Randomized Algorithms, Summer 2016 Lecture 13 (12 pages)

Limited Randomness

Instructor: Thomas Kesselheim and Kurt Mehlhorn

Contents

1 Introduction

Random bits are costly. Note that we never discussed how one can generate random bits or where one can buy them. If random bits are costly, it makes sense to use the thriftily and to ask how many one needs in order to obtain a certain effect.

Consider randomized quicksort. One way of describing it is as follows. One first permutes the input randomly and then applies deterministic quicksort to it. This algorithm requires to choose a random permutation of n items. As there are $n!$ such permutations, it requires $\Omega(n \log n)$ random bits. Can we do with less?

We will see today that $O(\log n)$ random bits suffice to guarantee the $O(n \log n)$ running time of randomized quicksort. I am basing my lecture on the paper

"M. Knudsen, M. Stöckel: Quicksort, Largest Bucket, and Min-Wise Independent Hashing with Limited Independence, ESA 2015".

The paper shows that 4-wise independence suffices to guarantee the $O(n \log n)$ expected running time of randomized quicksort. It leaves four open problems:

- the result holds for a variant of quicksort, not for the randomized quicksort as we usually know it;
- the constants are not worked out.
- is three-wise independence enough?
- it only deals with the expected running time, and does not also give a bound on the variance or a tail bound. Remember that randomized quicksort runs in time $O(n \log n)$ with high probability. I conjecture that going to higher independence will decrease the variance. My former PhD-student Adrian Neumann has performed experiments that confirm this intuition.

  An earlier paper on the subject is by Karloff and Raghavan.

2 Limited Independence

Definition 13.1. *Let X_0 to X_{n-1} be a sequence of random variables with values in a set R. The sequence is 4-independent if for any subset $I = \{i_1, i_2, i_3, i_4\}$ of four distinct variables and any four tuple $(y_1, \ldots, y_4)$ of values in R,*

$$\mathbf{Pr}[X_{i_1} = y_1,\ X_{i_2} = y_2,\ X_{i_3} = y_3,\ X_{i_4} = y_4] = \prod_{1 \le \ell \le 4} \mathbf{Pr}[X_{i_\ell} = y_\ell]\,.$$

Let $\mathcal{H}$ be a set of functions from $[n]$ to $[m]$; here $[n] = \{0, \ldots, n-1\}$. We will refer to the functions in $\mathcal{H}$ as hash functions. By a random hash function, we mean a random function $h \in \mathcal{H}$; each function is chosen with probability $1/|\mathcal{H}|$. We can generate a sequence X_0 to X_{n-1} by choosing a function $h \in \mathcal{H}$ and setting $X_i = h(i)$. The definition above plus the requirement that $h(i)$ should be a random element in the range of h then translates into:

Definition 13.2. *The family $\mathcal{H}$ is* k-independent *if for any k distinct elements x_1 to x_k in $[n]$ and any k elements y_1 to y_k in $[m]$*

$$|\{h \in \mathcal{H} \mid h(x_i) = y_i \text{ for } 1 \le i \le k\}| = \frac{|\mathcal{H}|}{m^k}.$$

Note that the requirement that $h(i)$ is a random element of $[m]$ means $\mathbf{Pr}[h(i) = y_i] = 1/m$. So for up to four arguments, a k-independent class behaves like the class of all functions from $[n]$ to $[m]$. Note that there are m^n functions in total and m^{n-k} functions mappings x_i to y_i for $1 \le i \le k$.

You will see the definition above also phrased as: $\mathcal{H}$ is k-independent if for any k distinct elements x_1 to x_k in $[n]$ and any k elements y_1 to y_k in $[m]$

$$\mathbf{Pr}_{h \in \mathcal{H}}\left[h(x_i) = y_i \text{ for } 1 \le i \le k\right] = \prod_{1 \le i \le k} \mathbf{Pr}_{h \in \mathcal{H}}\left[h(x_i) = y_i\right] = \frac{1}{m^k}.$$

For prime n, k-independent families exist.

Lemma 13.3. *Let n be a prime and let $\mathcal{H}$ be the set of polynomials of degree $k-1$. For $a = (a_0, \ldots, a_{k-1}) \in [n]^k$ the polynomial h_a defines the function*

$$x \mapsto \sum_{0 \le i \le k-1} a_i x^i \bmod n$$

from $[n]$ to $[n]$. The family $\mathcal{P}_k$ of polynomials of degree $k-1$ is k-independent.

Proof. The family has size n^k. There is exactly one polynomial that maps x_i to y_i for $1 \le i \le k$. □

k-independent hash functions share some properties with random functions, but also differ from them in important aspects. For a random function h, an integer ℓ, and a set $A \subseteq [m]$ of size a, we have $\mathbf{Pr}[h(i) \in A \text{ for } 0 \le i < \ell] = (a/m)^\ell$. For a k-independent function, this is only true for $\ell \le k$ (Lemma 13.4). We can nevertheless derive a non-trivial bound for the event "$h(i) \in A$ for $0 \le i < \ell$" provided $\ell \gg a/m \gg 1$ (Lemma 13.5).

Lemma 13.4. *Let $A \subseteq [m]$ and let $X_i = [h(i) \in A]$ be the indicator variable for the event $h(i) \in A$. The variables X_0 to X_{n-1} are k-independent. In particular, if $\ell \le k$, and $i_1, i_2, \ldots, i_\ell$ are distinct indices then $\mathbf{Pr}[h(i) \in A \text{ for } 1 \le i \le \ell] = (a/m)^\ell$.*

Proof. Let $a = |A|$ and let $A^1 = A$ and $A^0 = [m] \setminus A$. Then $\mathbf{Pr}[X_i = 1] = \mathbf{Pr}[h(i) \in A] = a/m$. For four distinct indices i_1 to i_4 and four values y_1 to y_4 in $\{0, 1\}$, we have

$$\begin{aligned}
\mathbf{Pr}[X_i = y_i \text{ for } 1 \le i \le 4] &= \sum_{a_i \in A^{y_i} \text{ for } 1 \le i \le 4} \mathbf{Pr}[h(i) = a_i \text{ for } 1 \le i \le 4] \\
&= \sum_{a_i \in A^{y_i} \text{ for } 1 \le i \le 4} \prod_{1 \le i \le 4} \mathbf{Pr}[h(i) = a_i] \\
&= \prod_{1 \le i \le 4} \sum_{a_i \in A^{y_i}} \mathbf{Pr}[h(i) = a_i] \\
&= \prod_{1 \le i \le 4} \mathbf{Pr}[X_i = y_i]\,.
\end{aligned}$$

