The entropy of the longest increasing subsequences: typical and extreme sequences

Phil Krabbe[1] Hendrik Schawe[1,2] Alexander K. Hartmann[1]

[1]Carl von Ossietzky Universität Oldenburg

[2]CY Cergy Paris Université

March 18, 2020

Outline

The Model: LIS

COunting LIS efficiently

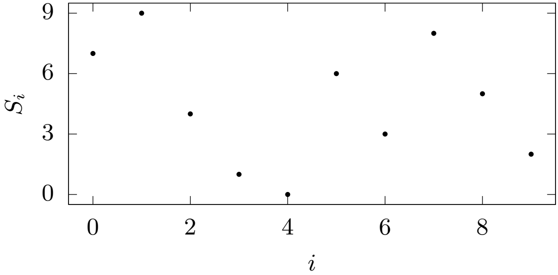
Distribution of the Entropy

Sampling the Far Tails

Longest Increasing Subsequence (LIS)

"Mark the most elements, such that all marked elements left of a marked element are smaller (or equal) than it"

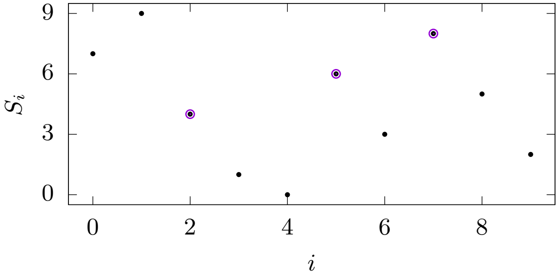
$$7 \quad 9 \quad 4 \quad 1 \quad 0 \quad 6 \quad 3 \quad 8 \quad 5 \quad 2$$



Longest Increasing Subsequence (LIS)

"Mark the most elements, such that all marked elements left of a marked element are smaller (or equal) than it"

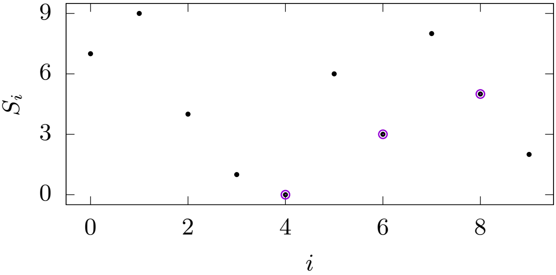
7 9 (4) 1 0 (6) 3 (8) 5 2



Longest Increasing Subsequence (LIS)

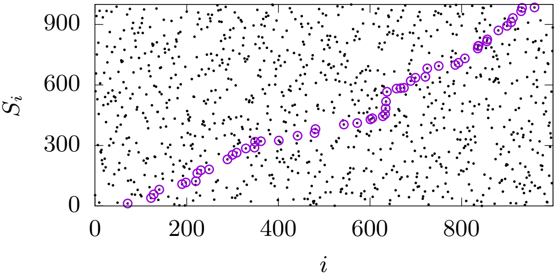
"Mark the most elements, such that all marked elements left of a marked element are smaller (or equal) than it"

7 9 4 1 (0) 6 (3) 8 (5) 2



Longest Increasing Subsequence (LIS)

"Mark the most elements, such that all marked elements left of a marked element are smaller (or equal) than it"



Longest Increasing Subsequence (LIS)

"Mark the most elements, such that all marked elements left of a marked element are smaller (or equal) than it"

With a permutation of length n:

- What is the expected length of the LIS? $\Rightarrow 2\sqrt{n}$ [1, 2, 3, 4]
- Why is this interesting?
  - random matrix theory [4]
  - surface growth (KPZ) [5, 6]
  - applications in computer science and bioinformatics
- How many possibilities are there? $\Rightarrow \mathcal{O}(\exp(n))$ [7]

---

[1] SM Ulam (1961); [2] K Johansson (1998); [3] J Deuschel, O Zeitouni (1999); [4] J Baik, P Deift, K Johansson (1999); [5] M Prähofer, H Spohn (2000); [6] SN Majumdar, S Nechaev (2004); [7] JM Hammersley (1972)

Counting Longest Increasing Subsequences

Number of different LIS grows exponentially

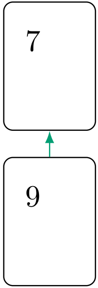
⇒ We can not just enumerate, we have to count cleverly.

7

$7 \quad 9 \quad 4 \quad 1 \quad 0 \quad 6 \quad 3 \quad 8 \quad 5 \quad 2$

Counting Longest Increasing Subsequences

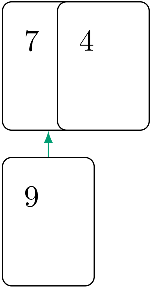
Number of different LIS grows exponentially
⇒ We can not just enumerate, we have to count cleverly.



