RANDOMIZED QUICK SORT: A COMPLETE ANALYSIS

Kapu Nirmal Joshua **Aravind Seshadri** **Aatman Jain**
Indian Institute of Technology Kanpur
{nirmalj21, aravinds21, aatman21}@iitk.ac.in

ABSTRACT

This paper provides a detailed explanation of the proof from "Strong Concentration for Quicksort" by Colin McDiarmid and Ryan Hayward (1), which establishes the tightest known bound on the deviation of the comparison count in basic Randomized QuickSort from its expected value. Using advanced concentration inequalities and a careful analysis of the recursion tree, the original proof shows that the probability of significant deviations decreases exponentially. We also conducted experiments to assess how closely the actual deviation probabilities align with the theoretical bounds and to measure the extent of deviation from the mean in Randomized QuickSort. All experimental code and plots are available on GitHub at `https://github.com/kn-joshua/CS648-Randomized-QuickSort`.

1 INTRODUCTION

Randomized QuickSort is a widely studied sorting algorithm that achieves an expected comparison count of approximately $2n \ln n$ when sorting a list of n distinct keys by selecting pivots uniformly at random. A key question in its probabilistic analysis is how tightly the actual number of comparisons concentrates around this expected value. In their seminal work, "Strong Concentration for Quicksort" (1), Colin McDiarmid and Ryan Hayward derive the tightest known bound on this deviation. This paper aims to provide a clear and comprehensive explanation of their proof for the basic Randomized QuickSort algorithm.

To achieve this, we break down the original proof, which leverages the method of bounded differences and martingale-based concentration inequalities, into accessible steps. Additionally, we ran experiments to evaluate how closely the actual deviation probabilities match the theoretical exponential bound and to quantify the typical deviations from the mean. The paper is organized as follows: Section 3 describes the Randomized QuickSort algorithm and defines the notation used. Section 4 outlines the experimental setup and results. Section 5 presents the detailed proof, including supporting lemmas. Section 6 concludes with a summary of the findings.

2 PREVIOUS WORK

The study of QuickSort's performance began with Hoare's introduction of the algorithm (6), laying the groundwork for subsequent analyses. Early efforts by Sedgewick (5) established bounds on the variance of the comparison count in QuickSort, suggesting that deviations are manageable but not sharply characterized. Roura and Vallejo (3) later refined these bounds using generating functions, achieving concentration results that, while improved, did not reach the tightest possible precision. Fill and Janson (4) applied branching process techniques to derive subexponential tail bounds for the comparison count, advancing the understanding of QuickSort's probabilistic behavior. Devroye (2) analyzed the structure of binary search trees, providing insights into the recursion tree's height, which relates to QuickSort's performance but does not directly address concentration. The tightest bound, however, was established by McDiarmid and Hayward (1), whose proof of exponential decay in deviation probabilities forms the focus of this paper.

Algorithm 1 Randomized QuickSort

```
procedure RandomizedQuickSort(A, p, r)
    if p < r then
        q ← RandomPartition(A, p, r)
        RandomizedQuickSort(A, p, q - 1)
        RandomizedQuickSort(A, q + 1, r)
    end if
end procedure

procedure RandomPartition(A, p, r)
    i ← Random(p, r)                                    ▷ Select a random index between p and r
    Swap(A[i], A[r])
    x ← A[r]                                            ▷ Pivot element
    j ← p − 1
    for k ← p to r − 1 do
        if A[k] < x or (A[k] == x and k < i) then
            j ← j + 1
            Swap(A[j], A[k])
        end if
    end for
    Swap(A[j + 1], A[r])
    return j + 1
end procedure
```

3 Randomized QuickSort

3.1 Algorithm

The Randomized QuickSort algorithm refines the classic QuickSort by introducing randomness in pivot selection to ensure average-case efficiency regardless of input ordering. In this basic version, a single pivot is chosen uniformly at random from the subarray at each recursive step. The pseudocode above outlines this process succinctly, where RandomizedQuickSort recursively sorts the array, and RandomPartition handles the pivot selection and partitioning.

The algorithm operates as follows: given an array A and indices p and r defining a subarray, RandomPartition selects a random pivot, repositions it to the end, and partitions the remaining elements around it, returning the pivot's final position q. The subarrays $A[p..q-1]$ and $A[q+1..r]$ are then recursively sorted. This randomization ensures that no specific input consistently triggers the worst-case quadratic time, making the expected performance robust. Further, no additional space is required to store sorted arrays as in the case of merge sort.

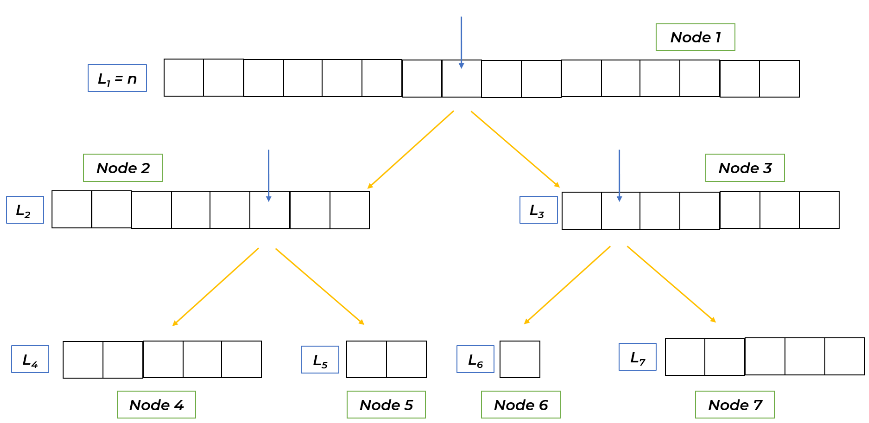

Figure 1: Notations and Terminologies Definition

3.2 Notation and Terminology

To analyze Randomized QuickSort rigorously, we establish a consistent framework for the variables and concepts used in our lemmas and proofs, aligning with the notation in McDiarmid and Hayward (2).

Consider an infinite binary tree where nodes are numbered level by level, left to right, starting with the root as node 1, its left child as node 2, its right child as node 3, and so forth. Each execution of Randomized QuickSort labels a finite subtree of this structure, with each node corresponding to a sublist processed during recursion. The root, node 1, is assigned the initial unsorted list of n distinct keys, and its list length is denoted $L_1 = n$. For any node j, its left child (node $2j$) receives the sublist of keys less than the pivot chosen at node j, while its right child (node $2j + 1$) receives the sublist of keys greater than the pivot. The list length L_j represents the number of keys in the sublist at node j, capturing the size of each recursive call. Refer to Figure 1 for the visualization of notations.

The recursion tree's structure allows us to study list lengths at specific depths. For a given depth k, the nodes at that level are numbered from 2^k to $2^{k+1} - 1$, and we define $M_k = \max\{L_{2^k+i} : i = 0, 1, \ldots, 2^k - 1\}$ as the maximum list length among the 2^k nodes at depth k. This quantity is crucial for bounding the worst-case behavior within a level. We also track the algorithm's performance through comparison counts. Let Q_n denote the total number of comparisons performed by Randomized QuickSort on a list of n keys, with its expectation $q_n = \mathbb{E}[Q_n]$ known to be approximately $2n \ln n$.

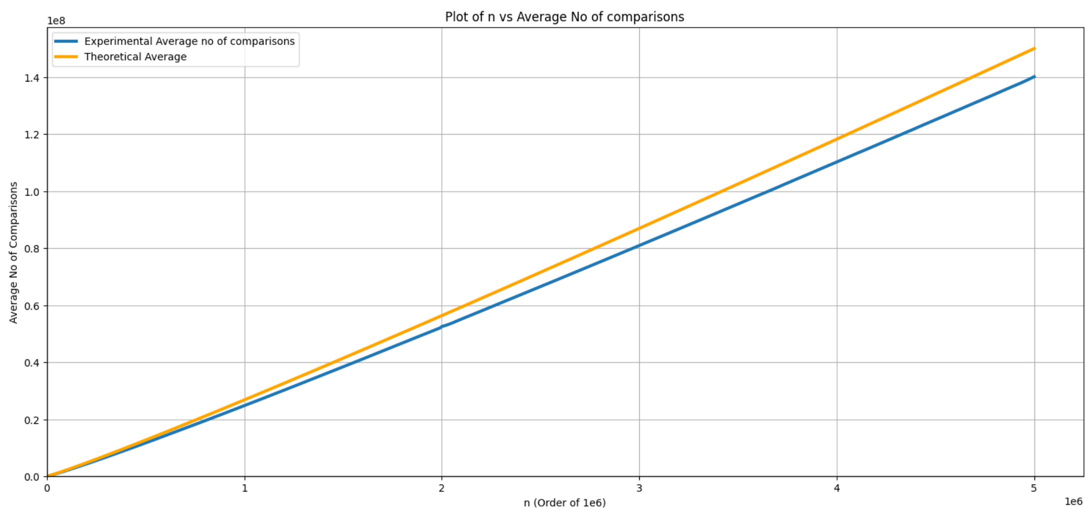

Figure 2: Exp. Average vs Theoretical Average

To dissect the recursive process, we introduce H_k, the random "history" of comparisons executed at level k of the recursion tree, encapsulating the decisions made at that depth. The k-history, $\mathcal{H}^{(k)} = (H_0, H_1, \ldots, H_{k-1})$, aggregates the comparison history across the first k levels, providing a cumulative view of the algorithm's progression.

These definitions—node numbering, list lengths L_j, maximum depth-specific length M_k, comparison count Q_n, expected comparisons q_n, level history H_k, and k-history $\mathcal{H}^{(k)}$—form the backbone of our analysis. They enable precise statements in lemmas 2.1 through 2.9, which collectively build toward the tightest bound on the deviation of Q_n from q_n, as detailed in subsequent sections.

4 Experiments

4.1 Experimental Details

We perform an experimental analysis to evaluate the deviation of randomized quicksort runtimes from their mean. First, the mean runtime is computed over a set of trials. Then, for a given percentage threshold, we count the number of trials whose runtimes exceed the mean by that percentage. Dividing this count by the total number of trials yields the empirical probability of such deviations. The experiment is conducted using the parameters summarized in Table 1.

Input Array	Random Permutation of $\{1, 2, \ldots, n\}$
Array Size n	500 to 5,000,000
Step Size	Varied between 500 to 2000
Trials	5000 overall
Deviation ε	1% to 100%

Table 1: Experimental setup

4.2 Results

4.2.1 Average no of comparisons

Figure 2 shows the plot of input size n versus the average number of comparisons made by randomized quicksort. The experimental average, computed over multiple trials, is plotted alongside the theoretical expectation given by $2n \log n + \mathcal{O}(n)$. As expected, the number of comparisons grows approximately linearly with $n \log n$. This trend highlights the relatively slow growth of the $\log n$ factor compared to n, making it appear almost constant even for large input sizes such as 5 million. The experimental curve closely follows the theoretical one, with a slight underestimation. This minor deviation can be attributed to constant factors and lower-order terms in practical implementations, which are typically ignored in asymptotic analysis. Additionally, increasing the number of trials could help the experimental average converge more closely to the theoretical value.

4.2.2 Probability of deviation

Figures 3 and 4 illustrate the probability of the number of comparisons exceeding the mean by 5% and 10%, respectively, as a function of input size n. The experimental probability is computed by running multiple trials and counting how often the number of comparisons crosses the specified deviation threshold. This is compared against the theoretical probability, which is expected to decrease exponentially as n increases due to the concentration of measure.

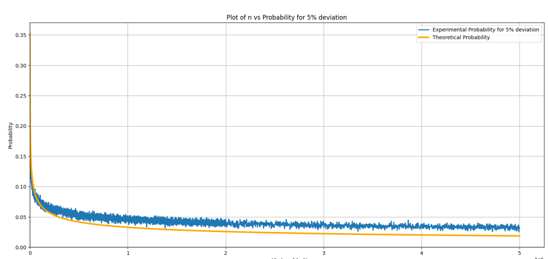

Figure 3: Exceeding Mean by 5%

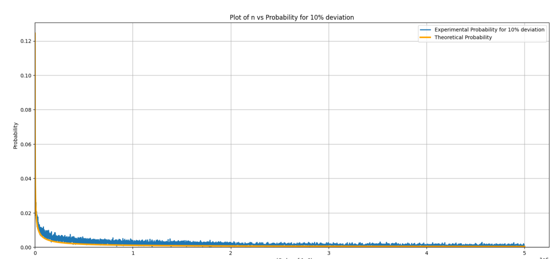

Figure 4: Exceeding Mean by 10%

As observed in both plots, the probability of deviation decreases with increasing input size. The experimental curves are nearly bounded by the theoretical predictions. We notice a decrease in the probability values from 5% to 10%. This conveys to us the reliability of quicksort in practical scenarios. Despite it having situations where the time taken is large, the probability of this happening—especially for large values of n—is very small.

