Lesson 1. Notes on Ramdomized Quicksort - a Geometric Vision

M. Pellegrini*

January 14, 2004

1 Disclaimer

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

2 Introduction

This lecture is a mix of [MR95, Chapter 1] and [Mul93, Chapter 1].

2.1 Analysis from [MR95, Chapter 1]

Consider the set S of n distinct real numbers.

Consider the permutation $P : [1, n] \to S$ which is the temporal sequence of pivot choices of randomized quicksort. This permutation is chosen uniformly at random in the set $Perm(n)$ of all permutation of the numbers in $[1, n]$.

Consider the permutation $Q : [1, n] \to S$ in which each element of S is associated to its rank.

Observation 1: Elements $Q(i)$ and $Q(j)$ are compared by Randomized Quicksort if and only if in P no element $Q(l)$ for $i < l < j$ appears before $Q(i)$ and $Q(j)$.

Observation 2: Any element in the set $Q(i), Q(i + 1), ..Q(j)$ has the same probability of appearing before the others in P.

Since there are $j - i + 1$ elements between $Q(i)$ and $Q(j)$ extremes included and we have a comparison between $Q(i)$ and $Q(j)$ only if the first element to appear in P is either $Q(i)$ or $Q(j)$. The probability of comparing them is: $p_{ij} = 2/(j - i + 1)$.

*Istituto di Informatica e Telematica del CNR, Via G. Moruzzi 1, 56100-Pisa (Italy). pellegrini@iit.cnr.it

Define X_{ij} the indicator random variable (depending on the choice of P that has value 1 id $Q(i)$ and $Q(j)$ are compared by randomized quicksort, 0 otherwise.

The number of comparison is thus $\sum_{i=1}^{n}\sum_{j>i} X_{ij}$. The expected number is:

$$E[\sum_{i=1}^{n}\sum_{j>i} X_{ij}] = \sum_{i=1}^{n}\sum_{j>i} E[X_{ij}]$$

by linearity of the expectation operator. But $E[X_{ij}] = p_{ij}$ so

$$\sum_{i=1}^{n}\sum_{j>i} p_{ij} = \sum_{i=1}^{n}\sum_{j>i} \frac{2}{j-i+1} \leq 2\sum_{i=1}^{n}\sum_{k=1}^{n} \frac{1}{k}$$

Thus calling $H_n = \sum_{k=1}^{n} \frac{1}{k}$ the n-th Harmonic number we have that the expected number of comparisons is bounded by $2nH_n$. Since $H_n \approx \log n + \Theta(1)$ the number of comparisons is $O(n \log n)$.

2.2 Analysis from [Mul93, Chapter 1]

[See Slides]

Note that the analysis of [Mul93, Chapter 1] is somewhat more cumbersome, but is not dependent on the notion of rank which might not be well defined in higher dimensional problems.

2.3 Tools used

1. Linearity of Expectation.

2. Asymptotics of harmonic numbers.

2.4 Notes

This backward analysis of Quicksort has been popularized in a seminal paper by R. Seidel [Sei92]. In this work Seidel gives credit for the first application of such an analytic tool to Paul Chew who used it to analyze randomized incremental algorithm for the Delaunay Triangulation of points in convex position in the plane (this algorithm uses expected linear time). This paper shows the analysis for the following algorithms : computing intersecting line segments in output sensitive time, planar convex hull, randomized quick-sort (with also a tail bound), Seidel's LP algorithm, Welzl's Min-disk algorithm. In Section 9 of [Sei92] gives the expected time bound for convex hull in any dimension simplifying the one in [CS89]. Actually such analysis can be extended easily to any incremental construction.

References

[CS89] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete and Computational Geometry*, 4(1):387–421, 1989.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[Mul93] K. Mulmuley. *Computational Geometry, an Introduction through Randomized Algorithms.* Prentice Hall, 1993.

[Sei92] Raimund Seidel. Backwards analysis of randomized geometric algorithms. Technical Report Technical Report TR92 -014,, International Computer Science Institute,, University of California at Berkeley, Berkeley, California, February 1992.