Expected Entropy of Random Range-Min&Max

Sebastian Wild*

November 14, 2022

Let $\mathcal{S}_n$ denote the set of permutations from $[n] = \{1, \ldots, n\}$ to $[n]$. Let $\mathcal{T}_n$ denote the set of binary trees on n vertices, i.e., trees where each node has a left and a right child; each of which could be the empty tree $\Lambda \in \mathcal{T}_0$.

Define $\mathrm{minTree}(x_1, \ldots, x_n)$ recursively as follows. $\mathrm{minTree}() = \Lambda$. $\mathrm{minTree}(x_1, \ldots, x_n)$ for $n \geq 1$ is a new root with $\mathrm{minTree}(x_1, \ldots, x_i)$ and $\mathrm{minTree}(x_{i+1}, \ldots, x_n)$ as left resp. right subtree, where $i = \arg\min_{1 \leq j \leq n} x_j$. maxTree is defined similarly using $i = \arg\max_{1 \leq j \leq n} x_j$.

We write $\mathrm{minTree}(\pi)$ for $\mathrm{minTree}(\pi(1), \pi(2), \ldots, \pi(n))$ for $\pi \in \mathcal{S}_n$.

The connection between minTree and range-minimum queries is explained in detail here *https://www.wild-inter.net/publications/entropy-trees*.

Warmup: Range-min only

For $T \in \mathcal{T}_n$, we define

$$p(T) = \frac{|\{\pi \in \mathcal{S}_n : \mathrm{minTree}(\pi) = T\}|}{n!}.$$

Goal: What is the entropy of the distribution over $\mathcal{T}_n$

$$H_1(n) = \sum_{T \in \mathcal{T}_n} p(T) \log_2(1/p(T)) = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \log_2(1/p(\mathrm{minTree}(\pi)))?$$

In this case, one can express $p(T)$ explicitly as the product of reciprocals of subtree sizes

$$p(T) = \prod_{v \in T} \frac{1}{\mathrm{nrDescendents(v)}};$$

inserting this into the sum above, one can obtain a recurrence relation for $H_1(n)$:

$$H_1(0) = H_1(1) = 0 \tag{1}$$

$$H_1(n) = \lg n + \frac{1}{n} \sum_{i=1}^{n} (H_1(i-1) + H_1(n-i)), \qquad (n \geq 2). \tag{2}$$

*University of Liverpool, UK, `wild@liverpool.ac.uk`

Kieffer, Yang and Szpankowski [3] resp. Hwang and Neininger [2] shows that this solves to

$$\begin{aligned} H_1(n) &= \lg(n) + 2(n+1)\sum_{i=2}^{n-1} \frac{\lg i}{(i+2)(i+1)} \\ &\sim 2n\sum_{i=2}^{\infty} \frac{\lg i}{(i+2)(i+1)} \\ &\approx 1.7363771n \end{aligned}$$

Open Problem: Range-min and max

Here, for $T_{\min}, T_{\max} \in \mathcal{T}_n$, define

$$p(T_{\min}, T_{\max}) = \frac{|\{\pi \in \mathcal{S}_n : \text{minTree}(\pi) = T_{\min} \text{ and } \text{maxTree}(\pi) = T_{\max}\}|}{n!}.$$

What is

$$H_2(n) = \frac{1}{n!} \sum_{T \in \mathcal{T}_n} \log_2(1/p(\text{minTree}(\pi), \text{maxTree}(\pi)))?$$

A great result would be a (somewhat) explicit form for $p(T_{\min}, T_{\max})$.

What is known.. One can uniquely construct from $(T_{\min}, T_{\max})$ a Baxter permutation π so that $(T_{\min}, T_{\max}) = (\text{minTree}(\pi), \text{maxTree}(\pi))$ [1]. Hence $H_2(n) \le \lg|\text{Baxter}_n| \sim 3n$.

Empirically, we should have $H_2(n) \approx 2.64n$, (I am not sure if the accuracy of that estimate really is two decimal places; it's expensive to sample).

References

[1] Paweł Gawrychowski and Patrick K. Nicholson. Optimal encodings for range top-k, selection, and min-max. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 593–604, 2015. doi:10.1007/978-3-662-47672-7_48.

[2] Hsien-Kuei Hwang and Ralph Neininger. Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM Journal on Computing*, 31(6):1687–1722, jan 2002. doi:10.1137/s009753970138390x.

[3] John C. Kieffer, En-Hui Yang, and Wojciech Szpankowski. Structural complexity of random binary trees. In *2009 IEEE International Symposium on Information Theory*. IEEE, jun 2009. doi:10.1109/isit.2009.5205704.