The Ulam-Hammersley problem for heapable sequences.

Gabriel Istrate

West University of Timişoara and e-Austria Research Institute, Timişoara, Romania

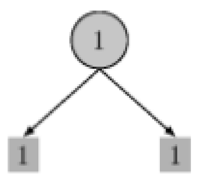
gabrielistrate@acm.org

1 Motivation: Heapability of integer sequences.

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) as leaves only.

- Polynomial time algorithm to decide heapability.
- Complete heapability NP-complete.
- If $\pi \in S_n$ is a random permutation, w.h.p. $LHS(\pi) = n - o(n)$.

2 Slots



- A node comes with two slots of the same value
- Slots may be occupied by larger numbers.
- If $A =$ (e.g.) 1 3 2 6 5 4 then a free slot is always available, and the sequence is heapable.
- If $A =$ (e.g.) 5 1 . . . then there is no good slot for 1.

3 Decomposition into heapable subsequences via Patience Heaping.

$MHS_k(A) =$ minimum number of k-ary heapable sequences one can decompose A into.

Algorithm 3.1: PATIENCE-HEAPING(W)

```
INPUT W = (w_1, w_2, ..., w_n) a list of integers.
Start with empty heap forest T = ∅.
for i in range(n):
  if (there exists a slot where X_i can be inserted):
      insert X_i in the slot with the lowest value.
  else :
      start a new heap consisting of X_i only.
```

THEOREM [3]: "Patience heaping" computes $MHS_k(A)$.

Proof Idea: (a). Define domination relation between multisets of slots. (b). Greedy insertion dominates any other insertion+ induction. If GREEDY creates new heap then any other algorithm does.

4 Extensions to Intervals/Partial orders of Finite Dimension.

Theorem [5]): PatienceHeaping still computes the decomposition of a sequence of intervals into a minimal number of heaps.

Proof Idea: Slot value of an interval: larger end.

- Longest increasing subequence of intervals: greedy.
- Justicz, Scheinerman, Winkler (AMM 1990): random intervals on [0,1]. $E[LIS(\pi)] \sim \frac{2\sqrt{n}}{\sqrt{\pi}}$.
- Width: the minimum # chains in a chain decomposition. $\boxed{\text{Dilworth's Thm: } w(P) = \text{size of largest antichain.}}$
- height: the length of the longest chain.
- Dimension: the minimum number of permutations $P_1, P_2, \ldots, P_k$ s.t. $P = P_1 \cap P_2 \cap \ldots P_k$.
- Random p.orders of dimension k: $P_k(n)$ (Winkler, 1985). Random permutations: $k = 2$.
- Width of random $P \in P_k(n)$: $\Theta(n^{1/k})$, Winkler (1985), Brightwell (1992).
- Height of random $\pi \in P_k(n)$: Winkler(1985), Bollobás and Brightwell (1992).

THEOREM [4]: The following IP for $MHS_k(P)$ has a totally unimodular matrix:

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

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

THEOREM [4]: For every fixed $k, d \geq 1$, $E_{P \in P_{d,n}}[MHS_k(P)] = \Omega(\log^{d-1}(n))$.
Proof Idea: Sequence minima start new heaps. Expected # minima analyzed by Winkler (1988).

5 Longest Heapable Subsequence

Complexity: open [2]

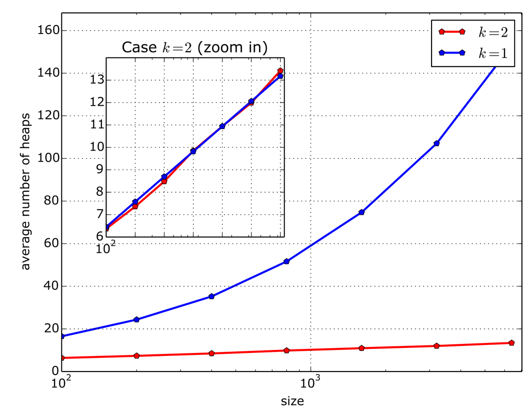
THEOREM [4]: For every fixed $k, d \geq 1$, $E_{P \in P_{d,n}}[LHS(P)] = n - o(n)$.
Proof Idea: Straightforward adaptation of argument from [2].

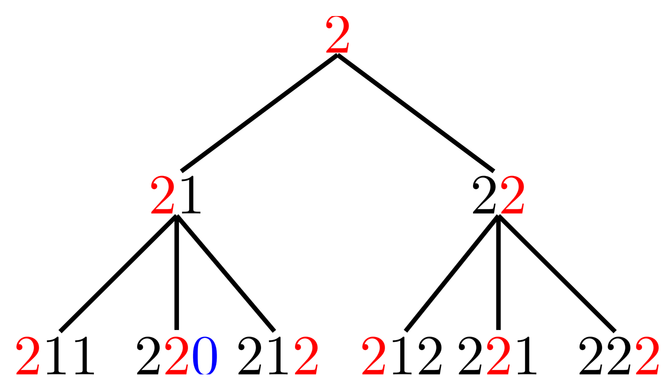
6 Scaling of $E[MHS_2(\pi)]$: random permutations and random intervals.