The second equality holds due to k-independence. □

Lemma 13.5. *Let k be an even integer, let $A \subseteq [m]$ with $a = |A|$, and let h be a k-independent hash function. Let S be the number of i, $0 \le i < \ell$, such that $h(i) \in A$. Then $\mathbf{E}[S] = \ell a/m$ and*

$$\mathbf{Pr}[S = \ell] = O(\frac{\mathbf{E}[S] + \mathbf{E}[S]^{k/2}}{(\ell - \mathbf{E}[S])^k}).$$

In particular, if $\ell a/m > 1$ the probability is bounded by $O(1)/((m/a-1)^k \mathbf{E}[S]^{k/2})$.

Proof. Let $X_i = [h(i) \in A]$ be the indicator variable for the event $h(i) \in A$. The variables X_0 to $X_{\ell-1}$ are k-independent by Lemma 13.4. Let $S = \sum_i X_i$ be their sum. The expectation of S is $\ell a/m$ by linearity of expectations. If $h(i) \in A$ for $0 \le i < \ell$, then $S = \ell$ and hence $|S - \mathbf{E}[S]| \ge \ell - \ell a/m$. Thus

$$\begin{aligned}
\mathbf{Pr}[S = \ell] &\le \mathbf{Pr}[|S - \mathbf{E}[S]| \ge \ell - \ell a/m] \\
&\le \frac{\mathbf{E}\big[(S - \mathbf{E}[S])^k\big]}{(\ell - \ell a/m)^k} && \text{Markov inequality (Fact 13.10)} \\
&\le \frac{O(\mathbf{E}[S] + \mathbf{E}[S]^{k/2})}{(\ell - \ell a/m)^k} && \text{Fact 13.8} \\
&= O(\frac{\ell a/m + (\ell a/m)^{k/2}}{(\ell - \ell a/m)^k}).
\end{aligned}$$

Assume now that $\ell a/m > 1$ and let $\ell = (x+1)\ell a/m$. Then $x = m/a - 1$ and the bound becomes

$$\frac{O(\mathbf{E}[S]^{k/2})}{(x\mathbf{E}[S])^k} = \frac{O(1)}{x^k \mathbf{E}[S]^{k/2}}.$$

□

Note that the bounds are much weaker than for the case of random functions. For a good bound, we need $\ell \gg \ell a/m \gg 1$. For example, if $a = m/2$ and $\ell \ge 4$, the bound becomes $O(1)/(\ell/2)^{k/2}$.

3 Quicksort with Limited Independence

We use x_0 to x_{n-1} to denote our input. We assume the keys to be pairwise distinct.

Let $\mathcal{H}$ be a 4-independent class of hash function. We choose $h \in \mathcal{H}$ at random and generate the sequence $i_0 = h(0)$ to $i_{n-1} = h(n-1)$. In general, this is not a permutation. We split the input at x_{i_0} (= use x_{i_0} as a pivot), then the sequence containing x_{i_1} at x_{i_1} and so on. Since the sequence of split indices is in general not a permutation, we will be left with some unsorted sequences of more than one element. We sort any such sequence with a quadratic sorting algorithm, e.g., selection sort.

An alternative view of this process is that we sort by insertion into a binary tree. We start with the empty tree, insert x_{i_0}, then x_{i_1}, and so on. Repeated elements are skipped over. The remaining elements are inserted in any order.

Yet another view is: turn the sequence into a partial permutation by removing duplicates, more precisely, keep only the first occurrence of any number. Then add the missing elements in any order. Use this permutation to insert the elements into a binary tree.

Then analysis rests on the following Lemma.

Lemma 13.6. *Let A and B be disjoint subsets of $[n]$ with $|A| \le |B|$. Let*

$$C = \{i \in [n] \mid h(i) \in A,\ h(0), \ldots, h(i-1) \notin B\},$$

where h is chosen randomly from a 4-independent class. Then $\mathbf{E}[|C|] = O(1)$.

We postpone the proof of the Lemma and first show how to use it to bound the expected running time of randomized quicksort.

Theorem 13.7. *The expected running time of randomized quicksort with a 4-independent random hash function is $O(n \log n)$.*

Proof. We first assume that our input is sorted, i.e., $x_0 < x_1 < \ldots < x_n$. At the end of the proof, we will remove this assumption.

Comparisons are always between a pivot element and a non-pivot; the non-pivot may be the pivot in a later split. We charge the comparison to the non-pivot. In the tree-insertion view of the algorithms, comparisons are charged to the inserted element and not to the elements already in the tree.

Consider x_i. We want to bound the number of times x_i is compared with a pivot. We estimate the number of comparisons with a pivot in $\{x_{i+1}, \ldots, x_{n-1}\}$; the number of comparisons with pivots in $\{x_0, \ldots, x_{i-1}\}$ is analyzed analogously. We split this set into exponentially increasing subsets, i.e., for $0 \le \ell \le \log n$, let

$$P_\ell = \{x_{i+j} \mid 2^\ell < j \le 2^{\ell+1} \text{ and } i+j \in [n]\}.$$

Then $P_0 = \{i+2\}$, $P_1 = \{i+3, i+4\}$, and so on. Note that $i+1$ is not contained in any of the sets. If element x_i is compared with a pivot $i+j \in P_\ell$ then $i+j$ appears in the sequence of indices before any index in $\{i+1, \ldots, i+2^\ell\}$. Thus, we can apply Lemma 13.6 with $A = P_\ell$ and $B = \{i+1, \ldots, i+2^\ell\}$ to conclude that the expected number of pivots in P_ℓ with which x_i is compared is $O(1)$. Thus the total expected number of pivots with which x_i is compared is bounded by

$$1 + \sum_{0 \le \ell \le \log n} O(1) = O(\log n),$$

where the first 1 counts the comparison with the pivot x_{i+1} (this comparison may or may not happen).

