CS271 Randomness & Computation **Spring 2018**

Lecture 21: April 3

Instructor: Alistair Sinclair

Disclaimer: *These notes have not been subjected to the usual scrutiny accorded to formal publications. They may be distributed outside this class only with the permission of the Instructor.*

21.1 Sharp Concentration for Quicksort

In this lecture we use Azuma's inequality to analyze the randomized *Quicksort* algorithm. Quicksort takes as input a set S of numbers, which can be assumed to be distinct without loss of generality, and sorts the set S as follows: it picks a pivot $x \in S$ uniformly at random, then it partitions the set S as $S_{<x} = \{y \in S \mid y < x\}$ and $S_{>x} = \{y \in S \mid y > x\}$, and recursively sorts $S_{<x}$ and $S_{>x}$. The formal description of the algorithm is as follows:

Input: A set S of distinct numbers.
Output: S in sorted order.

1. Pick an element $x \in S$ uniformly at random.
2. Partition S into $S_{<x} = \{y \in S \mid y < x\}$ and $S_{>x} = \{y \in S \mid y > x\}$ by comparing each element of S to x.
3. *Quicksort*($S_{<x}$) and *Quicksort*($S_{>x}$).

The worst case running time of the above algorithm for a set S of size n is $O(n^2)$, since in every invocation of the *Quicksort* algorithm the pivot chosen could be the least or the greatest element. However, it is well known that the expected running time of the algorithm is $O(n \ln n)$, and moreover this is the behavior typically observed in practice. In other words, it is widely believed that the running time is *sharply concentrated* about its mean value. In this lecture we will give a rigorous proof of this fact using Azuma's inequality. Surprisingly, despite the fact that Quicksort has been around since the early 1960s [Hoa62] and has had whole books written about it [Sed80], this result was proved only much more recently, by Hayward and McDiarmid [HM96]. It remains one of the more sophisticated uses of martingales in the analysis of algorithms.

Notation. We start with some notation.

$$Q_n = \#\text{ of comparisons of the } \textit{Quicksort} \text{ algorithm on a set } S \text{ of size } n.$$

Thus Q_n is a random variable. We write $q_n = \mathrm{E}[Q_n]$ for its expectation.

Since each element of $x \in S$ is equally likely to be the pivot, the size of the set $S_{<x}$ is distributed uniformly in $\{0, 1, \ldots, n-1\}$. Thus we obtain the following recurrence for q_n:

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

The precise solution of the above recurrence is given by

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

where γ is Euler's constant.

We are interested in evaluating the tail probability $\Pr[|Q_n - q_n| \geq \varepsilon \cdot q_n]$, as a function of ε and n, i.e., we want to determine the probability that Q_n deviates more than a fraction ε from its mean q_n as a function of ε and n. Using the second moment method and the fact that $\mathrm{Var}(Q_n) = \Theta(n^2)$, Knuth [Knu73] and Sedgewick [Sed80] had observed that this probability was $O((\epsilon \ln n)^{-2})$. Later, Hennequin [Hen89] refined this result to $O((\epsilon \ln n)^{-4})$. In this lecture we will prove the following tight result due to Hayward and McDiarmid [HM96], which completely resolves the issue.

Theorem 21.1 (Hayward/McDiarmid) *For any fixed* $\varepsilon > 0$*,* $\Pr[|Q_n - q_n| \geq \varepsilon q_n] = n^{-(2+o(1))\varepsilon \ln \ln n}$.

Actually we will just prove the (more interesting) upper bound, and will omit the proof that the bound is tight.

We can view the evolution of *Quicksort* as a random binary tree (Figure 21.1), where at each node we branch to two subtrees corresponding to the two sublists to be sorted. Note that the composition of these sublists is random and depends on the choice of pivot element at the node. We shall label the nodes of the tree top to bottom, left to right, as shown in the figure. Given a node i in the tree we denote by L_i the length of the list at node i (i.e., the length of the set to be sorted at node i). Observe that the lists at any level k are disjoint, so for all levels k we have

$$\sum_{j \text{ at level } k} L_j \leq n.$$

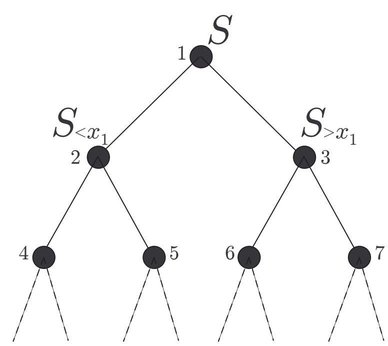

Figure 21.1: *Quicksort*

We now state some basic facts which we will use repeatedly in the main proof.