Further, we have plotted a probability heatmap in Figure 5, showing the percentage deviation of quicksort from its mean with respect to the size of the array n. The percentage deviation ranges from 1 to 15%; beyond this, the deviation is nearly negligible. We observe that the high-probability regions, shown by the colour yellow/green, are concentrated between 1–3%, and reduce beyond this, denoted by blue/black—especially for large n values. This conveys the fact that quicksort mostly remains concentrated around the mean.

4.2.3 Distribution of comparisons

In Figure 6, we plot a histogram of the actual number of comparisons. We chose the value of $n = 30000$ and performed the experiment over 100000 trials. We observe that the mode of the distribution is slightly shifted to the left of the mean. Additionally, the percentage of trials in which the number of comparisons exceeded a 10% deviation is less than 0.25%. This tells us that quicksort is highly concentrated around the mean.

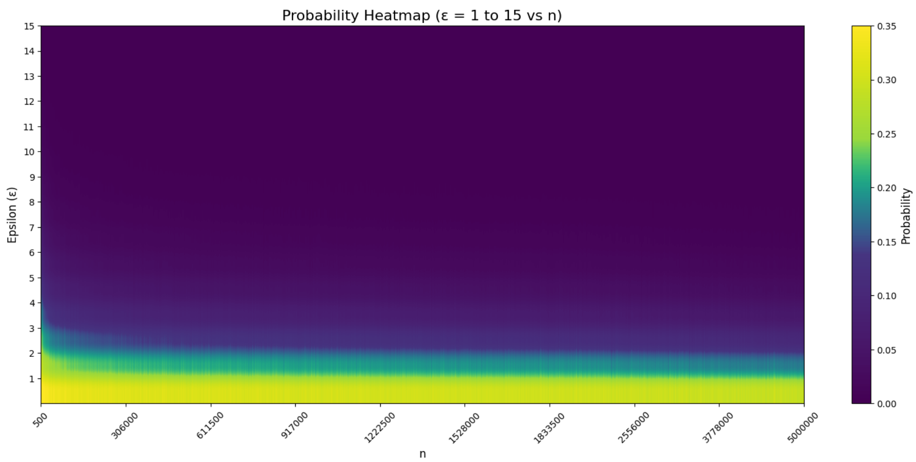

Figure 5: Probability Heatmap showing the Distribution

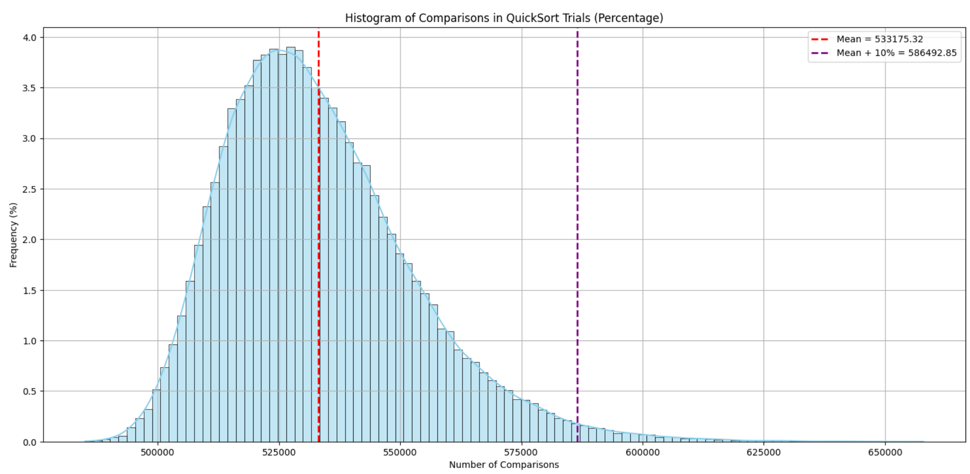

Figure 6: Frequency of Comparisons

5 Algorithm Analysis

This section derives the concentration (1) of Q_n around q_n using a sequence of lemmas. We first bound subproblem sizes (Lemma 2.1), then analyze conditional expectations (Lemmas 2.2–2.3), establish martingale properties (Lemmas 2.4–2.6), and synthesize these into tight probability bounds (Lemmas 2.7–2.9). The culmination is an exponential tail bound, refining prior results.

Lemma 2.1

Let M_k^n be the maximum subproblem size at depth k in the recursion tree of *basic* Quicksort acting on n distinct keys. Formally, if L_j is the size of the sublist at node j, then

$$M_k^n = \max\{ L_{2^k+i} : i = 0, 1, \ldots, 2^k - 1 \}$$

For a fixed real $0 < \alpha < 1$ and any integer $k \geq \ln(1/\alpha)$, we claim:

$$\Pr\big(M_k^n \geq \alpha\, n\big) \leq \alpha \left(\frac{2\, e\, \ln(\frac{1}{\alpha})}{k} \right)^k$$

Proof. We couple the random sizes of Quicksort's subproblems with partial products of independent $\mathrm{Uniform}(0,1)$ variables. Intuitively, each time Quicksort picks a pivot in a subproblem, it is as if we choose a fraction X_i of that subproblem's size to go "left," and $1 - X_i$ to go "right." By chaining these random splits level by level, one can show that the event $M_k^n \geq \alpha n$ is controlled by the maximum among 2^k products of i.i.d. $\mathrm{Uniform}(0,1)$ variables.

Let $X_1, X_2, \dots$ be i.i.d. $\mathrm{Uniform}(0,1)$ random variables, and build a sequence $(\widetilde{L}_j)$ by setting

$$\widetilde{L}_1 = n, \quad \widetilde{L}_{2i} = \lfloor X_i \, \widetilde{L}_i \rfloor, \quad \widetilde{L}_{2i+1} = \lfloor (1 - X_i) \, \widetilde{L}_i \rfloor, \quad i \geq 1$$

We know that, $(L_1, L_2, \dots)$ and $(\widetilde{L}_1, \widetilde{L}_2, \dots)$ have the *same* joint distribution. This is because both of them convey the same meaning of fraction/no of elements in each sublist, as mentioned above. Hence

$$\Pr\bigl(M_k^n \geq \alpha n\bigr) = \Pr\bigl(\widetilde{M_k^n} \geq \alpha n\bigr), \quad \text{where} \quad \widetilde{M_k^n} = \max_{\text{depth}=k} \widetilde{L}_j$$

Next define $Z_1 = 1$ and for $i \geq 1$,

$$Z_{2i} = X_i \, Z_i, \quad Z_{2i+1} = (1 - X_i) \, Z_i$$

Then each $\widetilde{L}_j$ is bounded by $n \, Z_j$. Thus

$$\widetilde{M_k^n} \;=\; \max_{j:\,\text{depth}(j)=k} \widetilde{L}_j \;\leq\; n \max_{j:\,\text{depth}(j)=k} Z_j \;=\; n \, Z_k^*,$$

where Z_k^* is the maximum Z_j at depth k. Therefore,

$$\{\widetilde{M_k^n} \;\geq\; \alpha n\} \;\subseteq\; \{n \, Z_k^* \;\geq\; \alpha \, n\} \;=\; \{Z_k^* \;\geq\; \alpha\}.$$

Since $\{E_1 \subseteq E_2\} \Rightarrow \Pr(E_1) \leq \Pr(E_2)$, we have

$$\Pr\bigl(M_k^n \geq \alpha n\bigr) = \Pr\bigl(\widetilde{M_k^n} \geq \alpha n\bigr) \leq \Pr\bigl(Z_k^* \geq \alpha\bigr)$$

It remains to bound $\Pr(Z_k^* \geq \alpha)$. At depth k, there are 2^k nodes/paths, each with a product of k random fractions X_i or $(1 - X_i)$. Thus

$$Z_k^* \;=\; \max_{1 \leq \ell \leq 2^k} \prod_{i=1}^{k} U_{\ell,i} \quad (U_{\ell,i} \text{ are uniform in } (0,1)).$$

Hence

$$\Pr\bigl(Z_k^* \geq \alpha\bigr) \;=\; \Pr\Bigl(\bigcup_{\ell=1}^{2^k} \{\prod_{i=1}^{k} U_{\ell,i} \geq \alpha\}\Bigr) \;\leq\; 2^k \, \Pr\Bigl(\prod_{i=1}^{k} U_i \geq \alpha\Bigr),$$

by a union bound. Next we note that if $U_i \sim \mathrm{Uniform}(0,1)$, then $-\ln U_i \sim \mathrm{Exp}(1)$, so

$$\prod_{i=1}^{k} U_i \;=\; \exp\Bigl(\sum_{i=1}^{k} \ln(U_i)\Bigr) \;=\; \exp\bigl(-\,G_k\bigr),$$

where $G_k = \sum_{i=1}^{k} (-\ln U_i) \sim \mathrm{Gamma}(k,1)$. Thus

$$\Pr\bigl(\prod_{i=1}^{k} U_i \geq \alpha\bigr) \;=\; \Pr\bigl(\,G_k \leq -\ln(\alpha)\bigr).$$

$$\begin{aligned}
\Pr\Bigl(\prod_{i=1}^{k} U_i \;\geq\; \alpha\Bigr) &= \Pr\Bigl(e^{-G_k} \;\geq\; \alpha\Bigr) \;=\; \Pr\Bigl(e^{-G_k}/\alpha \;\geq\; 1\Bigr) \;=\; \Pr\Bigl(e^{-tG_k}/\alpha^t \;\geq\; 1\Bigr) \quad (\text{for } t > 0) \\
&\leq\; \mathbb{E}\Bigl[e^{-t\,G_k}/\alpha^t\Bigr] \quad (\text{Markov's inequality}) \\
&=\; \frac{1}{\alpha^t} \, \mathbb{E}\bigl[e^{-t\,G_k}\bigr].
\end{aligned}$$

Since G_k is a sum of k i.i.d. $\mathrm{Exp}(1)$ random variables, it is $\Gamma(k,1)$, and thus

$$\mathbb{E}[\, e^{-t\, G_k} \,] = (1+t)^{-k}.$$

Hence

$$\Pr\Big(\prod_{i=1}^{k} U_i \;\geq\; \alpha\Big) \;\leq\; \frac{(1+t)^{-k}}{\alpha^t}.$$

Next, by a union bound over the 2^k paths (each having a product of k i.i.d. Uniform(0,1)),

$$\Pr\big(Z_k^* \;\geq\; \alpha\big) \;\leq\; 2^k \;\Pr\Big(\prod_{i=1}^{k} U_i \;\geq\; \alpha\Big) \;\leq\; 2^k\,\frac{(1+t)^{-k}}{\alpha^t}.$$

We choose t to minimize $2^k(1+t)^{-k}/\alpha^t$. By differentiation, we get t as $\frac{k}{\ln(1/\alpha)} - 1$, giving an upper bound of

$$\alpha \left(\frac{2\,e\,\ln\big(1/\alpha\big)}{k} \right)^{k}.$$

Thus

$$\Pr\big(Z_k^* \;\geq\; \alpha\big) \;\leq\; \alpha \left(\frac{2\,e\,\ln\big(\frac{1}{\alpha}\big)}{k} \right)^{k},$$

as claimed. □

LEMMA 2.2

Recall that $q_n = \mathbf{E}[Q_n]$ satisfies:

$$q_n \;=\; (n-1) \;+\; \sum_{j=1}^{n-1} (q_{\,j-1} \;+\; q_{\,n-j})$$

and we also know $q_n = 2\,n\,\ln(n) + O(n)$.

Define

$$A_n \;=\; \Big\{ n-1 \;+\; q_{k-1} \;+\; q_{n-k} \;-\; q_n \;:\; k = 1,2,\ldots,n \Big\}.$$

We claim that for every $x \in A_n$, we have $|x| \leq n$.

Proof First we rewrite the difference for each k as

$$f(k) \;=\; \Big(n-1\Big) \;+\; q_{\,k-1} \;+\; q_{\,n-k} \;-\; q_n \quad \text{for} \quad k = 1,2,\ldots,n.$$

Our goal is to show $f(k) \in [-\,n,\, +n]$, that is, $f(k)$ never exceeds $+\,n$ and never goes below $-\,n$.