We come to the clean-up phase. For simplicity, we assume that n is a power of two. The clean-up phase deals with subsequences not containing any pivot. Let us call a sequence *special* if it is form $\{x_{k\cdot 2^\ell}, \ldots, x_{(k+1)2^\ell - 1}\}$ for some $\ell < \log n$ and some k. Every subsequence contains a special subsequence of at least 1/4th its length. The expected cost of the clean-up phase is therefore bounded by

$$16 \cdot \sum_{S \text{ is special}} \mathbf{Pr}[\text{no pivot in } S] \cdot |S|^2$$

Consider a special set S of size 2^ℓ. For $i \in [n]$ consider the event $[h(i) \in S]$ and let $X = \sum_{i \in [n]} [h(i) \in S]$. Then $\mathbf{Pr}[h(i) \in S] = 2^\ell/n$ and hence $\mathbf{E}[X] = 2^\ell$. Thus

$$\mathbf{Pr}[\text{no pivot in } S] \le \mathbf{Pr}[|X - \mathbf{E}[X]| \ge \mathbf{E}[X]] \le \frac{\mathbf{E}\big[(X - \mathbf{E}[X])^2\big]}{\mathbf{E}[X]^2} = \frac{O(\mathbf{E}[X])}{\mathbf{E}[X]^2} = O(1) \cdot 2^{-\ell},$$

and hence

$$\sum_{S \text{ is special}} \mathbf{Pr}[\text{no pivot in } S] \cdot |S|^2 \le \sum_{0 \le \ell \le \log n} \frac{n}{2^\ell} O(1) \cdot 2^{-\ell} 2^{2\ell} = O(n \log n).$$

We still need to extend the argument to non-sorted inputs. Let $x_0, \ldots, x_{n-1}$ be arbitrary and let π be its order type, i.e., $x_{\pi(1)} < x_{\pi(2)} < \ldots$. In the preceding paragraphs, replace any indexed occurrence x_h of x by $x_{\pi(h)}$. So the first two sentences read. Consider $x_{\pi(i)}$. We want to bound the number of times $x_{\pi(i)}$ is compared with a pivot. It suffices to estimate the number of comparisons with a pivot in $\{x_{\pi(i+1)}, \ldots, x_{\pi(n-1)}\}$. □

4 Useful Facts

Fact 13.8. *Let k be an even positive integer and let X_1 to X_n be k-independent random variables with values in $[0,1]$. Let S be their sum. Then*

$$\mathbf{E}\big[(S - \mathbf{E}[S])^k\big] = O(\mathbf{E}[S] + \mathbf{E}[S]^{k/2}).$$

I give a proof in the Appendix. We use the bound for $k = 4$.

Fact 13.9. *Let $k \ge 2$ and r_0 be positive integers. Then*

$$\sum_{i \ge 1} \frac{1}{(r_0 + i)^k} = O(\frac{1}{r_0^{k-1}}).$$

Proof. We estimate the sum by the corresponding integral.

$$\sum_{i\geq 1}\frac{1}{(r_0+i)^k}\leq\int_{x\geq r_0}\frac{1}{x^k}=-\frac{x^{-k+1}}{k-1}\Big|_{r_0}^{\infty}=\frac{r_0^{-k+1}}{k-1}=O(\frac{1}{r_0^{k-1}}).$$

□

Fact 13.10 (Markov's inequality)**.** *Let k be an even integer and let Z be any random variable.*

$$\mathbf{Pr}[|Z-\mathbf{E}[Z]|\geq a]\leq\frac{\mathbf{E}\Big[(Z-\mathbf{E}[Z])^k\Big]}{a^k}.$$

Proof. Observe that

$$\mathbf{E}\Big[(Z-\mathbf{E}[Z])^k\Big]\geq a^k\mathbf{Pr}[|Z-\mathbf{E}[Z]|\geq a]\,.$$

□

5 Proof of Lemma 13.6

For an event E, we use $[E]$ to denote the indicator variable of E. It is one if E occurs and zero otherwise.

We may assume $|A|=|B|$. Otherwise, drop some elements from B. Let $m=|A|=|B|$ and let $\alpha=m/n$. For simplicity, we assume that n is a power of two.

As in the preceding proof, we consider exponentially increasing sets. For any non-negative integer ℓ, let

$$X_\ell=\{j\in[n]\mid j<2^\ell\text{ and }h(j)\in A\}\setminus(X_0\cup\ldots\cup X_{\ell-1}),$$

and let E_ℓ be the event that $h(i)\notin B$ for $i<2^\ell$. Note that $X_0\subseteq\{0\}$ and $X_\ell\subseteq\{2^{\ell-1},\ldots,2^\ell-1\}$. Consider any $j\in C$. Then $j\in X_\ell$ for some ℓ; also $E_{\ell-1}$ occurs. Thus

$$\mathbf{E}[C]\leq\sum_{0\leq\ell\leq\log n}\mathbf{E}[|X_\ell|\cdot[E_{\ell-1}]]\,. \tag{1}$$

We need to estimate the expectation of a product of random variables. Since the random variables are not independent, this is a challenge. We start by investigating the factors. Clearly,

$$\mathbf{E}[|X_\ell|]=\begin{cases}\alpha & \text{if }\ell=0\\ \alpha 2^{\ell-1} & \text{if }\ell\geq 1.\end{cases}$$

$|X_\ell|$ is a sum of $2^{\ell-1}$ random variables $[h(j)\in A]$ for $\ell\geq 1$.