Fact 1. Let $M_k^n = \max\{L_j \ : \ j \text{ at level } k\}$. For any α such that $0 < \alpha < 1$, and $k \geq \ln(\frac{1}{\alpha})$, we have

$$\Pr[M_k^n \geq \alpha n] \leq \alpha \cdot \left(\frac{2e \ln(\frac{1}{\alpha})}{k}\right)^k.$$

This Fact, which is due to Devroye [Dev86] and whose proof is fairly standard but omitted, provides a bound on the probability that the number of elements at any node at a given level will be unusually large. (Note that the expected value of L_j for a node at level k is about $n/2^k$, so the ratio $\frac{\ln(1/\alpha)}{k}$ appearing in the bound is natural.)

Fact 2. Let $V_n = \{[(n-1) + q_{j-1} + q_{n-j}] - q_n \ : \ j \in \{1, 2, \ldots, n\}\}$. Then $V_n \subseteq [-n, n]$.

This Fact bounds the change in the expected number of comparisons after fixing the first pivot to be j: the choice of pivot can affect this expectation by at most $\pm n$. Note that the expression $[(n-1)+q_{j-1}+q_{n-j}]$ is the expected number of comparisons required by the algorithm *given* the first choice of j (cf. the original recurrence relation for q_n). [Exercise: Prove Fact 2 using the recurrence for q_n.]

We need one final piece of notation before embarking on the main proof. Define random variables

$$\begin{aligned} H_k &= \text{outcomes of comparisons at level } k \text{ of tree;} \\ \mathcal{H}^{(k)} &= (H_0, H_1, \ldots, H_{k-1}), \end{aligned}$$

i.e., $\mathcal{H}^{(k)}$ consists of the history of the process down to level k, including the composition of the lists at level k, but not the outcome of the comparisons at level k. We call $\mathcal{H}^{(k)}$ a *k-history.*

Two-phase analysis of the algorithm. We will analyze the evolution of the binary tree of the *Quicksort* algorithm in two phases (Figure 21.2). In the first phase we consider the tree down to level k_1 and use Fact 2 to crudely bound the deviation of the expected number of comparisons (conditioned on the history $\mathcal{H}^{(k_1)}$), from q_n. Then we analyze the tree from level k_1 down to level $k_2 > k_1$, where k_2 is chosen large enough so that w.h.p. the algorithm has completed by then. The analysis in this second phase is more sensitive, and will make use of a martingale argument and Azuma's inequality. Finally we will combine the analyses of the two phases to obtain the desired result.

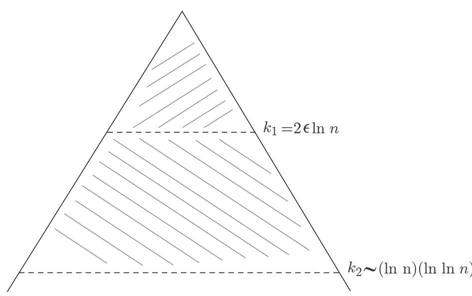

Figure 21.2: Two-phase analysis

We now prove a lemma for the analysis of the first phase.

Lemma 21.2 *For any k-history $\underline{h}$ we have*

$$\left|\mathrm{E}[Q_n \mid \mathcal{H}^{(k)} = \underline{h}] - q_n\right| \le kn.$$

Proof: The proof follows from Fact 2. Consider a node j at level k and let the length of the list at node j be L_j. By Fact 2, the maximum change in the expectation if we go down a level (i.e., to the children of L_j) is bounded by $\pm|L_j|$. Since the sum of the lengths of the lists at nodes of level k is at most n, it follows that when we go down from level k to $k+1$, the change in the expectation is bounded by $\pm n$. The result then follows easily by induction on the levels. ∎

Note that the statement of Lemma 21.2 is deterministic and not probabilistic. Also, in order to use the lemma in the proof of Theorem 21.1, we will certainly need to chose the depth k_1 so that $k_1 n \le \varepsilon q_n$, i.e., k_1 will have to be at most $2\varepsilon \ln n$.

We now prove the main lemma, which forms the heart of the analysis.