To see why $f(k)$ cannot exceed $+\,n$, let us compare $f(k)$ to a simpler continuous function. Suppose we approximate each q_m by $C\,(m \ln m)$ for some constant $C > 0$ (the same C for all n), consistent with the fact that $q_m = O(m \ln m)$. Then define

$$g(x) \;=\; \big(n-1\big) \;+\; C\Big[\,(x-1)\,\ln(x-1) \;+\; (n-x)\,\ln(n-x) \;-\; n\,\ln(n)\Big] \quad \text{for} \quad x \in [2, n-1].$$

We differentiate $g(x)$ to locate extrema. For $x \in (2,n)$, write

$$g'(x) \;=\; C\left[\ln\big(x-1\big) + (x-1)\,\frac{1}{x-1} \;-\; \ln\big(n-x\big) - (n-x)\,\frac{1}{n-x}\right],$$

which simplifies to

$$g'(x) = C\left[\ln\big(\tfrac{x-1}{n-x}\big)\right].$$

Hence $g'(x) = 0$ exactly when $x - 1 = n - x$, i.e. $x = \frac{n+1}{2}$. By standard continuity arguments on $(1,n)$, $g(x)$ reaches its maximum either at $x = 1$, at $x = n$, or at this critical point. But $x = \frac{n+1}{2}$ in fact yields a minimum because $g''(x)$

is positive there, so the maximum must occur at the endpoints $x = 1$ or $x = n$. Translating back to discrete indices, we only need to check $k = 1$ and $k = n$ to bound $f(k)$ from above. When $k = 1$ or $k = n$, we have

$$f(k) \;=\; (n-1) \;+\; q_0 \;+\; q_{n-1} \;-\; q_n \;=\; n - 1 + q_{n-1} - q_n,$$

since $q_0 = 0$. Because $q_{n-1} \approx C\,(n-1)\ln(n-1)$ and $q_n \approx C\,n \ln n$, the difference $q_{n-1} - q_n$ is negative.Thus $\max_k f(k) \le n$.

We also need to show $f(k) \ge -n$ for all k. Because $g'(x) = 0$ at $x = \frac{n+1}{2}$ yields a minimum in the continuous analogy, we check $x = \frac{n+1}{2}$. Substituting $x = \frac{n+1}{2}$ into $g(x)$, we find that

$$g\Big(\tfrac{n+1}{2}\Big) \;=\; (n-1) \;+\; C\,\Big[\Big(\tfrac{n+1}{2}-1\Big)\ln\Big(\tfrac{n+1}{2}-1\Big) \;+\; \Big(n-\tfrac{n+1}{2}\Big)\,\ln\Big(n-\tfrac{n+1}{2}\Big) \;-\; n\,\ln(n)\Big].$$

Now simplifying the expression, we get:

$$\begin{aligned} g\Big(\tfrac{n+1}{2}\Big) &= \;(n-1) \;+\; C\,\big\{(n-1)\,\ln\big(\tfrac{n-1}{2}\big) \;-\; n\,\ln(n)\big\} \\ &= \;(n-1) \;+\; C\,\big\{n\,\ln\Big(\tfrac{n-1}{2n}\Big) \;-\; \ln\Big(\tfrac{n-1}{2}\Big)\big\} \\ &= \;n\Big\{1 \;+\; C\,\ln\Big(\tfrac{n-1}{2n}\Big)\Big\} \;-\; \Big\{1 \;+\; C\,\ln\Big(\tfrac{n-1}{2}\Big)\Big\}. \end{aligned}$$

We shall now take $C = 2$ for the rest of the proof, without loss of generality. Then:

$$g\Big(\tfrac{n+1}{2}\Big) \;=\; n\Big\{1 + 2\ln\Big(\tfrac{n-1}{2n}\Big)\Big\} \;-\; \Big\{1 + 2\ln\Big(\tfrac{n-1}{2}\Big)\Big\}.$$

Consider the continuous counterpart:

$$h(x) \;=\; x\Big\{1 + 2\ln\Big(\tfrac{x-1}{2x}\Big)\Big\} \;-\; \Big\{1 + 2\ln\Big(\tfrac{x-1}{2}\Big)\Big\}, \quad \text{defined for } x > 1, \text{ and } h(1) = 0.$$

Now define

$$\ell(x) \;=\; h(x) + x.$$

Differentiate:

$$\ell'(x) \;=\; h'(x) + 1,$$

and one finds that $\ell(x) = 0$ at some point $x_0 \approx 1.2, 9.9$, which can be computed by solving $\ell(x) = 0$. Checking the first derivative,$\ell(x)$ increases after $x \approx 10$, Hence

$$\ell(x) \;=\; h(x) + x \;>\; 0 \quad \text{for all } x \ge 10,$$

which implies

$$h(x) \;>\; -x.$$

Translating back to the discrete scenario, for $n > 10$, we have:

$$g\Big(\tfrac{n+1}{2}\Big) \;>\; -n$$

<u>Note:</u> We do not consider $n < 10$ as it is trivial, and we wish to be working in the limit of large n.

Equivalently, the minimum of $f(k)$ over $k = 1, 2, \ldots$ is strictly above $-n$ in the discrete setting.

Combining these observations, $f(k)$ lies in the interval $[-n, n]$ for all $k = 1, \ldots, n$. Hence every element x of

$$A_n \;=\; \{\, f(k) : k = 1, \ldots, n\}$$

satisfies $|x| \le n$. This completes the proof that $A_n \subseteq [-n, n]$. $\square$

LEMMA 2.3

Let n and k be positive integers, and let h be any possible k-history for Q_n. Then

$$\Big|\,\mathbf{E}\big[Q_n \;\big|\; \underline{H}^{(k)} = \underline{h}\big] \;-\; q_n\Big| \;\;\le\;\; k\,n.$$

Proof We prove this by induction on k.

Base Case: $k = 1$. When $k = 1$, the "history" $\underline{H}^{(1)}$ just corresponds to which pivot was chosen at the root. Suppose the pivot splits the array into subarrays of sizes $(j-1)$ and $(n-j)$. Then

$$\mathbf{E}\big[Q_n \;\big|\; \underline{H}^{(1)}\big] \;=\; (n-1) \;+\; q_{j-1} \;+\; q_{n-j}.$$

Subtracting q_n gives

$$\Big[(n-1) \;+\; q_{j-1} \;+\; q_{n-j}\Big] \;-\; q_n \;=\; x, \quad \text{say.}$$

From Lemma 2.2, we know that this difference x lies in $[-n, n]$. Hence

$$\big|\,\mathbf{E}[\,Q_n \mid \underline{H}^{(1)}] \;-\; q_n\big| \;\le\; n.$$

Thus the base case $k = 1$ is established.

Inductive Step. Assume the lemma is true for $k = t$, i.e. for any t-history $h^{(t)}$ and any n,

$$\big|\,\mathbf{E}[\,Q_n \;\mid\; \underline{H}^{(t)} = h^{(t)}] \;-\; q_n\big| \;\le\; t\,n.$$

We show it holds for $k = t+1$. Suppose $\underline{H}^{(t+1)} = h^{(t+1)}$ is a $(t+1)$-history: we then have an *additional* pivot choice at depth $t+1$ beyond what was known in $h^{(t)}$. Let us denote the subproblems formed at depth 1 by sizes $(j-1)$ and $(n-j)$. Then

$$\mathbf{E}\big[\,Q_n \;\big|\; \underline{H}^{(t+1)} = \underline{h}^{(t+1)}\big] \;=\; (n-1) \;+\; \mathbf{E}\big[\,Q_{j-1} \;\big|\; \underline{H}^{(t+1)}\big] \;+\; \mathbf{E}\big[\,Q_{n-j} \;\big|\; \underline{H}^{(t+1)}\big].$$

Next, by induction on the subproblems themselves which each have size at most n but *one* fewer level of new pivot selection to consider. This is because the information given by $H^{(1)}$ is used to split the array ,

$$\Big|\mathbf{E}[\,Q_{j-1} \mid \underline{H}^{(t+1)}] \;-\; q_{j-1}\Big| \;\le\; t\,(j-1), \qquad \Big|\mathbf{E}[\,Q_{n-j} \mid \underline{H}^{(t+1)}] \;-\; q_{n-j}\Big| \;\le\; t\,(n-j).$$

Now,

$$\big|\,\mathbf{E}[\,Q_{j-1} \mid \underline{H}^{(t+1)}] + \mathbf{E}[\,Q_{n-j} \mid \underline{H}^{(t+1)}] \;+\; \big(n-1\big) - q_n\big|$$

$$= \big|\,\mathbf{E}[\,Q_{j-1} \mid \underline{H}^{(t+1)}] + \mathbf{E}[\,Q_{n-j} \mid \underline{H}^{(t+1)}] \;+\; \big(n-1\big) - q_n + q_{j-1} + q_{n-j} - q_{j-1} - q_{n-j}\big|$$

Using triangle inequality,

$$\le \big|\,\mathbf{E}[\,Q_{j-1} \mid \underline{H}^{(t+1)}] - q_{j-1}\big| + \big|\mathbf{E}[\,Q_{n-j} \mid \underline{H}^{(t+1)}] - q_{n-j}\big| \;+\; \big|\big(n-1\big) - q_n + q_{j-1} + q_{n-j}\big|$$

Using lemma 2.2 and the other inequalities from above we have

$$\le t(j-1) + t(n-j) + n = (t+1)n - t \le (t+1)n$$

That completes the induction. Therefore,

$$\Big|\mathbf{E}[\,Q_n \mid \underline{H}^{(k)} = \underline{h}] \;-\; q_n\Big| \;\le\; k\,n$$

for all k and for any k-history h, as required. $\square$

Lemma 2.4

Let X be any real-valued random variable with $a \le X \le b$ almost surely and $\mathbb{E}[X] = 0$. Then for every $s \ge 0$,

$$\mathbb{E}\big[e^{\,sX}\big] \;\;\le\; \exp\!\Big(\tfrac{s^2\,(b-a)^2}{8}\Big).$$

Proof The function e^{sx} is convex for any real s. For a convex function, the chord connecting two points on the curve lies above or on the curve itself. Consider the points (a, e^{sa}) and (b, e^{sb}), where $a \le x \le b$. The equation of the line joining these points can be written as a linear interpolation:

$$e^{sx} \le e^{sa} \cdot \frac{b-x}{b-a} + e^{sb} \cdot \frac{x-a}{b-a}$$

This inequality holds for all x in the interval $[a, b]$ due to convexity.

Since X is a random variable with values in $[a, b]$, we take the expected value of both sides:

$$\mathbb{E}[e^{sX}] \le \mathbb{E}\left[e^{sa} \cdot \frac{b-X}{b-a} + e^{sb} \cdot \frac{X-a}{b-a}\right]$$

Using the linearity of expectation, this becomes:

$$\mathbb{E}[e^{sX}] \le e^{sa} \cdot \frac{b-\mathbb{E}[X]}{b-a} + e^{sb} \cdot \frac{\mathbb{E}[X]-a}{b-a}$$

Here, e^{sa} and e^{sb} are constants (since a and b are fixed), and we compute the expectations of the terms involving X.

Let's define:

$$p = \frac{\mathbb{E}[X]-a}{b-a}$$

Since $a \le \mathbb{E}[X] \le b$, we have $0 \le p \le 1$. Additionally:

$$1-p = 1 - \frac{\mathbb{E}[X]-a}{b-a} = \frac{b-a-(\mathbb{E}[X]-a)}{b-a} = \frac{b-\mathbb{E}[X]}{b-a}$$

These expressions match the coefficients in our inequality.

Substitute p and $1-p$ into the expression:

- $\frac{b-\mathbb{E}[X]}{b-a} = 1-p$
- $\frac{\mathbb{E}[X]-a}{b-a} = p$

Thus, the inequality simplifies to:

$$\mathbb{E}[e^{sX}] \le pe^{sb} + (1-p)e^{sa}$$

For the above inequality to be satisfied X has to be a two point distribution at a and b with probability p and 1-p respectively.To satisfy the condition E[X] = 0, we enforce $p\,b + (1-p)\,a = 0$ gives $p = \frac{-a}{b-a}$.

Define

$$M(s) \;=\; \mathbb{E}\big[e^{\,sX}\big] \;=\; p\,e^{sb} \;+\; (1-p)\,e^{sa}.$$

We want to prove $M(s) \le \exp\!\big(\frac{s^2\,(b-a)^2}{8}\big)$. Equivalently, let $F(s) = \ln\!\big(M(s)\big)$ and show

$$F(s) \;\le\; \frac{(b-a)^2}{8}\,s^2.$$

Observe

$$F(0) = \ln\!\big(p + (1-p)\big) = \ln(1) = 0, \quad F'(s) = \frac{M'(s)}{M(s)}.$$

Since $M'(s) = p\,b\,e^{sb} + (1-p)\,a\,e^{sa}$, we have $M'(0) = p\,b + (1-p)\,a = 0$ (by construction, $p\,b + (1-p)\,a = 0$). Hence $F'(0) = 0$. Next,

$$F''(s) \;=\; \frac{M''(s)\,M(s) - \big(M'(s)\big)^2}{\big(M(s)\big)^2}.$$

At $s = 0$, $M(0) = 1$ and $M'(0) = 0$, so $F''(0) = M''(0)$. But $M''(s) = p\,b^2\,e^{sb} + (1-p)\,a^2\,e^{sa}$, so $M''(0) = p\,b^2 + (1-p)\,a^2$. A direct calculation shows

$$p\,b^2 + (1-p)\,a^2 \;\le\; \frac{(b-a)^2}{4}.$$

Thus $F''(0) \le (b-a)^2/4$. A standard argument about functions with bounded second derivative near $s = 0$ then implies

$$F(s) \;\le\; \frac{(b-a)^2}{8}\, s^2 \quad \text{for all } s \ge 0.$$

Exponentiating,

$$M(s) \;=\; e^{F(s)} \;\le\; \exp\Bigl(\tfrac{s^2\,(b-a)^2}{8}\Bigr).$$

Hence $p\, e^{sb} + (1-p)\, e^{sa} \le \exp\bigl(\frac{s^2\,(b-a)^2}{8}\bigr)$.