We can now already bound (1) for small ℓ, namely

$$\sum_{\ell;\ \alpha 2^\ell\leq 1}\mathbf{E}[|X_\ell|\cdot[E_{\ell-1}]]\leq\sum_{\ell;\ \alpha 2^\ell\leq 1}\mathbf{E}[|X_\ell|]\leq\sum_{\ell;\ \alpha 2^\ell\leq 1}\alpha 2^\ell=O(1).$$

We next concentrate on the ℓ with $\alpha 2^\ell>1$. Here, we use the following strategy which is also useful in other contexts. We write

$$\begin{aligned}\sum_{\ell;\ \alpha 2^\ell>1}\mathbf{E}[|X_\ell|\cdot[E_{\ell-1}]]&=\sum_{\ell;\ \alpha 2^\ell>1}\mathbf{E}[(|X_\ell|-2\mathbf{E}[|X_\ell|]+2\mathbf{E}[|X_\ell|])\cdot[E_{\ell-1}]]\\&\leq\sum_{\ell;\ \alpha 2^\ell>1}\mathbf{E}[\max(0,|X_\ell|-2\mathbf{E}[|X_\ell|])]+\sum_{\ell;\ \alpha 2^\ell>1}2\mathbf{E}[|X_\ell|]\cdot\mathbf{E}[[E_{\ell-1}]]\\&=\sum_{\ell;\ \alpha 2^\ell>1}\mathbf{E}[\max(0,|X_\ell|-2\mathbf{E}[|X_\ell|])]+\sum_{\ell;\ \alpha 2^\ell>1}\alpha 2^\ell\cdot\mathbf{E}[[E_{\ell-1}]]\,.\end{aligned}$$

Recall that the expectation of $|X_\ell|$ is $\alpha 2^{\ell-1}$. For the second term, we will show that $\mathbf{Pr}[E_{\ell-1}]$ is sufficiently small to cancel the factor $\alpha 2^\ell$.

Let r be a non-negative integer. Then

$$\mathbf{Pr}\big[|X_\ell|\geq\alpha 2^\ell+r\big]\leq\frac{\mathbf{E}\big[(|X_\ell|-\mathbf{E}[|X_\ell|])^4\big]}{(\alpha 2^{\ell-1}+r)^4}\leq\frac{O(\mathbf{E}[|X_\ell|]+\mathbf{E}[|X_\ell|]^2)}{(\alpha 2^{\ell-1}+r)^4}\leq\frac{O(1)}{(\alpha 2^{\ell-1}+r)^2},$$

and hence

$$\begin{aligned}\sum_{\ell;\ \alpha 2^\ell>1} \mathbf{E}\big[\max(0,|X_\ell|-\alpha 2^\ell)\big] &\le \sum_{\ell;\ \alpha 2^\ell>1}\sum_{r\ge 0}\mathbf{Pr}\big[|X_\ell|\ge \alpha 2^\ell + r\big]\\ &= \sum_{\ell;\ \alpha 2^\ell>1}\sum_{r\ge 0}\frac{O(1)}{(\alpha 2^{\ell-1}+r)^2}\\ &\le O(1)\cdot\sum_{\ell;\ \alpha 2^\ell>1}\frac{1}{\alpha 2^{\ell-1}}\\ &= O(1).\end{aligned}$$

The event E_ℓ occurs if the first 2^ℓ hash values do not belong to B. Let $Z_i = [h(i)\in B]$ and $Z = \sum_{0\le i<2^\ell} Z_i$. Then $\mathbf{E}[Z_i] = m/n$ and hence $\mathbf{E}[Z] = \alpha 2^\ell$. Also E_ℓ is equivalent to $Z=0$, and we have

$$\begin{aligned}\mathbf{Pr}[E_\ell] = \mathbf{Pr}[Z=0] \le \mathbf{Pr}[|Z-\mathbf{E}[Z]|\ge \mathbf{E}[Z]] &\le \frac{\mathbf{E}\Big[(Z-\mathbf{E}[Z])^4\Big]}{\mathbf{E}[Z]^4}\\ &= \frac{O(\mathbf{E}[Z]+\mathbf{E}[Z]^2)}{\mathbf{E}[Z]^4} = \frac{O(1)}{\mathbf{E}[Z]^2} = \frac{O(1)}{\alpha^2 2^{2\ell}}.\end{aligned}$$

Thus

$$\sum_{\ell;\ \alpha 2^\ell>1}\mathbf{E}\big[\alpha 2^\ell\cdot[E_{\ell-1}]\big] \le O(1)\cdot\sum_{\ell;\ \alpha 2^\ell>1}\frac{1}{\alpha 2^\ell} = O(1).$$

This completes the proof of Lemma 13.6.

6 2-Independence

The analysis above can also be carried out with 2-independence. What changes?

The clean-up phase was analyzed with 2-independence. Nothing changes.

Small ℓ, i.e., ℓ with $\alpha 2^\ell \le 1$. Nothing changes.

Large ℓ, i.e., ℓ with $\alpha 2^\ell > 1$.

Let r be a non-negative integer. Then

$$\mathbf{Pr}\big[|X_\ell|\ge \alpha 2^\ell + r\big] \le \frac{\mathbf{E}\big[(|X_\ell|-\mathbf{E}[|X_\ell|])^2\big]}{(\alpha 2^{\ell-1}+r)^2} \le \frac{O(\mathbf{E}[|X_\ell|]+\mathbf{E}[|X_\ell|])}{(\alpha 2^{\ell-1}+r)^2} \le \frac{O(1)}{(\alpha 2^{\ell-1}+r)^2},$$

and hence

$$\begin{aligned}\sum_{\ell;\ \alpha 2^\ell>1} \mathbf{E}\big[\max(0,|X_\ell|-\alpha 2^\ell)\big] &\le \sum_{\ell;\ \alpha 2^\ell>1}\sum_{r\ge 0}\mathbf{Pr}\big[|X_\ell|\ge \alpha 2^\ell + r\big]\\ &= \sum_{\ell;\ \alpha 2^\ell>1}\sum_{r\ge 0}\frac{O(1)\alpha 2^{\ell-1}}{(\alpha 2^{\ell-1}+r)^2}\\ &\le O(1)\cdot\sum_{\ell;\ \alpha 2^\ell>1}\frac{\alpha 2^{\ell-1}}{\alpha 2^{\ell-1}}\\ &= O(\log n).\end{aligned}$$