Lemma 21.3 *Let $0 < k_1 < k_2$ be integers, $0 < \alpha < 1$, and $\underline{h}$ be a k_1-history such that $M^n_{k_1} \le \alpha n$. Then we have*

$$\Pr\left(\big|\mathrm{E}\big[Q_n \mid \mathcal{H}^{(k_2)}\big] - \mathrm{E}\big[Q_n \mid \mathcal{H}^{(k_1)} = \underline{h}\big]\big| \ge \lambda \mid \mathcal{H}^{(k_1)} = \underline{h}\right) \le 2\exp\left(-\frac{\lambda^2}{2(k_2-k_1)\alpha n^2}\right).$$

Before the formal proof of Lemma 21.3 we describe the proof structure and a few key points. We will analyze the process between step k_1 and k_2 as a martingale of length $k_2 - k_1$, indexed by levels $\{k_1 + i : 0 \le i \le k_2 - k_1\}$. Note the striking similarity of the tail bound in Lemma 21.3 and the bound in Azuma's inequality (which is of the form $2\exp\big(-\frac{\lambda^2}{2\sum_i c_i^2}\big)$, where the c_i's are bounds on the martingale differences $|X_i - X_{i-1}|$). The interesting term in the bound of Lemma 21.3 is the denominator of the exponential term: the factor $(k_2 - k_1)$ is just the length of the martingale, and the term αn^2 will play the role of the c_i's as in Azuma's inequality. However, rather than arguing that the martingale differences are absolutely bounded by αn^2, we will instead only be able to show that their conditional expectations are suitably bounded; this additional flexibility is allowed by the proof of Azuma's inequality.

We now present the proof of Lemma 21.3.

Proof of Lemma 21.3: We define the Doob martingale as follows:

$$X_0 = \mathrm{E}\big[Q_n \mid \mathcal{H}^{(k_1)} = \underline{h}\big]; \qquad X_i = \mathrm{E}\big[Q_n \mid \mathcal{H}^{(k_1+i)}, \mathcal{H}^{(k_1)} = \underline{h}\big].$$

We first observe that the bound of Azuma's inequality can be obtained under the weaker assumption that

$$\mathrm{E}\big[\exp\big(t(X_i - X_{i-1})\big) \mid \mathcal{F}_i\big] \le \exp\big(\tfrac{1}{2}t^2 c_i^2\big), \tag{21.1}$$

instead of the bounded difference assumption (i.e., $|X_i - X_{i-1}| \le c_i$). (Exercise: go back to the proof of Azuma's inequality in Lecture 19 and check this! Recall also that we made a similar detour in our use of Azuma in our analysis of the TSP problem in the last lecture.)

Now we need to verify that (21.1) holds for our martingale, with $c_i = \alpha n^2$ for all i. To this end, let us write

$$X_i - X_{i-1} = \sum_{j \text{ at level } i-1} T_j,$$

where T_j denotes the contribution to $X_i - X_{i-1}$ coming from node j. Observe that the T_j's are independent. Also, by symmetry $\mathrm{E}[T_j] = 0$, and if the length of the list at node j is L_j, then by Fact 2 we have $T_j \in [-L_j, L_j]$. We may therefore use the same convexity argument as in the proof of Azuma's inequality (Lemma 19.4 in Lecture 19) to obtain

$$\mathrm{E}[\exp(tT_j) \mid L_j] \le \exp\big(\tfrac{1}{2}t^2 L_j^2\big).$$

Hence

$$\begin{aligned}
\mathrm{E}[\exp(t(X_i - X_{i-1})) \mid \{L_j\}] &= \prod_{j \text{ at level } i-1} \mathrm{E}[\exp(tT_j) \mid L_j] \qquad \text{(by independence)}\\
&\le \prod_{j \text{ at level } i-1} \exp\big(\tfrac{1}{2}t^2 L_j^2\big)\\
&= \exp\Big(\tfrac{1}{2}t^2 \textstyle\sum_{j \text{ at level } i-1} L_j^2\Big)\\
&\le \exp\big(\tfrac{1}{2}t^2 \alpha n^2\big).
\end{aligned}$$

The last inequality is obtained as follows: observe that by the assumption of the lemma on the k_1-history $\underline{h}$ we have $M^n_{k_1} \le \alpha n$, and as we go down the tree the maximum length of the lists can only decrease. Hence we have

$$\max_{j \text{ at level } i-1} L_j \le \alpha n.$$

Therefore,

$$\sum_{j \text{ at level } i-1} L_j^2 \leq (\max_{j \text{ at level } i-1} L_j) \cdot (\sum_{j \text{ at level } i-1} L_j) \leq \alpha n \cdot n = \alpha n^2.$$

Lemma 21.3 now follows by Azuma's inequality, using αn^2 in place of each of the c_i. ■