$7 \quad 9 \quad 4 \quad 1 \quad 0 \quad 6 \quad 3 \quad 8 \quad 5 \quad 2$

Counting Longest Increasing Subsequences

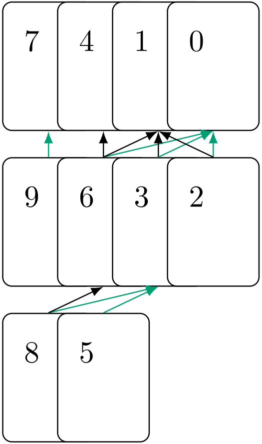
Number of different LIS grows exponentially
⇒ We can not just enumerate, we have to count cleverly.



7 9 4 1 0 6 3 8 5 2

Counting Longest Increasing Subsequences

Number of different LIS grows exponentially
$\Rightarrow$ We can not just enumerate, we have to count cleverly.

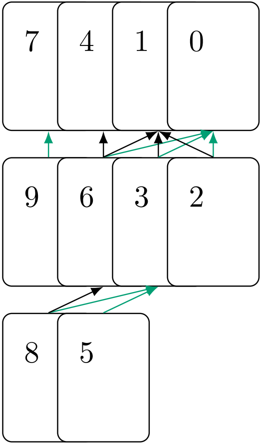


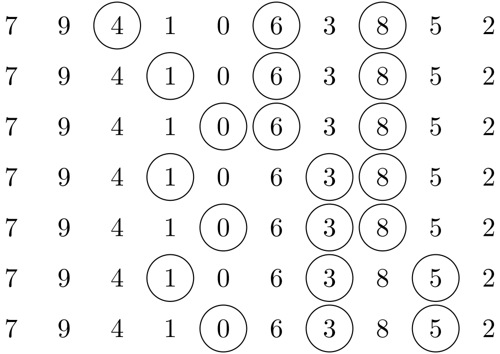
7 9 4 1 0 6 3 8 5 2

Counting Longest Increasing Subsequences

Number of different LIS grows exponentially

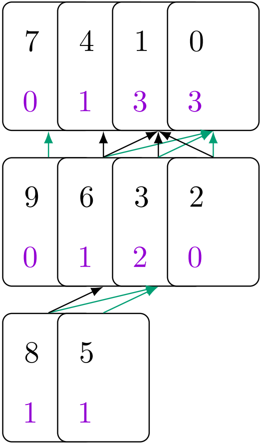
⇒ We can not just enumerate, we have to count cleverly.

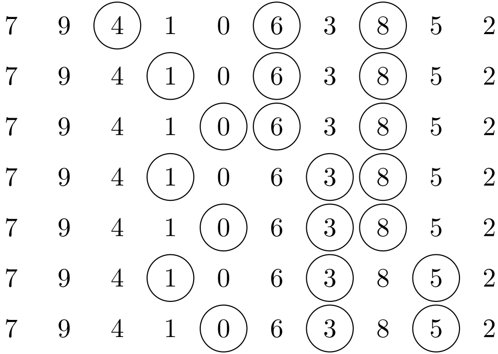




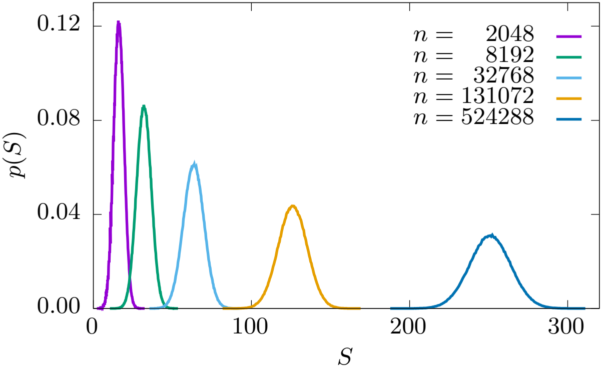
Counting Longest Increasing Subsequences

Number of different LIS grows exponentially
⇒ We can not just enumerate, we have to count cleverly.

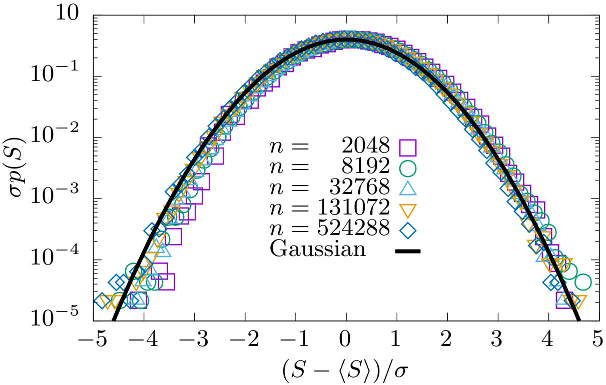




Distribution of the Entropy



Distribution of the Entropy



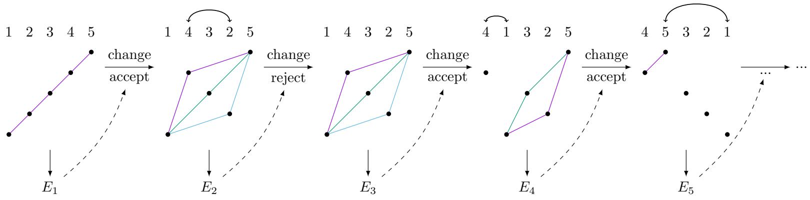
$$\langle S \rangle \approx 0.347\sqrt{n}$$