Hnece any $X \in [a,b]$ with $\mathbb{E}[X] = 0$, we get

$$\mathbb{E}\bigl[\, e^{\,sX}\bigr] \;\le\; \exp\Bigl(\tfrac{s^2\,(b-a)^2}{8}\Bigr) \quad \text{for all } s \ge 0,$$

which completes the proof of the Lemma. $\square$

LEMMA 2.5

A *filter* on a probability space is a nested sequence of sub-σ-algebras

$$F_0 \;\subseteq\; F_1 \;\subseteq\; \cdots \;\subseteq\; F_n,$$

where each F_k represents the information available (or "revealed") up to stage k.
Let $(\Phi, \Omega) = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$ be a filter, let X be an integrable random variable, and let $X_0, X_1, \ldots, X_n$ be the martingale obtained by setting

$$X_k = \mathbb{E}[\, X \mid F_k],$$

for $k = 0, 1, \ldots, n$. Suppose that for each $k = 1, \ldots, n$ there is a constant c_k such that for any $s > 0$,

$$\mathbb{E}\bigl[\exp\{\, s(X_k - X_{k-1})\} \;\big|\; (F_{k-1})\bigr] \;\le\; \exp\Bigl\{\tfrac{c_k^2\, s^2}{8}\Bigr\}.$$

Then for any $s > 0$,

$$\Pr\bigl[\; \bigl|\mathbb{E}[X \mid F_n] - \mathbb{E}[X]\bigr| \;\ge s\bigr] \;\le\; 2\, \exp\Bigl\{-\, \tfrac{2\, s^2}{\sum_{k=1}^n c_k^2}\Bigr\}.$$

Proof: Let

$$M \;:=\; X_n \;-\; X_0 \;=\; \mathbb{E}[X \mid F_n] \;-\; \mathbb{E}[X].$$

We want to bound $\Pr(|M| \ge s)$ for $s > 0$. We use *Chernoff's method*, which applies Markov's inequality to $\exp(\lambda M)$ for $\lambda > 0$.

First note that

$$\Pr(M \ge s) \;=\; \Pr\bigl(e^{\lambda M} \;\ge\; e^{\lambda s}\bigr) \;\le\; \frac{\mathbb{E}[\, e^{\lambda M}\,]}{e^{\lambda s}} \;=\; \exp(-\,\lambda s)\, \mathbb{E}\bigl[e^{\lambda M}\bigr].$$

for any choice of $\lambda > 0$. This inequality might seem strange at first, but it is simply an application of Markov's inequality after exponentiating the event $\{\, M \ge s\}$.

Hence, if we can show $\mathbb{E}\bigl[\, e^{\lambda M}\bigr]$ is itself bounded by an exponential in λ^2, we can plug that in and then optimize over λ.

Observe that

$$M \;=\; X_n - X_0 \;=\; \sum_{k=1}^{n} \bigl(X_k - X_{k-1}\bigr).$$

Denote $\Delta_k := X_k - X_{k-1}$. Then

$$M \;=\; \Delta_1 + \Delta_2 + \cdots + \Delta_n, \quad \exp(\lambda M) \;=\; \exp\Bigl(\lambda \sum_{k=1}^{n} \Delta_k\Bigr) \;=\; \prod_{k=1}^{n} \exp\{\lambda\, \Delta_k\}.$$

Intuitively, by tuning λ we will penalize the large deviations of M in the exponential and obtain a sharp bound. The next step is to control $\mathbb{E}[e^{\lambda M}]$ using the assumption about the increments $X_k - X_{k-1}$. Recall that $M = X_n - X_0 = \sum_{k=1}^n (X_k - X_{k-1})$.

We will use the tower property of conditional expectation to handle this sum one step at a time. For clarity, let us write $\Delta_k := X_k - X_{k-1}$ for each increment. Then $M = \Delta_1 + \Delta_2 + \cdots + \Delta_n$. Consider the exponential $\exp(\lambda M) = \exp(\lambda(\Delta_1 + \cdots + \Delta_n)) = \prod_{k=1}^{n} \exp(\lambda \Delta_k)$. Although the increments Δ_k need not be independent, they are adapted to the filtration in a way that we can successively condition on F_{k-1} to bound this product's expectation.

Recall the *law of total expectation* (sometimes called the "tower property"), which states that for any integrable random variable Y,

$$\mathbb{E}[\,Y\,] \;=\; \mathbb{E}\Big[\,\mathbb{E}[Y \mid F_{n-1}]\Big].$$

By applying this identity to the random variable $Y = \exp\{\lambda(\Delta_1 + \cdots + \Delta_n)\}$, we obtain

$$\mathbb{E}\big[\exp\{\lambda(\Delta_1 + \cdots + \Delta_n)\}\big] \;=\; \mathbb{E}\Big[\mathbb{E}\big(e^{\lambda(\Delta_1+\cdots+\Delta_n)} \;\big|\; F_{n-1}\big)\Big].$$

In this step, $\Delta_1, \ldots, \Delta_{n-1}$ are considered F_{n-1}-measurable and hence can be viewed as constants inside the inner conditional expectation, while Δ_n remains random even after conditioning on F_{n-1}. Formally:

$$\mathbb{E}\big(e^{\lambda(\Delta_1+\cdots+\Delta_n)} \;\big|\; F_{n-1}\big) \;=\; e^{\lambda(\Delta_1+\cdots+\Delta_{n-1})} \;\times\; \mathbb{E}\big[e^{\lambda\,\Delta_n} \;\big|\; F_{n-1}\big].$$

Thus, conditioning on F_{n-1} isolates the final increment Δ_n in a single expectation factor, while everything up to Δ_{n-1} appears as a prefactor $e^{\lambda(\Delta_1+\cdots+\Delta_{n-1})}$.

Inside the inner conditional expectation, because $\Delta_1, \ldots, \Delta_{n-1}$ are F_{n-1}-measurable (they are determined by stages 1 through $n-1$ of the filtration) while Δ_n is F_n-measurable but independent of F_{n-1}, we can factor the product. In particular, on the event F_{n-1} we treat $\Delta_1, \ldots, \Delta_{n-1}$ as fixed, so

$$\mathbb{E}\Big(e^{\lambda(\Delta_1+\cdots+\Delta_n)} \;\Big|\; F_{n-1}\Big) \;=\; e^{\lambda(\Delta_1+\cdots+\Delta_{n-1})}\,\mathbb{E}\Big[e^{\lambda\Delta_n} \;\Big|\; F_{n-1}\Big].$$

This step uses the fact that if A is F_{n-1}-measurable, then $\mathbb{E}[A \cdot Z \mid F_{n-1}] = A\,\mathbb{E}[Z \mid F_{n-1}]$. We have pulled $\exp\{\lambda(\Delta_1 + \cdots + \Delta_{n-1})\}$ outside the conditional expectation because it is F_{n-1}-measurable. Now we apply the exponential moment assumption for the increment Δ_n: by hypothesis,

$$\mathbb{E}\big[e^{\lambda\Delta_n} \;\big|\; F_{n-1}\big] \;\le\; \exp\Big\{\tfrac{c_n^2\,\lambda^2}{8}\Big\}.$$

This is exactly where we use the given condition in the lemma: conditioned on F_{n-1}, the moment-generating function (mgf) of the increment Δ_n is bounded by an exponential with variance term $c_n^2/4$ in the exponent. Substituting this bound into the previous expression gives

$$\mathbb{E}\Big(e^{\lambda(\Delta_1+\cdots+\Delta_n)} \;\Big|\; F_{n-1}\Big) \;\le\; e^{\lambda(\Delta_1+\cdots+\Delta_{n-1})}\;\exp\Big\{\tfrac{c_n^2\,\lambda^2}{8}\Big\}.$$

Now take the expectation of both sides (i.e. unconditioned expectation). The right-hand side factors into a constant times $\mathbb{E}[e^{\lambda(\Delta_1+\cdots+\Delta_{n-1})}]$ because $\exp\{\frac{c_n^2\lambda^2}{8}\}$ does not depend on any randomness. We obtain:

$$\mathbb{E}\Big[e^{\lambda(\Delta_1+\cdots+\Delta_n)}\Big] \;\le\; \exp\Big\{\tfrac{c_n^2\,\lambda^2}{8}\Big\}\;\mathbb{E}\Big[e^{\lambda(\Delta_1+\cdots+\Delta_{n-1})}\Big].$$

Simplifying the product of exponentials proves the formula for k. By recursion, the formula holds for $k = n$. In particular, we have established

$$\mathbb{E}[\,e^{\lambda M}\,] \;=\; \mathbb{E}\Big[e^{\lambda(\Delta_1+\cdots+\Delta_n)}\Big] \;\le\; \exp\Big\{\tfrac{\lambda^2}{8}\sum_{k=1}^{n} c_k^2\Big\}.$$

This is a uniform bound on the mgf of the martingale difference $M = X_n - X_0$. Plugging this bound back into our earlier inequality from Markov's inequality, we get:

$$\Pr(M \ge s) \;\le\; \exp(-\lambda\,s)\;\exp\Big(\tfrac{\lambda^2}{8}\sum_{k=1}^{n} c_k^2\Big).$$

This holds for every $\lambda > 0$. Now, to get the strongest (smallest) possible upper bound, we optimize the right-hand side as a function of λ. The exponent in that expression is a quadratic function of λ:

$$f(\lambda) \;:=\; -\lambda s \;+\; \frac{\lambda^2}{8}\sum_{k=1}^{n} c_k^2.$$

We can find its minimum by elementary calculus or by completing the square. Differentiating with respect to λ and setting to zero,

$$f'(\lambda) \;=\; -\,s \;+\; \frac{\lambda}{4}\sum_{k=1}^{n} c_k^2 \;=\; 0,$$

which gives the critical point $\lambda^* \;=\; \dfrac{4\,s}{\sum_{k=1}^{n} c_k^2}$. This critical point indeed minimizes $f(\lambda)$ (the coefficient of λ^2 is positive, so f is a convex parabola). Substituting λ^* back into $f(\lambda)$, we get

$$f(\lambda^*) = -\,\frac{4\,s}{\sum_{k=1}^{n} c_k^2}\,s \;+\; \frac{\left(4\,s/\sum c_k^2\right)^2}{8}\,\sum_{k=1}^{n} c_k^2 \;=\; -\,\frac{2\,s^2}{\sum_{k=1}^{n} c_k^2}.$$

Therefore, choosing $\lambda = \lambda^*$ in the inequality yields the smallest (tightest) bound:

$$\Pr(M \geq s) \;\leq\; \exp\Bigl\{-\,\frac{2\,s^2}{\sum_{k=1}^{n} c_k^2}\Bigr\}.$$

This takes care of the upper tail $\Pr(X_n - X_0 \geq s)$. To bound the lower tail $\Pr(X_n - X_0 \leq -\,s)$, we apply the same argument to the martingale $-\,M \;=\; X_0 - X_n$. Observe that $-\,M$ satisfies the same exponential moment condition with the same constants c_k, because

$$\mathbb{E}\bigl[\exp\{\,\lambda\,(X_{k-1} - X_k)\}\;\big|\;F_{k-1}\bigr] \;=\; \mathbb{E}\bigl[\exp\{-\,\lambda\,(X_k - X_{k-1})\}\;\big|\;F_{k-1}\bigr],$$

and by the given assumption this is $\leq \exp\{\frac{c_k^2\,\lambda^2}{8}\}$ for all $\lambda > 0$ (the inequality holds with λ replaced by $-\lambda$ as well, since it was stated for "any $s > 0$"). Thus by the identical derivation,

$$\Pr(-\,M \geq s) \;=\; \Pr(X_0 - X_n \geq s) \;=\; \Pr(X_n - X_0 \leq -\,s) \;\leq\; \exp\Bigl\{-\,\frac{2\,s^2}{\sum_{k=1}^{n} c_k^2}\Bigr\}.$$

Finally, we combine the two tail bounds. The event $\{|M| \geq s\}$ is the union of $\{M \geq s\}$ and $\{M \leq -s\}$, which are disjoint events. Hence by the union bound (or simply summing the two probabilities) we obtain for any $s > 0$:

$$\Pr\bigl(|M| \geq s\bigr) \;=\; \Pr(M \geq s) \;+\; \Pr(M \leq -\,s) \;\leq\; \exp\Bigl\{-\,\tfrac{2\,s^2}{\sum_{k=1}^{n} c_k^2}\Bigr\} \;+\; \exp\Bigl\{-\,\tfrac{2\,s^2}{\sum_{k=1}^{n} c_k^2}\Bigr\}.$$

Note: The bound on $Pr(M \leq -\,s)$ can be achieved by following the same process by defining $(-M)$ as your random variable. Thus, each $\Delta_k := X_{k-1} - X_k$ in this case, as $-M = X_0 - X_n$.