CONJECTURE: We have $\lim_{n \to \infty} \frac{E[MHS_2[\pi]]}{\ln(n)} = \phi$, with $\phi = \frac{1+\sqrt{5}}{2}$ the golden ratio.

Note: Basdevant et al. [1] believe constant is slightly smaller.

Empirically $E[MHS_2(P)] \sim \frac{n}{3}$!





7 "Physics like argument" via a Multiset Hammersley Process.

Hammersley's process with $k \geq 1$ lives: particles arrive as random points in [0,1]. Each endowed with k lives. A new particle p takes one life from closest $q < p$.

- Live particles correspond to slots in Patience Heaping.
- New heaps: # of heaps = # of (local) maxima.

- Words over alphabet $0, 1, 2$.
- Start with $W_0 = \lambda$.
- Choose random position. Put there a 2. Remove 1 from the closest nonzero digit to the right (if any).

INTUITION: Process "converges to" a compound Poisson Process with densities $d_0 = d_2 \sim \frac{3-\sqrt{5}}{2} \sim 0.381\ldots, d_1 \sim \sqrt{5} - 2 \sim 0.236\ldots$.

Scaling: # minima in this limit process.

Assuming existence of constants and well-mixing, "mean-field" flow equations:

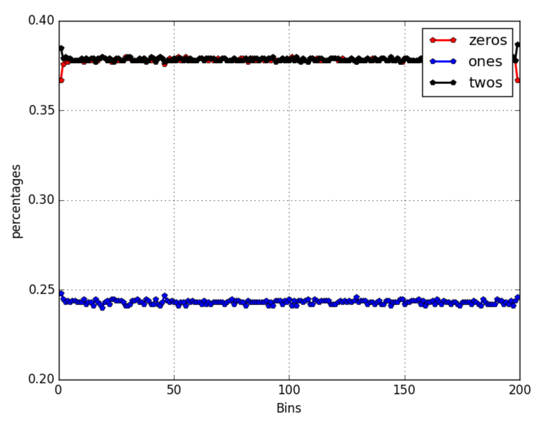
$$d_2 = 1 - \frac{d_2}{d_1 + d_2}, d_1 = \frac{d_2 - d_1}{d_1 + d_2}, d_0 + d_1 + d_2 = 1.$$

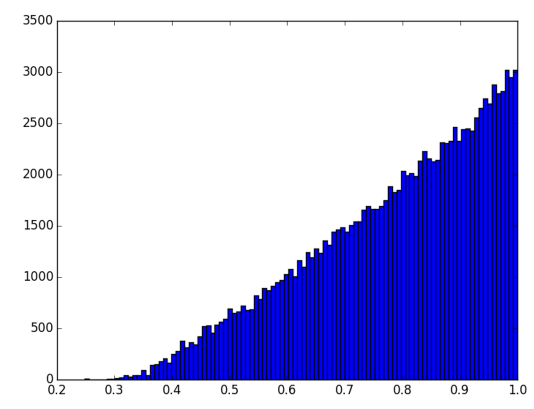
yield promised values for parameters of compound Poisson process.

8 ... and for intervals

New version of Hammesley process with k lifelines: new particle= random interval $I_n = [x_n, y_n]$. x_n kills lowest particle, but y_n is the new particle !

Limit of Hammersley process: far from Poisson jump process !





9 Relevant Publications

- [1] Basdevant, Gerin, Gouere, Singh. *arxiv/1605.02981*
- [2] J. Byers, B. Heeringa, M. Mitzenmacher, M., & G. Zervas, G. Heapable sequences and subseqeuences. In *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics (ANALCO'2011)*, pp. 33-44. Society for Industrial and Applied Mathematics.
- [3] G.Istrate, C.Bonchiş. Partition into heapable sequences, heap tableaux and a multiset extension of Hammersleyâs process, *Proceedings of the 26th Annual Symposium on Combinatorial Pattern Matching (CPM'2015)*, Lecture Notes in Computer Science vol. 9133, Springer Verlag, 2015.
- [4] G. Istrate, C. Bonchiş. Heapability, interactive particle systems, partial orders: results and open problems, *Proceedings of the 18th International Conference on Descriptive Complexity of Formal Systems (DCFS'2016)*, Lecture Notes in Computer Science, vol. 9777, Springer Verlag, 2016.
- [5] manuscript with C. Bonchiş, D. Diniş, J. Bálogh, I. Todinca. *arXiv.org:abs/1706.01230v3*, submitted for publication.

Building Bridges II, July 2018. Supported by IDEI Grant PN-III-P4-ID-PCE-2016-0842 ATCO, "Advanced techniques in optimization and computational complexity".