Finally, we complete the analysis to prove Theorem 21.1. By a union bound we have

$$\begin{aligned}
\Pr\left[|Q_n - q_n| \geq k_1 n + \lambda\right] \quad &\leq \quad \Pr\left[M_{k_2}^n \geq 2\right] + \Pr\left[M_{k_1}^n > \alpha n\right] + \\
&\qquad \Pr\left[\left|\mathrm{E}\left[Q_n|\mathcal{H}^{(k_2)}\right] - \mathrm{E}\left[Q_n|\mathcal{H}^{(k_1)}\right]\right| \geq \lambda \,\middle|\, \mathcal{H}^{(k_1)}, M_{k_1}^n \leq \alpha n\right] \\
&\leq \quad \frac{2}{n}\left(\frac{2\varepsilon \ln\left(\frac{n}{2}\right)}{k_2}\right)^{k_2} + \alpha\left(\frac{2\varepsilon \ln\left(\frac{1}{\alpha}\right)}{k_1}\right)^{k_1} + 2\exp\left(-\frac{\lambda^2}{2(k_2-k_1)\alpha n^2}\right). \qquad (21.2)
\end{aligned}$$

The three terms of the union bound can be explained as follows: the first term denotes the probability that the process does not end within k_2 steps; the second term denotes the probability that the maximum length of a list after k_1 steps is greater than αn, and the last term denotes the probability that the deviation of the expectation between k_1 and k_2 steps is at most λ, given that $M_{k_1}^n \leq \alpha n$ for the k_1-history. Inequality (21.2) is obtained from these three terms as follows: the first inequality follows from Fact 1 by substituting $\alpha = \frac{2}{n}$, the second term follows directly from Fact 1, and the last inequality follows from Lemma 21.3.

Finally, to obtain Theorem 21.1 from inequality (21.2) we need to choose the parameters to satisfy the following constraints:

$$\begin{aligned}
k_1 n + \lambda \quad &\leq \quad \varepsilon q_n \sim 2\varepsilon n \ln n; \\
\Pr[|Q_n - q_n| \geq k_1 n + \lambda] \quad &\leq \quad n^{-(2+o(1))\varepsilon \ln\ln n}.
\end{aligned}$$

After some slightly messy but straightforward calculations, we arrive at the following choices of the parameters to satisfy the constraints:

$$\begin{aligned}
k_1 \quad &= \quad 2\varepsilon \ln n - \frac{2\lambda}{n}; \\
k_2 \quad &= \quad (\ln n)(\ln\ln n); \\
\lambda \quad &= \quad \frac{\varepsilon n \ln n}{\ln\ln n}; \\
\alpha \quad &= \quad \frac{\varepsilon^2}{(\ln\ln n)^5}.
\end{aligned}$$

Plugging in the above values in each of the three terms above, we get (ignoring various constants and lower order terms):

$$\begin{aligned}
\Pr[M_{k_2}^n \geq 2] \quad &\sim \quad \exp(-(\ln n)(\ln\ln n)(\ln\ln\ln n)) \\
&= \quad \exp(-\omega(\ln n \ln\ln n)) \\
\Pr\left[\mathrm{E}\left[Q_n|\mathcal{H}^{(k_2)}\right] - \mathrm{E}\left[Q_n|\mathcal{H}^{(k_1)}\right] \geq \lambda \,\middle|\, \mathcal{H}^{(k_1)}, M_{k_1}^n \leq \alpha n\right] \quad &\sim \quad \exp(-(\ln\ln n)^2 \ln n) \\
&= \quad \exp(-\omega(\ln n \ln\ln n)) \\
\Pr[M_{k_1}^n \geq \alpha n] \quad &\sim \quad \exp\big(-2\epsilon \ln n \ln\ln n + O(\ln\ln\ln n)\big).
\end{aligned}$$

From this we see that the sum is dominated by the $\Pr[M_{k_1}^n \geq \alpha n]$ term, and this is of the order claimed in the statement of Theorem 21.1.

Exercise: Check the above calculations by plugging in the given values for $k_1, k_2, \lambda, \alpha$. And, to get a better understanding, reverse-engineer these parameter choices from equation (21.2), starting with k_1 and k_2.

References

[Dev86] L. Devroye, “A note on the height of binary search trees,” *Journal of the ACM*, 33(3):489–498, 1986.

[GS92] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press, 1992.

[HM96] R. Hayward and C. McDiarmid, “Large Deviations for Quicksort,” *J. Algorithms* , 21(3):476–507, 1996.

[Hen89] P. Hennequin, “Combinatorial analysis of quicksort algorithm,” *Theoretical Informatics and Applications*, 23(3):317–333, 1989.

[Hoa62] C.A.R. Hoare, “Quicksort,” *Computer Journal*, 5:10–15, 1962.

[Knu73] D. Knuth, *The Art of Computer Programming*, Addison-Wesley, 1973.

[Sed80] R. Sedgewick, *Quicksort*, Garland Publishing Inc., 1980.