This simplifies to

$$\Pr\Bigl[\bigl|\mathbb{E}[X \mid F_n] - \mathbb{E}[X]\bigr| \;\geq\; s\Bigr] \;=\; \Pr(|M| \geq s) \;\leq\; 2\,\exp\Bigl\{-\,\tfrac{2\,s^2}{\sum_{k=1}^{n} c_k^2}\Bigr\}.$$

This is exactly the desired concentration inequality. Intuitively, the reason we get a Gaussian-style $\exp(-(\text{constant})\cdot s^2)$ bound is that the sub-exponential increment condition $\mathbb{E}[e^{\lambda(X_k - X_{k-1})} \mid F_{k-1}] \leq \exp\{\frac{c_k^2\,\lambda^2}{8}\}$ behaves like a variance bound $\mathrm{Var}(X_k - X_{k-1}) \leq \frac{c_k^2}{4}$ in the limit of small λ. Chernoff's method effectively leverages all those individual bounds at once to control the tail of the sum $M = \sum_{k=1}^{n}(X_k - X_{k-1})$. This completes the proof. $\square$

LEMMA 2.6

Let $0 < k_1 < k_2$ be integers, and let $h^{(k_1)}$ be a k_1-history such that $M_{k_1}^n \leq a\,n$. Define A to be the event that $\underline{H}^{(k_1)} = \underline{h}^{(k_1)}$. Then for any $s > 0$,

$$\Pr\Bigl(\,\bigl|E[Q_n \mid \underline{H}^{(k_2)}] - E[Q_n \mid A]\bigr| \;\geq\; s \;\Big|\; A\Bigr) \;\leq\; 2\,\exp\Bigl(-\,\frac{s^2}{2\,(k_2 - k_1)\,a\,n^2}\Bigr).$$

Proof: Set up a Martingale

Define a filtration based on the history revealed from depth k_1 to k_2: For

$$k = k_1, k_1 + 1, \ldots, k_2,$$

let

$$\mathcal{F}_k = \sigma\big(\underline{H}(k)\big)$$

be the sigma-algebra generated by the history $H(k)$. Given an event A, $\underline{H}(k_1) = \underline{h}(k_1)$ is fixed.

Now, define the martingale sequence under the conditional probability $\Pr[\cdot \mid A]$:

$$X_k = \mathbb{E}\big[Q_n \mid \underline{H}(k)\big] \quad \text{for } k = k_1, \ldots, k_2.$$

Check the boundary conditions: At $k = k_1$,

$$X_{k_1} = \mathbb{E}\big[Q_n \mid \underline{H}(k_1)\big] = \mathbb{E}\big[Q_n \mid A\big],$$

which is a constant since A fixes $\underline{H}(k_1)$.

At $k = k_2$,

$$X_{k_2} = \mathbb{E}\big[Q_n \mid \underline{H}(k_2)\big],$$

which varies based on the pivot choices between depths k_1 and k_2.

Since

$$\mathbb{E}\big[X_{k+1} \mid \mathcal{F}_k, A\big] = X_k,$$

the sequence $\{X_k\}$ is a martingale with respect to $\{\mathcal{F}_k\}$ under $\Pr[\cdot \mid A]$.

Our goal is to bound:

$$\Pr\Big[\big|X_{k_2} - X_{k_1}\big| \geq s \mid A\Big] = \Pr\Big[\Big|\mathbb{E}\big[Q_n \mid \underline{H}(k_2)\big] - \mathbb{E}\big[Q_n \mid A\big]\Big| \geq s \mid A\Big].$$

<u>Why a Martingale?</u>

Check if $E[X_{k+1} \mid \underline{H}^{(k)}, A] = X_k$:

$$E[X_{k+1} \mid \underline{H}^{(k)}, A] = E\left[E[Q_n \mid \underline{H}^{(k+1)}] \mid \underline{H}^{(k)}, A\right].$$

- $\underline{H}^{(k+1)}$ is $\underline{H}^{(k)}$ plus the pivot choices at depth k.
- By the tower property (a rule in probability): if you know $\underline{H}^{(k)}$, averaging over all possible $\underline{H}^{(k+1)}$ gives you the expectation with just $\underline{H}^{(k)}$:

$$E[Q_n \mid \underline{H}^{(k)}, A] = X_k.$$

- Since A is always true here, this holds. So, X_k is a martingale.

<u>Understand Q_n and Differences</u>

Let us see how Q_n works in the tree:

- At each node with m elements, quicksort makes $m-1$ comparisons to split into left ($L_{m,l}$) and right ($R_{m,r}$) sublists.
- $Q_m = \sum$ of $m-1$ over all nodes where splitting happens.
- $E[Q_n \mid H^{(k)}]$:
  - Up to depth $k-1$, all comparisons are fixed by $H^{(k)}$.
  - At depth k, we have sublists $L_1^{(k)}, \ldots, L_s^{(k)}$.
  - For each sublist, we expect $E[Q_{|L_i^{(k)}|}]$ comparisons from there down.
- So:

$$E[Q_n \mid H^{(k)}] = (\text{comparisons up to depth } k-1) + \sum_{i=1}^{s} E[Q_{|L_i^{(k)}|}].$$

$$X_{k+1} - X_k = E[Q_n \mid \underline{H}^{(k+1)}] - E[Q_n \mid \underline{H}^{(k)}].$$

- Up to depth $k-1$, the comparisons are the same.

- At depth k, $\underline{H}^{(k)}$ gives sublists $L_1^{(k)}, \dots, L_s^{(k)}$.
- $H^{(k+1)}$ tells us the pivots picked at depth k:
  - For $L_i^{(k)} = m > 1$, pick pivot rank J_i (from 1 to m), left size $J_i - 1$, right size $m - J_i$.
  - $J_i \sim U(1, m)$, comparisons at that node = $m - 1$.
- Comparisons at node = $m - 1$.

So:

$$E[Q_n \mid \underline{H}^{(k+1)}] = (\text{comparisons up to depth } k) + \sum_{i:|L_i^{(k)}|>1} \Big[(|L_i^{(k)}| - 1) + E[Q_{J_i-1}] + E[Q_{|L_i^{(k)}|-J_i}]\Big].$$

- Difference:

$$X_{k+1} - X_k = \sum_{i:|L_i^{(k)}|>1} \Big[(|L_i^{(k)}| - 1) + E[Q_{J_i-1}] + E[Q_{|L_i^{(k)}|-J_i}] - E[Q_{|L_i^{(k)}|}]\Big].$$

Bound the Differences

To use a concentration inequality, we need to bound

$$X_{k+1} - X_k.$$

Refer to Lemma 2.2 from the paper, which states that for any sublist size m and pivot rank $j = 1, \dots, m$,

$$\Big|(m-1) + q_{j-1} + q_{m-j} - q_m\Big| \le m,$$

where

$$q_m = \mathbb{E}[Q_m].$$

Since $T_i(k)$ takes the form

$$T_i(k) = \big(L_i(k) - 1\big) + q_{J_i-1} + q_{L_i(k)-J_i} - q_{L_i(k)},$$

it follows that

$$|T_i(k)| \le L_i(k).$$

Thus,

$$|X_{k+1} - X_k| = \Big|\sum_{i:L_i(k)>1} T_i(k)\Big| \le \sum_{i:L_i(k)>1} |T_i(k)| \le \sum_{i:L_i(k)>1} L_i(k) \le \sum_i L_i(k),$$

where the last sum is over all sublists at depth k.

Since the sublists at depth k partition the original list of n elements, we have

$$\sum_i L_i(k) \le n.$$

Additionally, the condition

$$M_k^1 \le \alpha n$$

implies that at depth k_1, the largest sublist size is at most αn. As $k \ge k_1$, sublist sizes decrease (or remain the same) with depth due to partitioning, so

$$L_i(k) \le \alpha n \quad \text{for all } i \text{ and } k \ge k_1.$$

To bound the variance-like term, consider:

$$\sum_i \big(L_i(k)\big)^2 \le \max_i L_i(k) \cdot \sum_i L_i(k) \le \alpha n \cdot n = \alpha n^2.$$

This bound holds for each k from k_1 to $k_2 - 1$.

Applying Azuma's Inequality

We use Lemma 2.5 from the paper, a variant of Azuma's inequality for martingales. Suppose a martingale $\{X_k\}$ from $k = k_1$ to k_2 satisfies, for each k,

$$\mathbb{E}\Big[\exp\{s(X_{k+1} - X_k)\} \mid \mathcal{F}_k\Big] \le \exp\Big\{\frac{c_k^2 s^2}{8}\Big\},$$

for some constants c_k. Then:

$$\Pr\Big[\big|X_{k_2} - X_{k_1}\big| \geq s \mid A\Big] \leq 2\exp\Big\{-\frac{2s^2}{\sum_{k=k_1}^{k_2-1} c_k^2}\Big\}.$$

To apply this, compute the moment generating function of $X_{k+1} - X_k$. Since the $T_i(k)$ are conditionally independent given $\mathcal{F}_k$ (pivot choices in different sublists are independent), and

$$\mathbb{E}[T_i(k) \mid \mathcal{F}_k] = 0,$$

we use Lemma 2.4:

For a random variable X with $\mathbb{E}[X] = 0$ and $a \leq X \leq b$,

$$\mathbb{E}\big[\exp\{sX\}\big] \leq \exp\Big\{\frac{s^2(b-a)^2}{8}\Big\}.$$

For each $T_i(k)$, since

$$E[T_i(k)] = \frac{1}{L_i(k)} \sum (L_i - 1 + q_{i-j} + q_{j-1} - q_i) = 0$$

where q_i is the expected no of comparisons for each subarray above. And by lemma 2.2,

$$-L_i(k) \leq T_i(k) \leq L_i(k),$$

we have $b - a = 2L_i(k)$, and thus:

$$\mathbb{E}\Big[\exp\{sT_i(k)\} \mid \mathcal{F}_k\Big] \leq \exp\Big\{\frac{s^2(2L_i(k))^2}{8}\Big\} = \exp\Big\{\frac{s^2\big(L_i(k)\big)^2}{2}\Big\}.$$

Hence,

$$\begin{aligned}\mathbb{E}\Big[\exp\{s(X_{k+1} - X_k)\} \mid \mathcal{F}_k\Big] &= \prod_i \mathbb{E}\Big[\exp\{sT_i(k)\} \mid \mathcal{F}_k\Big] \\ &\leq \prod_i \exp\Big\{\frac{s^2\big(L_i(k)\big)^2}{2}\Big\} \\ &= \exp\Big\{\frac{s^2}{2}\sum_i \big(L_i(k)\big)^2\Big\}.\end{aligned}$$

Using the bound

$$\sum_i \big(L_i(k)\big)^2 \leq \alpha n^2,$$

we obtain:

$$\mathbb{E}\Big[\exp\{s(X_{k+1} - X_k)\} \mid \mathcal{F}_k\Big] \leq \exp\Big\{\frac{s^2}{2}\alpha n^2\Big\}.$$

