Goro Sort

1 Problem

This problem has been given as one of the qualification problems of the 2011 edition of Google CodeJam.

2 Analysis

3 Partial Proof

We will assume for now that one optimal strategy is to hold down all elements except one cycle at each iteration, and we will prove that the expected number of hits to sort one cycle is simply the length of the cycle. Proving that this is indeed the optimal solution is a harder endeavor, and it is addressed in the contest analysis

Let $p(n)$ be the expected number of hits required to sort a random permutation of size n, and $q(n)$ be the expected number of hits to sort a cycle of size n. Let us first observe that $p(1) = q(1) = 0$.

3.1 $p(n)$

We consider a random permutation σ of size n. Let us focus on the first cycle of σ, i.e. the cycle containing the first element σ_1. Call that cycle c. For all permutations such that c has size k, the expected number of required hits is $q(k)+p(n-k)$ (hits required to sort the first cycle, and then hits required to sort the remaining elements, that form a permutation of size $n-k$). Therefore if we denote by $r(k)$ the probability that c has size k, we can write

$$p(n) = \sum_{k=1}^{n} r(k) \left(q(k) + p(n-k)\right) \tag{1}$$

Probability $r(k)$ can be computed by enumerating the permutations with first cycle of size k.

– The first cycle is determined by its first $k-1$ elements (the last element has to be 1). There are $P_{n-1}^{k-1} = \frac{(n-1)!}{(n-k)!}$ possible such cycles, where P_{n-1}^{k-1} is the number of $(k-1)$-permutations of $n-1$ elements (since each one of the first $k-1$ elements could be anything but 1)
– The remaining $n-k$ elements can form any one of $(n-k)!$ possible permutations

therefore, the total number of permutations with first cycle of length k is $\frac{(n-1)!}{(n-k)!}(n-k)! = (n-1)!$. And the probability we are looking for is

$$r(k) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

We can then rewrite equation 1 as

$$\begin{aligned}
p(n) &= \sum_{k=1}^{n} \frac{1}{n}(q(k) + p(n-k)) \\
n.p(n) &= \sum_{k=1}^{n} q(k) + \sum_{k=1}^{n} p(n-k) \\
n.p(n) &= \sum_{k=1}^{n} q(k) + \sum_{k=0}^{n-1} p(k) \\
n.p(n) &= \sum_{k=2}^{n} q(k) + \sum_{k=2}^{n-1} p(k) \qquad (2)
\end{aligned}$$

since $p(0) = p(1) = q(1) = 0$

3.2 $q(n)$

To sort a random cycle of size n, one has to perform a first hit, the result will be a random permutation of size n, and the expected number of remaining hits required to sort that permutation is simply $p(n)$. Therefore $q(n) = 1 + p(n)$. Note that this only holds for $n > 1$, since for $n = 1$ we observed that $p(1) = q(1) = 0$. We can now rewrite 2

$$\begin{aligned}
n.p(n) &= \sum_{k=2}^{n} q(k) + \sum_{k=2}^{n-1} p(k) \\
&= (n-1) + \sum_{k=2}^{n} p(k) + \sum_{k=2}^{n-1} p(k) \\
&= (n-1) + p(n) + 2\sum_{k=2}^{n-1} p(k)
\end{aligned}$$

then

$$\begin{aligned}
(n-1)p(n) &= (n-1) + 2\sum_{k=2}^{n-1} p(k) \\
p(n) &= 1 + \frac{2}{n-1}\sum_{k=2}^{n-1} p(k) \qquad (3)
\end{aligned}$$

3.3 $p(n) = n - 1$

Let us show by induction that $\forall n \geq 1$, $p(n) = n - 1$. This is true for $n = 1$ since $p(1) = 0$. Now let $n \geq 2$, and assume that $\forall k < n$ $p(k) = k - 1$. Using equation 3 we can write

$$\begin{aligned}
p(n) &= 1 + \frac{2}{n-1}\sum_{k=2}^{n-1} p(k) \\
&= 1 + \frac{2}{n-1}\sum_{k=2}^{n-1} k - 1 \\
&= 1 + \frac{2}{n-1}\frac{(n-2)(n-1)}{2} \\
&= n - 1
\end{aligned}$$

which completes the proof.