The event E_ℓ occurs if the first 2^ℓ hash values do not belong to B. Let $Z_i = [h(i)\in B]$ and $Z = \sum_{0\le i<2^\ell} Z_i$. Then $\mathbf{E}[Z_i] = m/n$ and hence $\mathbf{E}[Z] = \alpha 2^\ell$. Also E_ℓ is equivalent to $Z=0$, and we have

$$\begin{aligned}\mathbf{Pr}[E_\ell] = \mathbf{Pr}[Z=0] \le \mathbf{Pr}[|Z-\mathbf{E}[Z]|\ge \mathbf{E}[Z]] &\le \frac{\mathbf{E}\Big[(Z-\mathbf{E}[Z])^2\Big]}{\mathbf{E}[Z]^2}\\ &= \frac{O(\mathbf{E}[Z]+\mathbf{E}[Z])}{\mathbf{E}[Z]^2} = \frac{O(1)}{\mathbf{E}[Z]} = \frac{O(1)}{\alpha 2^\ell}.\end{aligned}$$

Thus

$$\sum_{\ell;\ \alpha 2^\ell > 1} \mathbf{E}\big[\alpha 2^\ell \cdot [E_{\ell-1}]\big] \leq O(1) \cdot \sum_{\ell;\ \alpha 2^\ell > 1} O(1) = O(\log n).$$

Theorem 13.11. *With 2-independence, the expected running time is $O(n \log^2 n)$.*

7 A Sketch of a Research Immersion Lab/Master Thesis: Randomized Algorithms with Limited Randomness

The goal is to investigate randomized algorithms, in particular quicksort and randomized incremental constructions under limited randomness.

7.1 Experimental Work: Quicksort

We implement standard randomized quicksort and the randomized quicksort of today's lecture. We do a statistic for different number of keys and different degrees of independence: number of comparisons, expectation, variance, higher moments. In particular, we would like to answer the following questions:

1. Is 2-independence enough? Is 3-independence enough? The best proven upper bounds are $O(n(\log n)^2)$. After dividing out n, one should see the difference to $\log n$.
2. What improves, if one goes to higher independence? The expectation? The variance?

7.2 Experimental Work: Randomized Incremental Constructions

We perform similar experiments for randomized incremental constructions. Our first experiments are for convex hulls in 2- and higher dimensions. Note that no proven bounds are known. We ask the same questions as above.

7.3 Theoretical Work

Constant Factors: We redo the analysis of Knudsen and Stöckel, but pay attention to constant factors.

The Different Views on Quicksort There are (at least) two views on randomized quicksort.

- Permute the input randomly and then run deterministic quicksort. Or insert the elements into a binary tree.
- For each subproblem choose the pivot uniformly at random from the subproblem.

In what sense are the two views isomorphic?

In order to choose a random permutation, we need $\log n! = \Theta(n \log n)$ bits.

Let $B(n)$ be the number of bits required for standard quicksort. Then $B(1) = 0$ and

$$B(n) = \log n + \frac{1}{n} \sum_{1 \leq i \leq n} (B(i-1) + B(n-i)) = \log n + \frac{1}{n} + 2 \cdot \sum_{2 \leq i \leq n-1} B(i).$$

Thus $nB(n) = n \log n + 2 \cdot \sum_{2 \leq i \leq n-1} B(i)$ or

$$nB(n) - (n-1)B(n-1) = n \log n - (n-1) \log(n-1) + 2B(n-1)$$

or

$$nB(n) = (n+1)B(n-1) + n \log n - (n-1) \log(n-1).$$

I conjecture $B(n) = O(n)$, i.e., the standard view of quicksort requires fewer random bits. Consider the special case, that we split perfectly in every step. Let $n = 2^k$. Then

$$B(n) = k + 2B(n-1) = k + 2(k-1) + 4(k-2) + \ldots + 2^k(k-k) = \sum_{0 \leq i \leq k} i 2^{k-i} = 2^k \cdot \sum_i i 2^{-i} = O(2^k).$$

Standard quicksort generates a random tree. If $n = 0$, the tree is empty, and if $n = 1$, the tree consists of a single node. If $n \geq 2$, the tree consists of a root and a random left subtree of size $i - 1$ and random right subtree of size $n - i$, where i is chosen uniformly from 1 to n. What does it mean for this process to be k-independent?

2-Independence

Analysis of RICs under Limited Independence Does the analysis by Knudsen and Stöckel generalize to RICs? Let us first try convex hulls in 2d.

k-independent Permutations I found only one reference on the subject. On Permutations with Limited Independence by TOSHIYA ITOH, YOSHINORI TAKEI and JUN TARUI.

The simplest problem in randomized analysis is the number of left-to-right maxima in a permutation. What does it mean to do this analysis under limited independence?

In the Knudsen/Stöckel paper, they analyze the number of comparisons between i and pivot elements. Consider pivots $j > i$. Pivots are elements j that are chosen before any element in $i+1$ to $j-1$ are chosen. In some sense, these are right-to-left minima right of i.

8 Appendix: Proof of Fact 13.8

The paper states Fact 13.8 without a proof and without giving a reference. I tried to prove it myself for almost a full day, but could not do it. Then I started asking around. My former postdocs Yi Li pointed me to the lecture notes of Jelani Nelson (`minilek@seas.harvard.edu`) for the Modalgo summer school on streaming algorithms. It still took me a day to work out the details.

Thomas proofread the appendix and observed that the case $k = 4$ has a simple proof.

8.1 The Complex Proof

For a vector $x \in \mathbb{R}^n$, $\|x\|_2 = \left(\sum_i x_i^2\right)^{1/2}$ denotes the 2-norm.

Fact 13.12 (Khintchine Inequality)**.** *Let k be an even integer and let $x \in \mathbb{R}^n$. Let r_1 to r_n be k-independent rademachers (random variables with $r_i = \pm 1$ with probability $1/2$ each). Then*[1]

$$\mathbf{E}\left[(\sum_i r_i x_i)^k\right] \leq c_k \cdot \left(\sum_i x_i^2\right)^{k/2} = c_k \cdot \|x\|_2^k,$$

where c_k is a constant only depending on k.