This expression matches the form required by Lemma 2.5 with

$$c_k^2 = 4\alpha n^2,$$

since

$$\frac{c_k^2 s^2}{8} = \frac{4\alpha n^2 s^2}{8} = \frac{s^2\alpha n^2}{2}.$$

<u>Computing the Final Bound</u>

Since,

$$(X_{k2} - X_{k1}) = \sum_{k=k_1}^{k_2-1} X_{k+1} - X_k$$

The number of steps is $k_2 - k_1$, so:

$$\sum_{k=k_1}^{k_2-1} c_k^2 = (k_2 - k_1) \cdot 4\alpha n^2.$$

By using Lemma 2.5, which provides a concentration bound for martingales, we apply it to the sequence $\{X_k\}$ defined under $\Pr[\cdot|A]$, where the bound $\Pr[X_{k_2} - X_{k_1} \geq s|A]$ holds within this conditional measure. Since A fixes $\underline{H}_{k_1}$, and the differences satisfy the moment condition $E[\exp\{s(X_{k+1} - X_k)\}|F_k] \leq \exp\{c_k^2 s^2/8\}$, Lemma 2.5's unconditional form extends to the conditional setting by applying the inequality to the restricted probability space,therefore we have the following bound:

$$\Pr\Big[\big|X_{k_2} - X_{k_1}\big| \geq s \mid A\Big] \leq 2\exp\Big\{-\frac{2s^2}{(k_2-k_1)\cdot 4\alpha n^2}\Big\} = 2\exp\Big\{-\frac{s^2}{2(k_2-k_1)\alpha n^2}\Big\}.$$

This is exactly the bound stated in Lemma 2.6, completing the proof. □

Lemma 2.7

Let n, k_1, and s be positive integers. Then, for any real α with $0 < \alpha \leq 1$ and any positive integer k_2 satisfying

$$\ln\Big(\frac{1}{\alpha}\Big) \leq k_1, \quad k_2 > k_1, \quad \text{and} \quad k_2 \geq \ln\Big(\frac{n}{2}\Big),$$

we have

$$\Pr\Big(|Q_n - q_n| \geq k_1\, n + s\Big) \;\leq\; \frac{2}{n}\left(\frac{2e\,\ln(n/2)}{k_2}\right)^{k_2} \;+\; \alpha\left(\frac{2e\,\ln(1/\alpha)}{k_1}\right)^{k_1} \;+\; 2\exp\left(-\frac{s^2}{2(k_2-k_1)\,\alpha\, n^2}\right).$$

Proof: Let

$$R_n = \mathbb{E}\big[Q_n \mid \underline{H}^{(k_2)}\big]$$

be the conditional expectation of Q_n given the (k_2)-history $\underline{H}^{(k_2)}$. Note that R_n is a random variable measurable with respect to $\underline{H}^{(k_2)}$. Also from earlier definition we know that

$$M_{k_1}^n = \max\{L_j : \text{node } j \text{ appears at depth } k_1\}$$

which is determined by the (k_1)-history $\underline{h}^{(k_1)}$. Define

$$H = \Big\{\, \underline{h}^{(k_1)} : M_{k_1}^n \leq \alpha\, n \Big\}.$$

Then, by partition theorem, we can write

$$\begin{aligned}\Pr\Big(|Q_n - q_n| \geq k_1 n + s\Big) &= \Pr\Big(|Q_n - q_n| \geq k_1 n + s \;\Big|\; Q_n \neq R_n\Big)\Pr(Q_n \neq R_n) \\ &\quad + \Pr\Big(|Q_n - q_n| \geq k_1 n + s \;\Big|\; Q_n = R_n\Big)\Pr(Q_n = R_n). \\ &\leq \Pr\{Q_n \neq R_n\} + \Pr\Big(|Q_n - q_n| \geq k_1 n + s \mid Q_n = R_n\Big),\end{aligned}$$

since probabilities are at most 1

Next, we condition further on the event that the (k_1)-history belongs or does not belong to H. In particular,

$$\begin{aligned}\Pr\Big(|Q_n - q_n| \geq k_1 n + s \mid Q_n = R_n\Big) &= \Pr\Big(|Q_n - q_n| \geq k_1 n + s \mid Q_n = R_n,\ \underline{H}^{(k_1)} \notin H\Big)\ \Pr\{\underline{H}^{(k_1)} \notin H\} \\ &\quad + \Pr\Big(|Q_n - q_n| \geq k_1 n + s,\ \underline{H}^{(k_1)} \in H \mid Q_n = R_n\Big) \\ &\leq \Pr\{\underline{H}^{(k_1)} \notin H\} + \Pr\Big(|R_n - q_n| \geq k_1 n + s,\ \underline{H}^{(k_1)} \in H\Big),\end{aligned}$$

since each probability is at most 1. Thus we we have,

$$\begin{aligned}\Pr\Big(|Q_n - q_n| \geq k_1 n + s\Big) &\leq \underbrace{\Pr\{Q_n \neq R_n\}}_{\text{(i)}} + \underbrace{\Pr\{\underline{H}^{(k_1)} \notin H\}}_{\text{(ii)}} \\ &\quad + \underbrace{\sum_{h \in H} \Pr\Big(|R_n - q_n| \geq k_1 n + s \mid \underline{H}^{(k_1)} = h\Big)\Pr\{\underline{H}^{(k_1)} = h\}}_{\text{(iii)}}.\end{aligned}$$

Next we bound each term in the above equation,

Term (i)

Define the event

$$E = \{M_{k_2}^n < 2\},$$

which implies that no comparisons are performed beyond depth k_2. This is because the maximum subarray size at depth k_2 is less than 2, so all subarrays are of size 1. Thus, if E occurs then Q_n is completely determined by the history $\underline{H}^{(k_2)}$. If this condition was not true, then there might be some randomness left beyond depth k_2, leading to different no of comparisons.Thus when this occurs,

$$Q_n = R_n.$$

Since both are determined by the $\underline{H}^{(k_2)}$ only, they are the same random variable. It follows that

$$\Pr\{E\} \le \Pr\{Q_n = R_n\}$$
$$\Pr\{Q_n \ne R_n\} \le \Pr\{E^c\} = \Pr\{M_{k_2}^n \ge 2\}.$$

By application of Lemma 2.1 , we have

$$\Pr\{M_{k_2}^n \ge 2\} = \Pr\{M_{k_2}^n \ge \frac{2}{n} n\} \le \frac{2}{n}\left(\frac{2e\ \ln(n/2)}{k_2}\right)^{k_2}.$$

Term (ii)

Note that by definition the event

$$\{\underline{H}^{(k_1)} \notin H\}$$

is equivalent to

$$\{M_{k_1}^n > \alpha\, n\}.$$

Thus, by Lemma 2.1,

$$\Pr\{M_{k_1}^n > \alpha\, n\} \le \alpha\left(\frac{2e\ \ln(1/\alpha)}{k_1}\right)^{k_1}.$$

Term (iii)

In this case we have $Q_n = R_n$ .By Lemma 2.3 we have

$$\left|\mathbb{E}\left[Q_n \mid \underline{H}^{(k_1)} = h\right] - q_n\right| \le k_1 n \quad \text{for every } h \in \underline{H}^{(k_1)}.$$

We define the event

$$A = \Big\{\, |R_n - q_n| \ge k_1 n + s \Big\},$$
$$C = \Big\{\, |R_n - E[Q_n \mid \underline{H}^{k_1} = h]| \ge s \Big\},$$

We now show that whenever A is true, C is also true. Let A be true, It implies the following is true,

$$|R_n - q_n + E[Q_n \mid \underline{H}^{k_1} = h] - E[Q_n \mid \underline{H}^{k_1} = h]| \ge k_1 n + s,$$

By using triangle inequality,

$$|R_n - q_n + E[Q_n \mid \underline{H}^{k_1} = h] - E[Q_n \mid \underline{H}^{k_1} = h]| \le |R_n - \mathbb{E}[Q_n \mid \underline{H}_{k_1} = h]| + |\mathbb{E}[Q_n \mid \underline{H}_{k_1} = h] - q_n|$$
$$\le |R_n - \mathbb{E}[Q_n \mid \underline{H}_{k_1} = h]| + k_1 n$$

Thus,

$$k_1 n + s \le |R_n - \mathbb{E}[Q_n \mid \underline{H}^{k_1} = h]| + k_1 n$$

Thus event C is true from above. Hence whenever event A occurs, event C also occurs.

$$Pr[A] \leq Pr[C]$$

$$\Pr\left(|R_n - q_n| \geq k_1 n + s \mid \underline{H}^{k_1} = h\right) \leq \Pr\left(\left|R_n - \mathbb{E}[Q_n \mid \underline{H}^{k_1} = h]\right| \geq s \mid \underline{H}^{k_1} = h\right)$$

By applying Lemma 2.6, we have for all $h \in H^{k_1}$

$$\Pr\Big(\left|Q_n - \mathbb{E}[Q_n \mid \underline{H}^{(k_1)} = h]\right| \geq s \mid \underline{H}^{(k_1)} = h\Big) \leq 2\exp\left(-\frac{s^2}{2\,(k_2 - k_1)\,\alpha\, n^2}\right).$$

Thus,

$$\Pr\Big(|R_n - q_n| \geq k_1 n + s \mid \ \underline{H}^{(k_1)} \in H\Big) \leq 2\exp\left(-\frac{s^2}{2\,(k_2 - k_1)\,\alpha\, n^2}\right).$$

Substituting each of these upper bounds we have the following result.

Now, combining all these bounds, by the partition (or total probability) theorem we obtain

$$\begin{aligned}
\Pr\Big(|Q_n - q_n| \geq k_1 n + s\Big) &\leq \Pr\{Q_n \neq R_n\} \ + \ \Pr\{\underline{H}^{(k_1)} \notin H\} \\
&\quad + \sum_{h \in H} \Pr\Big(|Q_n - q_n| \geq k_1 n + s \mid \underline{H}^{(k_1)} = h\Big) \Pr\{\underline{H}^{(k_1)} = h\} \\
&\leq \frac{2}{n}\left(\frac{2e\ \ln(n/2)}{k_2}\right)^{k_2} \ + \ \alpha\left(\frac{2e\ \ln(1/\alpha)}{k_1}\right)^{k_1} \ + \ 2\exp\left(-\frac{s^2}{2\,(k_2 - k_1)\,\alpha\, n^2}\right)
\end{aligned}$$

which is the desired result. □

LEMMA 2.8 - UPPER BOUND

For any $\varepsilon > 0$,

$$\Pr\left[\left|\frac{Q_n}{q_n} - 1\right| > \varepsilon\right] \leq n^{-2\varepsilon((\ln^{(2)} n) + O(\ln^{(4)} n))}$$

where:

- Q_n is the number of comparisons in QuickSort on an array of size n
- $q_n = 2(n+1)H_n - 4n$ is its expected value with $H_n = \sum_{k=1}^{n} \frac{1}{k}$
- $\ln^{(k)} n$ denotes the k-th iterated natural logarithm of n.

Proof: To prove this lemma, we aim to bound the probability $\Pr\left[\left|\frac{Q_n}{q_n} - 1\right| > \varepsilon\right] = \Pr\left[|Q_n - q_n| > \varepsilon q_n\right]$. We will use Lemma 2.7, which states that for positive integers n, k_1, s, real α with $0 < \alpha \leq 1$, and integer k_2 satisfying $\ln(1/\alpha) \leq k_1$, $k_2 > k_1$, and $k_2 \geq \ln(n/2)$,

$$\Pr\left[|Q_n - q_n| \geq k_1 n + s\right] \leq \frac{2}{n}\left(\frac{2e\ln(n/2)}{k_2}\right)^{k_2} + \alpha\left(\frac{2e\ln(1/\alpha)}{k_1}\right)^{k_1} + 2\exp\left\{-\frac{s^2}{2(k_2 - k_1)\alpha n^2}\right\}.$$

Our strategy is to choose parameters $s = s(n)$, $k_1 = k_1(n)$, $\alpha = \alpha(n)$, and $k_2 = k_2(n)$ such that $k_1 n + s$ approximates εq_n, and the right-hand side of the inequality yields the desired bound.

First, let's express q_n. Recall $q_n = 2(n+1)H_n - 4n$, where $H_n = \ln n + \gamma + O(1/n)$ and $\gamma \approx 0.57721$ is the Euler-Mascheroni constant. Expanding:

$$q_n = 2(n+1)(\ln n + \gamma + O(1/n)) - 4n = 2n\ln n + 2\ln n + 2\gamma n + 2\gamma + O(1) - 4n = 2n\ln n + (2\gamma - 4)n + 2\ln n + O(1).$$

For large n, the dominant term is $2n\ln n$, so $\varepsilon q_n \approx 2\varepsilon n \ln n$, with lower-order terms $O(\varepsilon n)$.

Following the paper's sketch, define:

- $s = s(n) = \frac{n \ln n}{\ln^{(2)} n}$, where $\ln^{(2)} n = \ln\ln n$. (We assume s is adjusted to an integer, negligible for large n.)

- $k_1 = k_1(n) = 2\varepsilon \ln n - 2\frac{s}{n}$, similarly an integer with $O(1)$ adjustment.

- $\alpha = \alpha(n) = (\ln^{(2)} n)^{-5}$.

- $k_2 = k_2(n) = (\ln n)(\ln^{(2)} n)$, also an integer.

