I.

Mean number of symbol comparisons for the Algorithm `InsertionSort`

We recall that the mean number $\widetilde{S}(Z)$ of symbol comparisons for an algorithm $\mathcal{A}$ is

$$\widetilde{S}(Z) = \int_{\mathcal{T}} [\gamma(u,t) + 1]\, \widetilde{\pi}_Z(u,t)\, du\, dt$$

where $\mathcal{T} := \{(u,t), \quad 0 \le u \le t \le 1\}$ is the unit triangle

$\gamma(u,t)$:= coincidence between $M(u)$ and $M(t)$

$\widetilde{\pi}_Z(u,t)\, du\, dt$:= Mean number of key-comparisons between $M(u')$ and $M(t')$ with $u' \in [u, u+du]$ and $t' \in [t-dt, t]$ performed by the algorithm $\mathcal{A}$.

An (easy) alternative expression for $\widetilde{S}(Z)$

$$\widetilde{S}(Z) = \sum_{w \in \Sigma^\star} \int_{\mathcal{T}_w} \widetilde{\pi}_Z(u,t)\, du\, dt$$

It involves the fundamental triangles

and separates the rôles of the source and the algorithm.

Question 1. Compute $\widetilde{\pi}_Z(u,t)$ for the algorithm `InsertionSort`.

Does $\widetilde{\pi}_Z(u,t)$ depend on the pair (u,t)? Such an algorithm is called an isotropic algorithm.

Question 2. Compute the mean number of symbol comparisons $\widetilde{S}(Z)$ for `InsertionSort` (in the Poisson model).

Question 3. Compute the mean number of symbol comparisons $S(n)$ for `InsertionSort` (in the Bernoulli model),

– first for a general source,

– then for a general memoryless source with probabilities $(p_1, p_2, \ldots, p_r)$.

Mean Path Length for a BST with repeated keys.

Consider an ordered alphabet $\sigma = \{a_1, a_2, \ldots, a_r\}$ and consider a slice s formed with N_i symbols equals to a_i and a total number of symbols N.

Question 1. Prove that the path length δ_A of the BST built on the slice s equals

$$\delta_A = \sum_{i \in \Sigma} N_i \sum_{j \neq i} \mathbf{1}_{[a_j \text{ ancestor of } a_i \text{ in bst}]}$$

Question 2. Denote by $\underline{\delta}_A$ the parameter defined by

$$\underline{\delta}_A(s) := \frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \delta_A(\tau(s))$$

where $\mathfrak{S}_N$ is the set of permutations on $[1..N]$.

Prove that the expectations of δ_A and $\underline{\delta}_A$ are the same in the Poisson model $\mathcal{P}_z$ and prove the equality

$$\underline{\delta}_A = 2 \sum_{i \in \Sigma} N_i \sum_{i < j} \frac{N_j}{N_{[i,j]}} = 2 \sum_{i \in \Sigma} \sum_{i < j} \frac{N_i N_j}{N_i + N_j + N_{]i,j[}}$$

Question 3. Prove that the expectation of δ_A in the Poisson model $\mathcal{P}_z$ is

$$\widetilde{\mathbb{E}}_z[\delta_A] = 2 \sum_{\substack{(i,j)\in\Sigma^2 \\ i<j}} \frac{p_i\, p_j}{{P_{[i,j]}}^2} \left[e^{-zP_{[i,j]}} - 1 + zP_{[i,j]}\right]$$

Question 4. Compute the mean path length of a `BST` with repeated keys in the Bernoulli model and prove the equality

$$\mathbb{E}_n[\delta_A] = 2 \sum_{\substack{(i,j)\in\Sigma^2 \\ i<j}} \frac{p_i\, p_j}{P_{[i,j]}} \; n \log n$$