Please e-mail padayasi.1@osu.edu if you were able to prove ~~t~~ either conjectures! No prizes, just curious ☺

Sorting Problem

Jaychandran Padayasi

October 2021

1 Statement

Consider a set of N unique ordered objects. The scoring algorithm an attempt at sorting the ordered objects is as such:

- For every object i in the correct spot in the sequence, award N points; $s_i = N$.
- For every object j that ends one spot away from its position in the ordered list, award $N - 1$ points; $s_j = N - 1$.
- (Generalizing the previous rule) For every object that ends k spots away from its correct position, award $N - k$ points.

The total score $\mathcal{S}$ for each permutation of the ordered objects is

$$\mathcal{S} = \sum_{i=1}^{N} s_i \tag{1}$$

2 Questions

- What is the range of attainable scores $\mathcal{S}$ as a function of N?
  - Conjecture: Placing the objects in the reverse order yields a score of

$$\mathcal{S}_{\min} = \begin{cases} \frac{N^2+1}{2}, & N \text{ odd} \\ \frac{N^2}{2}, & N \text{ even} \end{cases} \tag{2}$$

    and experimentally this seems to be the lowest achievable score based on performing many random shuffles. **However**, for low values of N it is easy to see that there is degeneracy, *i. e.* many configurations can have the lowest score. Is eq. 2 the lowest possible score? How to prove it?

- What is the expected score $\mathbb{E}[\mathcal{S}]$ for a random permutation? What is the variance? A simpler question may be what is $\lim_{N\to\infty} \mathbb{E}[\mathcal{S}(N)]$?
  - Using `np.random.shuffle()` 10000 times and calculating the average for high values of N gives the conjecture

$$\lim_{N\to\infty} \mathbb{E}[\mathcal{S}] = \frac{2N^2}{3}.$$

Solution to Jaychandran's Questions

Shi Feng

August 27, 2022

It is convenient to note that the deviation from the optimal score (N^2) is equivalent to the L_1 distance between a pair of points $\mathbb{1}, \sigma(\mathbb{1}) \in \mathbb{S}^N$. Here $\mathbb{1}$ denotes the optimal ordered list (i.e. a reference point on the N-sphere), and $\sigma(\mathbb{1})$ denotes permutations of $\mathbb{1}$ by σ, with σ an element in the S_N group. The distance reads:

$$d(\mathbb{1}, \sigma(\mathbb{1})) = \sum_i |i - \sigma(i)|, \quad i \in \mathbb{N} \tag{1}$$

To see why this is an equivalent form, we can imagine a 2D plane where the x axis represents the ordered list $\mathbb{1}$ and the y axis represents the ordered list $\sigma(\mathbb{1})$. The deviation from the optimal score (N^2) is then given by the $L_1(x)$ distance (the sum of L_1 distances only in the x coordinate). But this is equivalent to the $L_1(y)$ distance because of the symmetry about $y = x$. Therefore, the total deviation from the optimal score is equivalent to the sum of the pairwise difference between the two lists, which is the same as the L_1 distance between two points in a N-sphere with the constraint that all accessible points are generated by S_N group.

This distance is known in statistical and the computer science community as the *Spearman's Footrule Distance* and is well studied by Diaconis and Graham (1977). There the authors proved the following statistical property:

$$\mathbb{E}[d(\mathbb{1}, \sigma(\mathbb{1}))] = \frac{1}{3}N^2 + O(N) \tag{2}$$

$$\mathrm{Var}[d(\mathbb{1}, \sigma(\mathbb{1}))] = \frac{2}{45}N^3 + O(N^2) \tag{3}$$

which means the expectation of the score function is

$$\mathbb{E}(S) = N^2 - \mathbb{E}[d(\mathbb{1}, \sigma(\mathbb{1}))] = \frac{2}{3}N^2 - O(N) \tag{4}$$

which is consistent with your conjecture; and since the variance is attributed only to the variation in deviation, we have

$$\mathrm{Var}(S) = \frac{2}{45}N^3 + O(N^2) \tag{5}$$