Proof. This proof uses some concepts which we have not introduced in class. Let g_i be an gaussian with mean zero and variance 1. Expand $\mathbf{E}\left[(\sum_i r_i x_i)^k\right]$ and $\mathbf{E}\left[(\sum_i g_i x_i)^k\right]$ into a sum of expectations of monomials. For a monomial in which one of the x_i has an odd exponent, the expectation is zero (for the rademachers[2] and the gaussians). For a monomial in which all exponents are even, the rademacher monomial is dominated by the corresponding gaussian monomial; k-independence suffices for this argument as all monomials have degree at most k. Thus

$$\mathbf{E}\left[(\sum_i r_i x_i)^k\right] \leq \mathbf{E}\left[(\sum_i g_i x_i)^k\right].$$

A weighted sum of gaussians is again a gaussian; in particular $\sum_i g_i x_i$ is a gaussian with mean zero and variance $\sum_i x_i^2$. Thus

$$\mathbf{E}\left[(\sum_i g_i x_i)^k\right] \leq \frac{k!}{2^{k/2}(k/2)!} \cdot \left(\sum_i x_i^2\right)^{k/2}.$$

□

Lemma 13.13 (Symmetrization)**.** *Let k be an even integer, let X_1 to X_n be k-independent random variables, and let r_1 to r_n be k-independent rademachers. Then*

$$\mathbf{E}\left[\left(\sum_i (X_i - \mathbf{E}[X_i])\right)^k\right] \leq 2^k \mathbf{E}\left[(\sum_i r_i X_i)^k\right].$$

[1] A naive bound would simply use $\sum_i r_i x_i \leq \sum_i |x_i|$, which is the 1-norm of x. Note that the 2-norm is usually smaller than the 1-norm. For example, if $x_i = 1$ for all i, then the 1-norm is n and the 2-norm is $\sqrt{n}$.

[2] Let p be an odd integer. Then $\mathbf{E}[(r_i x_i)^p] = x_i^p \mathbf{E}\left[r_i^p\right] = 0$.

Proof. Let Y_1 to Y_n be copies of the variables X_1 to X_n, i.e., identically distributed, but independent. Let r_1 to r_n be independent rademachers. Then (we are indicating by subscripts with respect to which variables we are taking expectations)

$$\begin{aligned}
\mathbf{E}\big[(S-\mathbf{E}[S])^k\big] &= \mathbf{E}_X\left[\left((\sum_i X_i) - \mathbf{E}_Y\left[\sum_i Y_i\right]\right)^k\right] && \text{replacing } \mathbf{E}\left[\sum_i X_i\right] \text{ by } \mathbf{E}\left[\sum_i Y_i\right]\\
&= \mathbf{E}_X\left[\left(\mathbf{E}_Y\left[(\sum_i X_i) - (\sum_i Y_i)\right]\right)^k\right] && \text{moving } \sum_i X_i \text{ into } \mathbf{E}_Y[.]\\
&= \mathbf{E}_X\left[\left(\mathbf{E}_Y\left[\sum_i (X_i - Y_i)\right]\right)^k\right] && \text{rearranging the inner sum}\\
&\le \mathbf{E}_X\left[\mathbf{E}_Y\left[\left(\sum_i (X_i - Y_i)\right)^k\right]\right] && \text{Jensen's inequality}\\
&= \mathbf{E}_{X,Y}\left[\left(\sum_i X_i - Y_i\right)^k\right] && \text{writing the two expectations as one}\\
&= \mathbf{E}_{X,Y,r}\left[\left(\sum_i r_i(X_i - Y_i)\right)^k\right] && (*)\\
&= \mathbf{E}_{X,Y,r}\left[\left(\sum_i r_i X_i - \sum_i r_i Y_i\right)^k\right] && \text{rearranging the inner sum}\\
&\le \mathbf{E}_{X,Y,r}\left[\left(\left|\sum_i r_i X_i\right| + \left|\sum_i r_i Y_i\right|\right)^k\right] && \text{since } x-y \le |x|+|y|\\
&= \mathbf{E}_r\left[\mathbf{E}_{X,Y}\left[\left(\left|\sum_i r_i X_i\right| + \left|\sum_i r_i Y_i\right|\right)^k\right]\right] && \text{writing one expectation as two}\\
&\le \mathbf{E}_r\left[\mathbf{E}_X\left[\left(2\sum_i r_i X_i\right)^k\right]\right] && (**)\\
&= 2^k \mathbf{E}_{X,r}\left[\left(\sum_i r_i X_i\right)^k\right] && \text{pulling out the 2.}
\end{aligned}$$

Jensen's inequality states $f(\mathbf{E}[X]) \le \mathbf{E}[f(X)]$ for a convex function f; we apply it to the function $x \mapsto x^k$. $(*)$ holds since the $X_i - Y_i$ are symmetric and independent across i; therefore $\sum_i (X_i - Y_i)$ is distributed as $\sum_i r_i(X_i - Y_i)$. k-independence suffices since expanding a k-th power into monomials generates only monomials of degree k. $(**)$ holds since k is even and the Y's are an independent copy of the X's. □

Lemma 13.14. *Let X_1 to X_n be k-independent random variables and let S be their sum. Then for any even positive integer k*

$$\mathbf{E}\left[(\sum_i (X_i - \mathbf{E}[X_i])^k\right] = d_k \mathbf{E}\left[\left(\sum_i X_i^2\right)^{k/2}\right] = d_k \mathbf{E}\big[\|X\|_2^k\big]$$

for some constant d_k.

