This is a supplementary material showing the analysis of Randomized Quicksort, which is denoted RQS. Here, our goal is to show that for *any* input, the running time of RQS is $O(n \log n)$ in expectation. What this means is the following. Let $T(I, r)$ denote the RT of RQS for input I with a sequence of random numbers r – if you recall RQS, it needs access to some random numbers to randomly choose the pivot index, and therefore, RQS is nothing but a deterministic quicksort with access to random numbers r. The formal statement we want to prove is the following:

$$\text{For any input } I \text{ of } n \text{ elements, we have, } E_r[T(I, r)] = O(n \log n)$$

Note that the expectation is taken over all possible random numbers r, *not over inputs.*

To show the claim, for simplicity, let's assume that all numbers are distinct. Further, as RQS is comparison based, it is wlog to assume that we have $1, 2, \ldots, n$, instead of n arbitrary numbers, $a_1 < a_2 < \ldots < a_n$. Let X_{ij} denote an indicator random variable that has value 1 if i is compared to j, and 0 otherwise. *We leave it as an exercise to show that the running time is asymptotically upper bounded by the number of comparisons made.* Knowing that a pair of distinct numbers i and j are compared at most once, our goal becomes to bound $E \sum_{i<j} X_{ij}$, which is equal to $\sum_{i<j} EX_{ij}$ due to the linearity of expectation; if you forgot this, it's time to revisit your probability textbook.

So what is the value of $EX_{ij} = Pr[X_{ij} = 1]$? Initially, i and j, in fact all numbers, belong to the same single partition. And when a pivot is applied, it is partitioned into two partitions—three if we also count the pivot itself as a singleton partition. At the end of day, we will have n singleton partitions which are sorted.

Note that at any point of time in the execution, a partition is a set of consecutive integers. That is, if a partition has $i < j$ and it must include all integers k such that $i < k$ and $k < j$.

Fix a pair of integers $i < j$. As discussed above, they belong to the same partition at the beginning and later separated into different partitions at a certain stage. Let's consider the last partition B where i and j both lie. Since i and j are separated at this moment, the pivot must be chosen from $\{i, i+1, \ldots, j\}$. Further, we observe that i and j are compared to each other, i.e. $X_{ij} = 1$, if and only if either i or j is chosen as the pivot, which occurs with probability $\frac{2}{j-i+1}$ because the pivot is chosen uniformly at random.

Thus, $\sum_{i<j} EX_{ij} = \sum_{1 \leq i < j \leq n} \frac{2}{j-i+1} \leq \sum_{1 \leq i \leq n} 2(\frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{n}) = O(n \log n)$.

Here we used a well-known fact that $\frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{n} \leq 1 + \int_{x=1}^{n} \frac{1}{x} dx = 1 + \ln n$