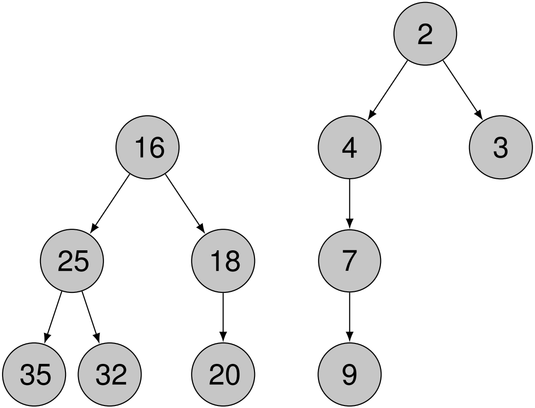
First result: Patience heaping

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)

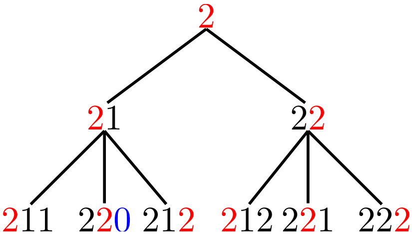
- studied in the area of interacting particle systems
- relative of a more famous process, the so-called Totally Asymmetric Exclusion Process (TASEP)

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

"Physics of patience heaping"

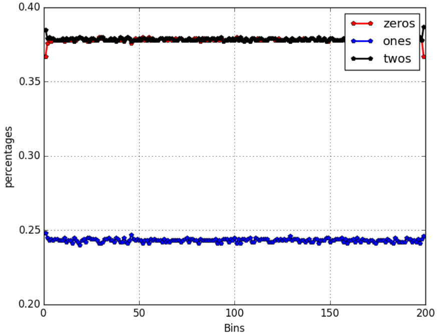
Hammersley's process with k lifelines (HAM_k):

- "Particles": random numbers $X_i \in (0, 1)$.
- each initially endowed with 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).
- $E[\Delta(\#\text{ of heaps})] = 1 + E[\#\text{ of trailing zeros of } w]$



A "physicist's explanation" for the dynamics of HAM$_k$

- *$n \to \infty$: 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: $E[\Delta(\#\ heaps.)\ at\ stage\ n] \sim \frac{1+\sqrt{5}}{2} \cdot \frac{1}{(n+1)}$.

The plot thickens: intervals/posets

- Longest increasing subsequence of intervals: another first-year problem.
- Justicz, Scheinerman, Winkler (AMM 1990): random intervals on [0,1].

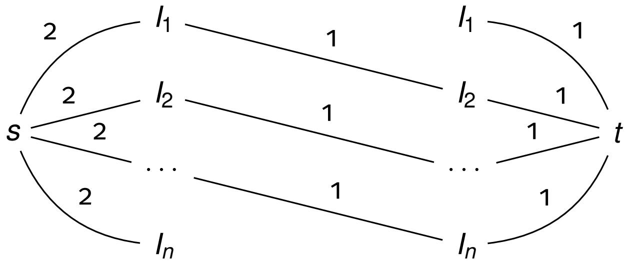
$$E[LIS(\pi)] \sim \frac{2\sqrt{n}}{\sqrt{\pi}}.$$

Theorem: PatienceHeaping still computes the decomposition of a sequence of intervals into a minimal number of heaps.

- Slot value of an interval: larger end.

Alternate solution

János Balogh (Szeged):



Theorem: number of heaps = n-maxflow.

Heapability and Dilworth's theorem.

- Width of poset: the minimum number of classes of a partition into chains.

Dilworth's Theorem: width= cardinality of largest antichain.

- k-chains: every item at most k direct successors.
- k-width(P): Minimum # classes of a partition into k-chains.

Theorem: The Network flow algorithm computes the k-width of arbitrary partial orders.

Essentially a generalization of (part of) the (network flow) proof of Dilworth's theorem.

IP computation of the k-width

Theorem

The following IP for k-width(P) has a totally unimodular matrix:

$$\begin{cases} max(1 + \sum_{p \prec q} X_{p,q}) \\ \sum_{q: p \prec q} X_{p,q} \leq k, \forall p \in P \\ \sum_{p: p \prec q} X_{p,q} \leq 1, \forall q \in P \\ X_{p,q} \in \{0, 1\} \end{cases}$$

Proof: if 1 instead of k on the right-hand side, IP for maximum matching in bipartite graph.

Further results on partial orders

THM: The k-width of a set of intervals is equal to that of the sequence of intervals obtained by listing intervals sorted by their right endpoint.

Trapezoid posets: induced by partial order among boxes in $\mathbf{R}^n$. Corresponding graphs: poly-time coloring, ind. set, etc.

Theorem: Greedy algorithm computes the k-width of trapezoid partial orders.

Theorem: For interval partial orders the longest heapable subsequence computable in poly-time.

Conclusions

Rich problem with many open questions:

- The complexity status of the longest heapable subsequence (Byers et al. 2011)

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 (G.I., Cosmin Bonchiş, 2017)

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 (G.I., Cosmin Bonchiş, 2017)
- The "golden ratio" conjecture (CPM'2015, also manuscript, early 2018)

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 (G.I., Cosmin Bonchiş, 2017)
- The "golden ratio" conjecture (CPM'2015, also manuscript, early 2018)
- Heapability of sets/seqs. of random intervals (2017 manuscript)

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

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 (G.I., Cosmin Bonchiş, 2017)
- The "golden ratio" conjecture (CPM'2015, also manuscript, early 2018)
- Heapability of sets/seqs. of random intervals (2017 manuscript)

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

- 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 ?