Compute $k_1 n + s$:

$$\frac{s}{n} = \frac{\frac{n \ln n}{\ln^{(2)} n}}{n} = \frac{\ln n}{\ln^{(2)} n},$$
$$k_1 = 2\varepsilon \ln n - 2\frac{\ln n}{\ln^{(2)} n},$$
$$k_1 n = n\left(2\varepsilon \ln n - 2\frac{\ln n}{\ln^{(2)} n}\right) = 2\varepsilon n \ln n - 2\frac{n \ln n}{\ln^{(2)} n} = 2\varepsilon n \ln n - 2s,$$
$$k_1 n + s = 2\varepsilon n \ln n - 2s + s = 2\varepsilon n \ln n - s = 2\varepsilon n \ln n - \frac{n \ln n}{\ln^{(2)} n} = n \ln n \left(2\varepsilon - \frac{1}{\ln^{(2)} n}\right).$$

Since $\varepsilon q_n \approx 2\varepsilon n \ln n$, compare:

$$\frac{k_1 n + s}{\varepsilon q_n} \approx \frac{2\varepsilon n \ln n - \frac{n \ln n}{\ln^{(2)} n}}{2\varepsilon n \ln n} = 1 - \frac{1}{2\varepsilon \ln^{(2)} n}.$$

For large n, $\ln^{(2)} n = \ln \ln n \to \infty$, so $\frac{1}{\ln^{(2)} n} \to 0$, and $k_1 n + s < \varepsilon q_n$. Thus, $\Pr\left[|Q_n - q_n| > \varepsilon q_n\right] \leq \Pr\left[|Q_n - q_n| > k_1 n + s\right]$ for sufficiently large n.

Verify Lemma 2.7 conditions:

- $\ln(1/\alpha) = \ln\left((\ln^{(2)} n)^5\right) = 5 \ln \ln^{(2)} n = 5 \ln \ln \ln n = 5 \ln^{(3)} n$

- $k_1 \approx 2\varepsilon \ln n - 2\frac{\ln n}{\ln^{(2)} n} \approx 2\varepsilon \ln n$, and $2\varepsilon \ln n \gg 5 \ln^{(3)} n$ since $\ln n / \ln^{(3)} n \to \infty$

- $k_2 = (\ln n)(\ln^{(2)} n) > 2\varepsilon \ln n \approx k_1$ for large n, as $\ln^{(2)} n \to \infty$

- $\ln(n/2) = \ln n - \ln 2 < (\ln n)(\ln^{(2)} n) = k_2$.

All conditions hold. Apply Lemma 2.7:

$$\Pr\left[|Q_n - q_n| > k_1 n + s\right] \leq T_1 + T_2 + T_3,$$

where:

$T_1 = \frac{2}{n}\left(\frac{2e \ln(n/2)}{k_2}\right)^{k_2}$ $\qquad$ $T_2 = \alpha\left(\frac{2e \ln(1/\alpha)}{k_1}\right)^{k_1}$ $\qquad$ $T_3 = 2\exp\left\{-\frac{s^2}{2(k_2 - k_1)\alpha n^2}\right\}$.

Calculate each term:

For T_1:

$$\frac{\ln(n/2)}{k_2} = \frac{\ln n - \ln 2}{(\ln n)(\ln^{(2)} n)} \approx \frac{\ln n}{(\ln n)(\ln^{(2)} n)} = \frac{1}{\ln^{(2)} n},$$
$$\frac{2e \ln(n/2)}{k_2} \approx \frac{2e}{\ln^{(2)} n},$$
$$T_1 = \frac{2}{n}\left(\frac{2e}{\ln^{(2)} n}\right)^{(\ln n)(\ln^{(2)} n)} = \frac{2}{n}\exp\left((\ln n)(\ln^{(2)} n)\ln\left(\frac{2e}{\ln^{(2)} n}\right)\right) = \frac{2}{n}\exp\left((\ln n)(\ln^{(2)} n)(\ln 2e - \ln \ln^{(2)} n)\right).$$

Since $\ln 2e \approx 1.945 < \ln \ln^{(2)} n$ for large n (e.g., $\ln \ln n > 2$), the exponent is negative and large, so $T_1 = o(n^{-k})$ for any fixed $k > 0$. We write T_1 in this way, to show that it is dominated by T_2 below.

For T_2:

$$\alpha = (\ln^{(2)} n)^{-5} = (\ln \ln n)^{-5},$$

$$
\begin{aligned}
\ln(1/\alpha) &= 5\ln^{(3)} n,\\
k_1 &\approx 2\varepsilon \ln n,\\
\frac{2e\ln(1/\alpha)}{k_1} &= \frac{2e\cdot 5\ln^{(3)} n}{2\varepsilon\ln n} = \frac{5e\ln^{(3)} n}{\varepsilon \ln n},\\
T_2 = (\ln^{(2)} n)^{-5}\left(\frac{5e\ln^{(3)} n}{\varepsilon\ln n}\right)^{2\varepsilon\ln n} &= (\ln^{(2)} n)^{-5}\exp\left(2\varepsilon\ln n\cdot\ln\left(\frac{5e\ln^{(3)} n}{\varepsilon\ln n}\right)\right),\\
\ln\left(\frac{5e\ln^{(3)} n}{\varepsilon\ln n}\right) &= \ln 5e + \ln\ln^{(3)} n - \ln\varepsilon - \ln\ln n \approx -\ln\ln n,\\
T_2 &\approx (\ln^{(2)} n)^{-5}\exp\left(-2\varepsilon(\ln n)(\ln^{(2)} n)\right) = (\ln^{(2)} n)^{-5} n^{-2\varepsilon\ln^{(2)} n}.
\end{aligned}
$$

Now, we shall upper bound this term by our desired expression:

$$
\begin{aligned}
&(\ln\ln n)^{-5}\, n^{-2\varepsilon\ln(\ln n)} \le n^{-2\varepsilon(\ln\ln n + C\ln\ln\ln\ln n)}\\
\text{if and only if}\quad &(\ln\ln n)^{-5}\, n^{-2\varepsilon\ln(\ln n)} \le n^{-2\varepsilon\ln\ln n}\, n^{-2\varepsilon C\ln\ln\ln\ln n}\\
\iff\quad &(\ln\ln n)^{-5} \le n^{-2\varepsilon C\ln\ln\ln\ln n}.
\end{aligned}
$$

Taking natural logarithms of both sides yields

$$
\begin{aligned}
-5\ \ln\big(\ln\ln n\big) &\le -2\varepsilon C\ \ln\ln\ln\ln n\cdot\ln n\\
2\varepsilon C\ \ln\ln\ln\ln n\cdot\ln n\ &\le 5\ \ln\big(\ln\ln n\big)\\
C &\le (5\ \ln\big(\ln\ln n\big))/(2\varepsilon\ln\ln\ln\ln(n).\ln(n))
\end{aligned}
$$

As $n\to\infty$, the right hand side of this expression approaches zero. Therefore, for any fixed $\varepsilon > 0$ one may choose a constant $C > 0$ (in fact, any positive C suffices for all sufficiently large n) so that

$$
2\varepsilon C\ \ln\ln\ln\ln n\cdot\ln n \le 5\ \ln(\ln\ln n)
$$

for all large n. This verifies that there exists a constant C such that

$$
(\ln\ln n)^{-5}\, n^{-2\varepsilon\ln(\ln n)}\ \le\ n^{-2\varepsilon(\ln\ln n + C\ln\ln\ln\ln n)}.
$$

Thus the claimed inequality holds, and we can now bound T_2 by $n^{-2\varepsilon(\ln^{(2)} n)+O((\ln^{(4)} n))}$ as required.

For T_3:

$$
\begin{aligned}
k_2 - k_1 \approx (\ln n)(\ln^{(2)} n) - 2\varepsilon\ln n &= (\ln n)(\ln^{(2)} n - 2\varepsilon) \approx (\ln n)(\ln^{(2)} n),\\
s^2 &= \left(\frac{n\ln n}{\ln^{(2)} n}\right)^2 = \frac{n^2(\ln n)^2}{(\ln^{(2)} n)^2},\\
2(k_2-k_1)\alpha n^2 &\approx 2(\ln n)(\ln^{(2)} n)(\ln^{(2)} n)^{-5} n^2 = 2n^2(\ln n)(\ln^{(2)} n)^{-4},\\
\frac{s^2}{2(k_2-k_1)\alpha n^2} = \frac{\frac{n^2(\ln n)^2}{(\ln^{(2)} n)^2}}{2n^2(\ln n)(\ln^{(2)} n)^{-4}} &= \frac{(\ln n)^2}{2(\ln^{(2)} n)^2}\cdot(\ln^{(2)} n)^4 = \frac{1}{2}(\ln n)(\ln^{(2)} n)^2,\\
T_3 &= 2\exp\left\{-\frac{1}{2}(\ln n)(\ln^{(2)} n)^2\right\},
\end{aligned}
$$

which is exponentially small.

Since T_1 and T_3 decay faster than any polynomial, and T_2 provides the leading term, we have:

$$
\Pr\left[|Q_n - q_n| > \varepsilon q_n\right] \le (\ln^{(2)} n)^{-5} n^{-2\varepsilon\ln^{(2)} n}
$$

which can also be written as:

$$
\Pr\left[|Q_n - q_n| > \varepsilon q_n\right] \le n^{-2\varepsilon(\ln^{(2)} n)+O((\ln^{(4)} n))}
$$

fitting the required bound, completing the proof. □

LEMMA 2.9 - LOWER BOUND

Let $\varepsilon > 0$. Then, as $n \to \infty$,

$$\Pr[Q_n > (1+\varepsilon)q_n] \geq \exp\{-2\varepsilon \ln n \ln^{(2)} n + O(\ln^{(3)} n)\},$$

where Q_n is the number of comparisons used by basic randomized quicksort on n elements, and $q_n = E[Q_n]$.

Proof We aim to establish a lower bound for the probability that the number of comparisons Q_n in basic quicksort exceeds its expectation q_n by a factor of $(1+\varepsilon)$, where $\varepsilon > 0$. Specifically, we prove that as $n \to \infty$,

$$\Pr\left[Q_n > (1+\varepsilon)q_n\right] \geq \exp\left\{-2\varepsilon \ln n \left(\ln\ln n + O(\ln\ln\ln n)\right)\right\}.$$

Our strategy involves constructing a specific event A, corresponding to a sequence of unfavorable pivot choices in the initial levels of the quicksort partition tree, which forces Q_n to be sufficiently large. We then compute the probability of this event and verify that it implies the desired bound on Q_n, using Chebyshev's inequality and the law of total probability to ensure rigor and completeness.

Consider the partition tree of quicksort, where each node represents a sublist. The root (node 1) corresponds to the initial list of n distinct keys. At each node, a pivot is selected uniformly at random, partitioning the sublist into a left child (keys less than the pivot) and a right child (keys greater than the pivot). To inflate the number of comparisons, we define the event A by enforcing small left sublists along the leftmost path for the first k levels.

Define the following parameters and sets:

- *Number of levels*: $k = k(n) = \lceil \kappa\varepsilon \ln n \rceil$, where $\kappa = 2 + \mu(n)$ and $\mu(n) = \frac{3 \ln\ln\ln n}{\ln\ln n}$,

- *Maximum left sublist size*: $l = l(n) = \left\lfloor \frac{\lambda n}{\varepsilon \ln n} \right\rfloor$, where $\lambda = \frac{\mu(n)}{3} = \frac{\ln\ln\ln n}{\ln\ln n}$,

- *Leftmost path nodes*: $J = \{2^0+1, 2^1+1, \ldots, 2^{k-1}+1\} \cup \{2^k\}$, which includes the leftmost nodes at depths 0 to $k-1$ (i.e., nodes 1, 3, 5, ..., up to $2^{k-1}+1$) and the right child at depth k (node 2^k),

- *Sublist size vectors*: $\mathcal{L}$ is the set of vectors $\underline{l} = (l_j : j \in J)$ such that $0 \leq l_j \leq l$ for $j \in J \backslash \{2^k\}$, and $\sum_{j \in J} l_j = n - k$,

- *Specific event*: For each $\underline{l} \in \mathcal{L}$, define $A(\underline{l})$ as the event where the left sublist size $L_j = l_j$ for all $j \in J$,

- *Total event*: $A = \bigcup_{\underline{l} \in \mathcal{L}} A(\underline{l})$.

Under event A, the left sublists at nodes along the leftmost path (depths 0 to $k-1$) have sizes at most l, and the remaining keys are distributed such that the total size of sublists at depth k equals $n-k$, with the majority concentrated at node 2^k.

