Expected cost of the Move-To-Front heuristic

Recall that if you know the probability distribution of searches, you know the optimal arrangement of your linked-list. If $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$, then the optimal arrangement is k_1 then k_2 and so on. The expected number of comparisons for the optimal arrangement is

$$C_{\mathrm{OPT}} = \sum_{j=1}^{n} j\, p_j.$$

Theorem 1. *Let C_{MTF} be the expected number of comparisons of self-organizing search using the move-to-front heuristic.*

Then

$$C_{\mathrm{MTF}} \leqslant 2\, C_{\mathrm{OPT}}.$$

Proof. Let k_1, k_2, ..., k_n be the set of keys, ordered by decreasing probability of search, *i.e.* $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$. Assume that we are in a steady state : roughly, every key has been searched for at least once. At this time, your keys could be in any order depending on the random lookups that you did before.

We know that

$$C_{\mathrm{MTF}} = \sum_{j=1}^{n} (\text{Expected}\,\#\,\text{comparisons to find}\, k_j)\; p_j.$$

The expected number of comparisons to find k_j is (the number of keys k_i before k_j) plus one, that is

$$\text{Expected}\,\#\,\text{comparisons to find}\, k_j = 1 + \left(\sum_{i \neq j} \mathrm{Prob}(k_i\,\text{is before}\,k_j) \right).$$

Lemma 2. $\mathrm{Prob}(k_i\,\text{is before}\,k_j) = \frac{p_i}{p_i + p_j}$.

Proof. If k_i is before k_j in the linked list, it means that it has been looked for more recently. The corresponding sequences of last searches can be any of the following:

1. last search is k_i
   Probability: p_i
2. last searches: One key k distinct from k_i and k_j then k_i
   Probability: $p_i\,(1 - p_i - p_j)$
3. last searches: Two keys distinct from k_i and k_j then k_i
   Probability: $p_i\,(1 - p_i - p_j)^2$
4. And so on and for forth with any number of key distinct from k_i and k_j

So our probability is

$$p_i + p_i\,(1 - p_i - p_j) + p_i\,(1 - p_i - p_j)^2 + \cdots = p_i \sum_{k \geqslant 0} (1 - p_i - p_j)^k = \frac{p_i}{p_i + p_j}.$$

□

Using Lemma 2, we get that

$$
\begin{aligned}
C_{\mathrm{MTF}} &= \sum_{j=1}^{n}\left(1+\left(\sum_{i\neq j}\frac{p_i}{p_i+p_j}\right)\right)p_j \\
&= \sum_{j=1}^{n} p_j + \sum_{j=1}^{n}\sum_{i\neq j}\frac{p_i p_j}{p_i+p_j} \\
&= 1 + \sum_{1\leqslant i\neq j\leqslant n}\frac{p_i p_j}{p_i+p_j} \\
&= 1 + 2\sum_{1\leqslant i<j\leqslant n}\frac{p_i p_j}{p_i+p_j} \\
&= 1 + 2\left(\sum_{j=1}^{n}\left(p_j\sum_{i<j}\frac{p_i}{p_i+p_j}\right)\right) \\
&\leqslant 1 + 2\left(\sum_{j=1}^{n}\left(p_j\sum_{i<j}1\right)\right) \\
&= 1 + 2\left(\sum_{j=1}^{n} p_j\,(j-1)\right) \\
&= 2\,C_{\mathrm{OPT}} - 1 \\
&\leqslant 2\,C_{\mathrm{OPT}}
\end{aligned}
$$

$\square$