$$\sigma \approx 0.49\sqrt[4]{n}$$

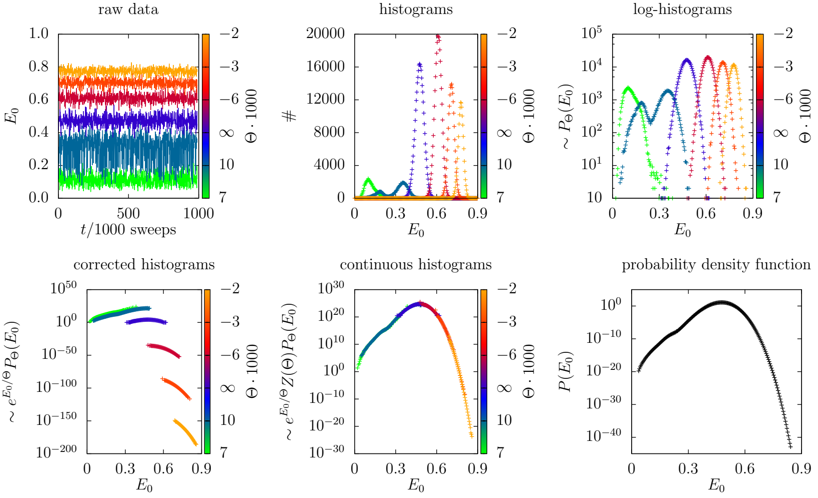
Markov Chain Monte Carlo for the Far Tails

- treat it as a canonical ensemble, i.e., weights $\sim e^{-E/\Theta}$
- artificial temperature Θ for the disorder ($\boldsymbol{\varepsilon}$)
- Markov chain of realizations $\boldsymbol{\varepsilon} = (\varepsilon_1, .., \varepsilon_N)$
- accept change with probability

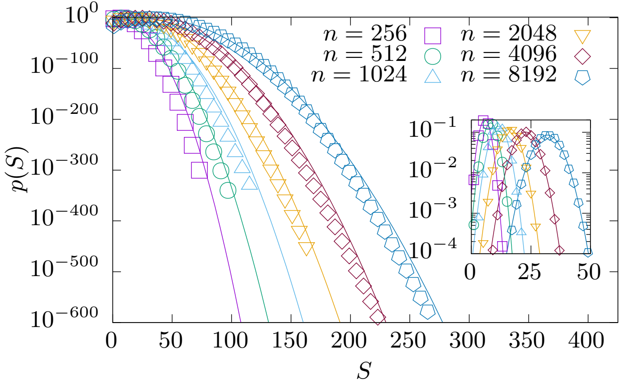
$$p_{\mathrm{acc}} = \min\left\{1, e^{-\Delta E_0/\Theta}\right\}$$



Markov Chain Monte Carlo for the Far Tails

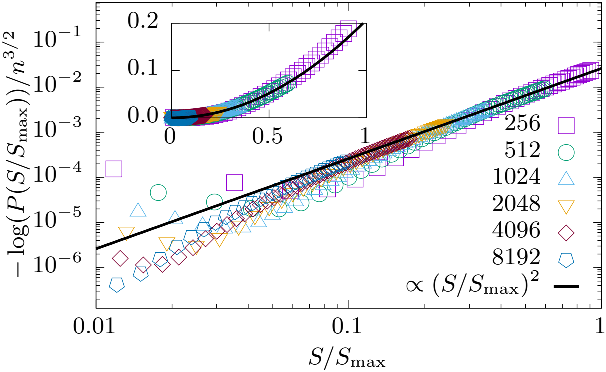


Full Distribution of the Entropy



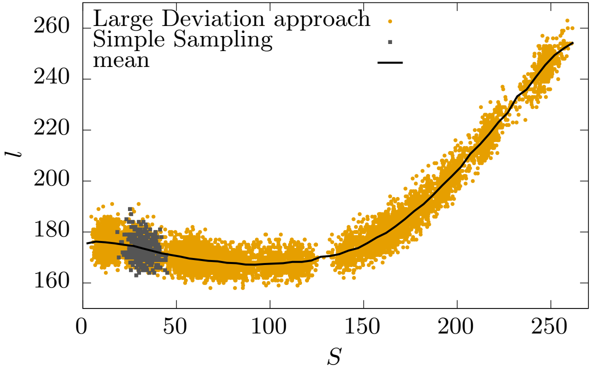
Deviations from Gaussians in the far tail

Full Distribution of the Entropy



Collapse of the far tail onto a rate function $\sim n^2$

Are Length and Entropy correlated?



Conclusions

- For Permutations
  - entropy scales as $S \approx \exp(0.347\sqrt{N})$
  - entropy distribution is well approximated by a Gaussian scaling form
  - far tails decay faster than Gaussian
  - hints for an unusual rate function