Proof. We first apply symmetrization and then Khintchine:

$$\mathbf{E}\left[\left(\sum_i (X_i - \mathbf{E}[X_i]\right)^k\right] \le 2^k \mathbf{E}_{X,r}\left[\left(\sum_i r_i X_i\right)^k\right] \qquad \text{symmetrization}$$

$$\le 2^k c_k \mathbf{E}_X\left[\left(\sum_i X_i^2\right)^{k/2}\right] \qquad \text{Khintchine}$$

□

Fact 13.15. *Let p be a positive integer. Then*

$$a^p \le 2^{p-1}\left((a-y)^p + y^p\right) \quad \textit{for all non-negative real } a \textit{ and } y.$$

Proof. The claim is obvious for $p = 1$. So assume $p \ge 2$. Consider $f(y) = 2^{p-1}\left((a-y)^p + y^p\right) - a^p$. Then

$$f'(y) = 2^{p-1}p\left(y^{p-1} - (a-y)^{p-1}\right) \quad \text{and} \quad f''(y) = 2^{p-1}p(p-1)\left(y^{p-2} + (a-y)^{p-2}\right).$$

and hence $f'(y) = 0$ if and only if $y = a/2$. Also, $f(a/2) = 0$ and $f''(y) \ge 0$ for all y. The latter is obvious for even p, and for odd p and $a \ge y \ge 0$. If p is odd and $y \ge a \ge 0$, $y^{p-2} + (a-y)^{p-2} = y^{p-2} - (y-a)^{p-2} \ge 0$. □

Fact 13.16. *Let p be a positive odd integer. Then $x^{p+1} + x^{p-1} \ge 2x^p$ for all real x.*

Proof. $x^{p+1} + x^{p-1} - 2x^p = x^{p-1}(x-1)^2 \ge 0$ for all x. □

We are now ready to prove Fact 13.8.

Fact 13.17. *Let k be an even positive integer and let X_1 to X_n be k-independent random variables with values in $[0,1]$, and let S be their sum. Then*

$$\mathbf{E}\left[(\sum_i (X_i - \mathbf{E}[X_i])^k\right] \le c_k \cdot \left(\mathbf{E}[S] + \mathbf{E}[S]^{k/2}\right)$$

for some constant c_k.

Proof. We use induction on k. Let $k = 2p$. Using Lemma 13.14 and Fact 13.15, we obtain

$$\mathbf{E}\left[\left(\sum_i (X_i - \mathbf{E}[X_i])\right)^k\right] \le d_k \mathbf{E}\left[\|X\|_2^k\right] \quad \text{by Lemma 13.14}$$

$$= d_k \mathbf{E}\left[\left(\sum_i (X_i^2 - \mathbf{E}\left[X_i^2\right]) + \sum_i \mathbf{E}\left[X_i^2\right]\right)^{k/2}\right] \quad \text{rewriting the sum}$$

$$\le d_k 2^{k/2-1}\left(\mathbf{E}\left[\left(\sum_i (X_i^2 - \mathbf{E}\left[X_i^2\right])\right)^{k/2}\right] + \mathbf{E}\left[\left(\sum_i \mathbf{E}\left[X_i^2\right]\right)^{k/2}\right]\right)$$

where the last inequality follows from Fact 13.15. We next bound the two terms on the right. For the second term, observe that $X_i \in [0,1]$ implies $X_i^2 \le X_i$ and that the outer expectation is the expectation of a scalar and hence can be dropped. Thus

$$\mathbf{E}\left[\left(\sum_i \mathbf{E}\left[X_i^2\right]\right)^{k/2}\right] \le \left(\sum_i \mathbf{E}[X_i]\right)^{k/2} = \mathbf{E}[S]^{k/2}.$$

We turn to the first term.

Let $Y_i = X_i^2$ and $T = \sum_i Y_i$ and write the first term as $\mathbf{E}\left[\left(\sum_i (Y_i - \mathbf{E}[Y_i])\right)^{k/2}\right]$. If $k/2 = 1$, the term is zero. If $k/2$ is even, by the induction hypothesis, the term is bounded by by $c_{k/2}(\mathbf{E}[T] + \mathbf{E}[T]^{k/4}) \le$

$2c_{k/2}(\mathbf{E}[S] + \mathbf{E}[T]^{k/2})$, where the last inequality uses $\mathbf{E}[T]^{k/4} \leq \max(\mathbf{E}[S], \mathbf{E}[S]^{k/2})$. If $k/2$ is odd, Fact 13.16 with $x = \sum_i (X_i^2 - \mathbf{E}[X_i^2])$ yields

$$\begin{aligned}\mathbf{E}\left[\left(\sum_i (X_i^2 - \mathbf{E}[X_i^2])\right)^{k/2}\right] &\leq \frac{1}{2}\left(\mathbf{E}\left[\left(\sum_i (X_i^2 - \mathbf{E}[X_i^2])\right)^{k/2+1}\right] + \mathbf{E}\left[\left(\sum_i (X_i^2 - \mathbf{E}[X_i^2])\right)^{k/2-1}\right]\right) \\ &\leq \frac{1}{2}\left(c_{k/2+1}(\mathbf{E}[T] + \mathbf{E}[T]^{(k/2+1)/2}) + c_{k/2-1}(\mathbf{E}[T] + \mathbf{E}[T]^{(k/2-1)/2})\right) \\ &\leq (c_{k/2+1} + c_{k/2-1})(\mathbf{E}[S] + \mathbf{E}[S]^{k/2}).\end{aligned}$$

Thus

$$\mathbf{E}\left[\left(\sum_i (X_i - \mathbf{E}[X_i])\right)^k\right] \leq c_k\left(\mathbf{E}[S] + \mathbf{E}[S]^{k/2}\right)$$

for some constant c_k. □

8.2 The Case $k = 4$

Fact 13.18. *Let X_1 to X_n be 4-independent random variables with values in $[0, 1]$, and let S be their sum. Then*

$$\mathbf{E}\left[(\sum_i (X_i - \mathbf{E}[X_i])^4\right] \leq c \cdot \left(\mathbf{E}[S] + \mathbf{E}[S]^2\right)$$

for some constant c.