We compute a lower bound for $\Pr[A]$. Pivot selections are independent across levels. For node 2^i+1 at depth i (where $i = 0, 1, \ldots, k-1$), with sublist size n_i, the pivot's rank is uniform over $\{0, 1, \ldots, n_i - 1\}$. The left sublist size $L_{2^{i+1}}$ is the pivot's rank, so:

$$\Pr[L_{2^{i+1}} \leq l] = \frac{l+1}{n_i},$$

since there are $l+1$ possible ranks (0 to l) yielding a left sublist of size at most l. Under A, $n_i = n - \sum_{j=0}^{i-1} l_{2^j+1} - i$, but for a conservative bound, note that $n_i \leq n$, and typically $n_i \geq n - i(l+1)$. Thus:

$$\Pr[L_{2^{i+1}} \leq l] \geq \frac{l+1}{n}.$$

For a specific $A(\underline{l})$, the probability is the product over the k splits:

$$\Pr[A(\underline{l})] = \prod_{i=0}^{k-1} \frac{l_{2^{i+1}} + 1}{n_i},$$

but since $l_{2^{i+1}} \leq l$, we bound:

$$\Pr[A(\underline{l})] \geq \left(\frac{l+1}{n}\right)^k.$$

Since $A = \bigcup_{\underline{l} \in \mathcal{L}} A(\underline{l})$, and the number of possible vectors is at most $(l+1)^k$ (an overestimate), we use the probability of a typical event adjusted by the constraint $\sum l_j = n - k$. For simplicity, we take:

$$\Pr[A] \geq \left(\frac{l+1}{n}\right)^k,$$

which holds as a lower bound by considering the contribution of all consistent $A(\underline{l})$ events.

Substitute the asymptotic expressions:

- $l + 1 \approx \frac{\lambda n}{\varepsilon \ln n} = \frac{n \ln \ln \ln n}{\varepsilon \ln n \ln \ln n}$
- $k \approx \kappa \varepsilon \ln n = \left(2 + \frac{3 \ln \ln \ln n}{\ln \ln n}\right) \varepsilon \ln n.$

Thus:

$$\frac{l+1}{n} \approx \frac{\ln \ln \ln n}{\varepsilon \ln n \ln \ln n},$$

$$\Pr[A] \geq \left(\frac{l+1}{n}\right)^k \approx \left(\frac{\ln \ln \ln n}{\varepsilon \ln n \ln \ln n}\right)^{\left(2 + \frac{3 \ln \ln \ln n}{\ln \ln n}\right) \varepsilon \ln n}.$$

Compute the exponent:

$$\ln \Pr[A] \geq k \ln\left(\frac{l+1}{n}\right) \approx \kappa \varepsilon \ln n \left(\ln \frac{\ln \ln \ln n}{\varepsilon \ln n \ln \ln n}\right),$$

$$= \kappa \varepsilon \ln n \left(\ln \ln \ln n - \ln(\varepsilon \ln n \ln \ln n)\right),$$

$$= \kappa \varepsilon \ln n \left(\ln \ln \ln n - \ln \varepsilon - \ln \ln n - \ln \ln \ln n - \ln 1\right).$$

Simplify:

$$= \kappa \varepsilon \ln n \left(-\ln \ln n + (\ln \ln \ln n - \ln \ln \ln n) - \ln \varepsilon\right),$$

$$= -\kappa \varepsilon \ln n \ln \ln n + \kappa \varepsilon \ln n (\ln \ln \ln n - \ln \ln \ln n - \ln \varepsilon).$$

Since $\kappa = 2 + \frac{3 \ln \ln \ln n}{\ln \ln n}$:

$$\ln \Pr[A] \geq -\left(2 + \frac{3 \ln \ln \ln n}{\ln \ln n}\right) \varepsilon \ln n \ln \ln n + O(\varepsilon \ln n \ln \ln \ln n),$$

$$= -2\varepsilon \ln n \ln \ln n - 3\varepsilon \ln n \frac{\ln \ln \ln n}{\ln \ln n} \ln \ln n + O(\varepsilon \ln n \ln \ln \ln n),$$

$$= -2\varepsilon \ln n \ln \ln n - 3\varepsilon \ln n \ln \ln \ln n + O(\varepsilon \ln n \ln \ln \ln n).$$

Thus:

$$\Pr[A] \geq \exp\left\{-2\varepsilon \ln n \left(\ln\ln n + O(\ln\ln\ln n)\right)\right\},$$

where the $O(\ln\ln\ln n)$ term absorbs lower-order contributions.

We now show that $Q_n > (1+\varepsilon)q_n$ under event A with high probability. Split Q_n into:

- Q'_n: Comparisons along the leftmost path up to depth $k-1$ (nodes $1, 3, \dots, 2^{k-1}+1$),

- Q''_n: Comparisons in all remaining subtrees (including the large subtree at node 2^k and smaller subtrees).

Bounding Q'_n:

$$Q'_n = \sum_{i=0}^{k-1} (n_i - 1),$$

where n_i is the sublist size at node $2^i + 1$. Under A: - $n_0 = n$, - $n_1 = n - l_1 - 1$, - $n_i = n - \sum_{j=0}^{i-1} l_{2^j+1} - i$.

Since $l_j \leq l$:

$$n_i \geq n - i(l+1),$$

$$Q'_n \geq \sum_{i=0}^{k-1} (n - i(l+1) - 1) = kn - \sum_{i=0}^{k-1} i(l+1) - k,$$

$$= kn - \frac{k(k-1)}{2}(l+1) - k.$$

Substitute $k \approx 2\varepsilon \ln n$, $l + 1 \approx \frac{n \ln\ln\ln n}{\varepsilon \ln n \ln\ln n}$:

$$\frac{k(k-1)}{2}(l+1) \approx \frac{(2\varepsilon \ln n)^2}{2} \cdot \frac{n \ln\ln\ln n}{\varepsilon \ln n \ln\ln n} = \frac{2\varepsilon n \ln n \ln\ln\ln n}{\ln\ln n},$$

$$Q'_n \geq kn - O\left(\frac{n \ln n \ln\ln\ln n}{\ln\ln n}\right).$$

Bounding Q''_n:

- *Size of the large subtree*: At node 2^k, $L_{2^k} = n - \sum_{j \in J \setminus \{2^k\}} l_j - (k-1) \geq n - kl$,

- *Smaller subtrees:* At most k subtrees of size $\leq l$.

The expected number of comparisons in a quicksort tree of size m is approximately $2m \ln m$. Thus:

$$E[Q''_n] \geq 2(n - kl)\ln(n - kl) + \sum_{\text{small subtrees}} 2 l_j \ln l_j,$$

$$\geq 2(n - kl)\ln(n - kl),$$

since the contribution of small subtrees is non-negative. Compute:

$$kl \approx (2\varepsilon \ln n) \cdot \frac{n \ln\ln\ln n}{\varepsilon \ln n \ln\ln n} = \frac{2n \ln\ln\ln n}{\ln\ln n},$$

$$n - kl \approx n\left(1 - \frac{2\ln\ln\ln n}{\ln\ln n}\right),$$

$$\ln(n - kl) = \ln n + \ln\left(1 - \frac{2\ln\ln\ln n}{\ln\ln n}\right) \approx \ln n - \frac{2\ln\ln\ln n}{\ln\ln n},$$

$$E[Q_n''] \geq 2n\ln n - O\left(\frac{n\ln\ln\ln n}{\ln\ln n}\right).$$

Since $q_n \approx 2n\ln n$, we need Q_n'' to concentrate around its expectation. The variance of quicksort comparisons for a tree of size m is $O(m^2)$, so:

$$\mathrm{Var}(Q_n'') \leq cn^2,$$

for some constant c, summing variances of independent subtrees (dominated by the subtree of size $n - kl \approx n$).

Apply *Chebyshev's inequality*:

$$\Pr\left[|Q_n'' - E[Q_n'']| \geq t\right] \leq \frac{\mathrm{Var}(Q_n'')}{t^2} \leq \frac{cn^2}{t^2}.$$

Choose $t = n\ln\ln n$:

$$\Pr\left[Q_n'' < E[Q_n''] - n\ln\ln n\right] \leq \frac{cn^2}{(n\ln\ln n)^2} = \frac{c}{(\ln\ln n)^2}.$$

As $n \to \infty$, $\frac{c}{(\ln\ln n)^2} \to 0$, so with probability approaching 1:

$$Q_n'' \geq E[Q_n''] - n\ln\ln n \geq 2n\ln n - n\ln\ln n - O\left(\frac{n\ln\ln\ln n}{\ln\ln n}\right).$$

Total Q_n:

$$Q_n = Q_n' + Q_n'' \geq kn - \frac{k(k-1)}{2}(l+1) + 2n\ln n - n\ln\ln n - O\left(\frac{n\ln\ln\ln n}{\ln\ln n}\right),$$

$$\geq 2\varepsilon\ln n \cdot n + 2n\ln n - O(n\ln\ln n).$$

Compare with $(1+\varepsilon)q_n \approx (1+\varepsilon)2n\ln n$:

$$Q_n - (1+\varepsilon)q_n \geq 2\varepsilon n\ln n - O(n\ln\ln n),$$

Since $2\varepsilon n\ln n \gg n\ln\ln n$ as $n \to \infty$, $Q_n > (1+\varepsilon)q_n$ holds with probability approaching 1 under A:

$$\Pr[Q_n > (1+\varepsilon)q_n \mid A] \to 1.$$

Using the law of total probability:

$$\Pr[Q_n > (1+\varepsilon)q_n] = \Pr[A] \cdot \Pr[Q_n > (1+\varepsilon)q_n \mid A] + \Pr[\text{not } A] \cdot \Pr[Q_n > (1+\varepsilon)q_n \mid \text{not } A].$$

Since all probabilities are non-negative:

$$\Pr[Q_n > (1+\varepsilon)q_n] \geq \Pr[A] \cdot \Pr[Q_n > (1+\varepsilon)q_n \mid A].$$

Substitute:

$\Pr[A] \geq \exp\left\{-2\varepsilon\ln n\,(\ln\ln n + O(\ln\ln\ln n))\right\}$, $\Pr[Q_n > (1+\varepsilon)q_n \mid A] \to 1$ as $n \to \infty$.

Thus, for sufficiently large n:

$$\Pr[Q_n > (1+\varepsilon)q_n] \geq \exp\{-2\varepsilon \ln n\,(\ln\ln n + O(\ln\ln\ln n))\} \cdot (1 - o(1)),$$

$$\geq \exp\{-2\varepsilon \ln n\,(\ln\ln n + O(\ln\ln\ln n))\},$$

where the $o(1)$ term vanishes asymptotically, completing the proof. □

6 Conclusion

In this report, we have conducted a comprehensive probabilistic analysis of the Randomized QuickSort algorithm, focusing on the concentration of its comparison count Q_n around its expected value $q_n \approx 2n\ln n$. Through a series of lemmas, we have derived tight bounds on the deviation $|Q_n - q_n|$, leveraging martingale concentration inequalities and a detailed examination of the recursion tree.

$$\Pr\left[\left|\frac{Q_n}{q_n} - 1\right| > \varepsilon\right] \leq n^{-2\varepsilon(\ln\ln n + O((\ln\ln\ln\ln n)))}$$

$$\Pr\left[Q_n > (1+\varepsilon)q_n\right] \geq \exp\{-2\varepsilon \ln n\,(\ln\ln n + O(\ln\ln\ln n))\},$$

These theoretical bounds are substantiated by our experimental results in Section 3. The distribution plots (e.g., Figure 3) and frequency histograms (e.g., Figure 6) illustrate that Q_n concentrates tightly around q_n, with empirical tail probabilities (e.g., Figure 4) aligning with the exponential decay predicted by Lemma 2.8. The average comparison counts (e.g., Figure 2) closely follow q_n, while the heatmap (e.g., Figure 5) confirms that large deviations are exceptionally rare, consistent with both upper and lower bounds.

In summary, our analysis refines the concentration properties of Randomized QuickSort beyond previous results, achieving the tightest known bounds on Q_n's deviation. By integrating advanced probabilistic tools—such as Azuma's inequality and careful asymptotic parameter tuning—we provide a deeper understanding of the algorithm's recursive behavior. These findings reinforce Randomized QuickSort's reliability and efficiency, solidifying its foundational role in algorithmic design and analysis.

References

[1] C. McDiarmid and R. Hayward. Strong concentration for quicksort. In Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 416–425, 1994.

[2] L. Devroye. A note on the height of binary search trees. Journal of the ACM, 33(3):489–498, July 1986.

[3] S. Roura and C. Vallejo. Concentration bounds for the quicksort algorithm. Theoretical Computer Science, 256(1-2):31–50, 2001.

[4] J. A. Fill and S. Janson. Quicksort asymptotics. Journal of Algorithms, 44(1):4–28, 2002.

[5] R. Sedgewick. The analysis of quicksort programs. Acta Informatica, 7(4):327–355, 1977.

[6] T. Hoare. Quicksort. The Computer Journal, 5(1):10–16, 1962.