Proof. Let $Y_i = X_i - \mathbf{E}[X_i]$. Then $\mathbf{E}[Y_i] = 0$ and hence

$$\mathbf{E}\left[(\sum_i Y_i)^4\right] = \sum_m \mathbf{E}\left[\prod_{i\in[n]} Y_i^{m(i)}\right] = \sum_m \prod_{i\in[n]} \mathbf{E}\left[Y_i^{m(i)}\right],$$

where the summation is over all non-negative vectors $m = (m_0, \ldots, m_{n-1})$ with $\sum_i m(i) = 4$. The second equality follows from 4-independence. If an exponent $m(i)$ is equal to one, $\mathbf{E}\left[Y_i^{m(i)}\right] = \mathbf{E}[Y_i] = 0$. Thus the only relevant terms are $Y_i^2 Y_j^2$ with $i \neq j$ and Y_i^4. Thus

$$\mathbf{E}\left[(\sum_i Y_i)^4\right] = \sum_{i\neq j} \mathbf{E}[Y_i^2]\,\mathbf{E}[Y_j^2] + \sum_i \mathbf{E}[Y_i^4] \leq \frac{1}{2}\left(\sum_i \mathbf{E}[Y_i^2]\right)^2 + \sum_i \mathbf{E}[Y_i^4].$$

Next observe

$$\mathbf{E}[Y_i^2] = \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 \leq \mathbf{E}[X_i^2] \leq \mathbf{E}[X_i]$$

and

$$\begin{aligned}\mathbf{E}[Y_i^4] &= \mathbf{E}[X_i^4] - 4\mathbf{E}[X_i^3]\,\mathbf{E}[X_i] + 6\mathbf{E}[X_i^2]\,\mathbf{E}[X_i]^2 - 4\mathbf{E}[X_i]\,\mathbf{E}[X_i]^3 + \mathbf{E}[X_i]^4 \\ &\leq \mathbf{E}[X_i^4] + 6\mathbf{E}[X_i^2]\,\mathbf{E}[X_i]^2 + \mathbf{E}[X_i]^4 \\ &\leq 8\mathbf{E}[X_i]\end{aligned}$$

since $X_i^4 \leq X_i^2 \leq X_i$ and $\mathbf{E}[X_i] \leq 1$. Thus

$$\mathbf{E}\left[(\sum_i Y_i)^4\right] \leq \frac{1}{2}\mathbf{E}[X_i]^2 + 8\mathbf{E}[X_i].$$

□

8.3 The Case $k = 6$

Fact 13.19. *Let X_1 to X_n be 6-independent random variables with values in $[0,1]$, and let S be their sum. Then*

$$\mathbf{E}\left[(\sum_i (X_i - \mathbf{E}[X_i])^6 \right] \leq c \cdot \left(\mathbf{E}[S] + \mathbf{E}[S]^3 \right)$$

for some constant c.

Proof. Let $Y_i = X_i - \mathbf{E}[X_i]$. Then $\mathbf{E}[Y_i] = 0$ and hence

$$\mathbf{E}\left[(\sum_i Y_i)^4 \right] = \sum_m \mathbf{E}\left[\prod_{i \in [n]} Y_i^{m(i)} \right] = \sum_m \prod_{i \in [n]} \mathbf{E}\left[Y_i^{m(i)} \right],$$

where the summation is over all non-negative vectors $m = (m_0, \ldots, m_{n-1})$ with $\sum_i m(i) = 6$. The second equality follows from 4-independence. If an exponent $m(i)$ is equal to one, $\mathbf{E}\left[Y_i^{m(i)} \right] = \mathbf{E}[Y_i] = 0$. Thus the only relevant terms are $Y_i^2 Y_j^2 Y_k$ with three distinct indices i, j, k, $Y_i^2 Y_j^4$ and $Y_i^3 Y_j^3$ with $i \neq j$ and Y_i^4. Thus

$$\mathbf{E}\left[(\sum_i Y_i)^6 \right] \leq (\sum_i \mathbf{E}\left[Y_i^2 \right])^3 + (\sum_i \mathbf{E}\left[Y_i^2 \right]) \cdot (\sum_i \mathbf{E}\left[Y_i^4 \right]) + (\sum_i \mathbf{E}\left[Y_i^3 \right])^2 + \sum_i \mathbf{E}\left[Y_i^6 \right].$$

Next observe

$$\mathbf{E}\left[Y_i^2 \right] = \mathbf{E}\left[X_i^2 \right] - \mathbf{E}[X_i]^2 \leq \mathbf{E}\left[X_i^2 \right] \leq \mathbf{E}[X_i]$$

and

$$\begin{aligned} \mathbf{E}\left[Y_i^4 \right] &= \mathbf{E}\left[X_i^4 \right] - 3\mathbf{E}\left[X_i^3 \right] \mathbf{E}[X_i] + 6\mathbf{E}\left[X_i^2 \right] \mathbf{E}[X_i]^2 - 3\mathbf{E}[X_i] \mathbf{E}[X_i]^3 + \mathbf{E}[X_i]^4 \\ &\leq \mathbf{E}\left[X_i^4 \right] + 6\mathbf{E}\left[X_i^2 \right] \mathbf{E}[X_i]^2 + \mathbf{E}[X_i]^4 \\ &\leq 8\mathbf{E}[X_i] \end{aligned}$$

since $X_i^4 \leq X_i^2 \leq X_i$ and $\mathbf{E}[X_i] \leq 1$. Thus ... □

This proof strategy should work for all even k. We first expand $\mathbf{E}\left[(\sum_i Y_i)^k \right]$ into a sum of monomials where all exponents are even. Then we collect all monomials of the same "type" and simplify. Then we use

$$\mathbf{E}[Y_i^p] = \mathbf{E}[(X_i - \mathbf{E}[X_i])^p] \leq \mathbf{E}[X_i^p] \leq ExX_i$$

for odd p and

$$\mathbf{E}\left[Y_i^{2p} \right] = \mathbf{E}\left[(X_i - \mathbf{E}[X_i])^{2p} \right] \leq \sum_{0 \leq j \leq 2p} \binom{2p}{j} \mathbf{E}\left[X_i^j \right] \mathbf{E}\left[X_i^{n-j} \right] \leq \left(\sum_{0 \leq j \leq 2p} \binom{2p}{j} \right) \cdot \mathbf{E}[X_i]$$

for even $2p$. Finally, we use $\mathbf{E}[S]^j \leq \max(\mathbf{E}[S], \mathbf{E}[S]^{k/2})$ for $1 \leq j \leq k/2$.