— 5 —

Chebyshev, Sampling and Selection

During a native rebellion in German East Africa, the Imperial Ministry in Berlin issued the following order to its representatives on the ground: "The natives are to be instructed that on pain of harsh penalties, every rebellion must be announced, in writing, six weeks before it breaks out."

Dead Funny: Humor in Hitler's Germany, Rudolph Herzog

5.1. Chebyshev's inequality

Here, we discuss Chebyshev's inequality, sometimes written as Chebychev's inequality using the French spelling of Chebyshev[5.1]

5.1.1. Example: A better inequality via moments

Let $X_1, \ldots, X_n \in \{-1, +1\}$ be mutually independent random variables, each taking value in $\{-1, +1\}$ with equal probability $1/2$. Let $Y = \sum_{i=1}^{n} X_i$. We have that

$$\mathbb{E}[Y] \;=\; \mathbb{E}\Big[\sum_i X_i\Big] \;=\; \sum_i \mathbb{E}[X_i] \;=\; n \cdot 0 \;=\; 0.$$

A more interesting quantity is

$$\begin{aligned}\mathbb{E}[Y^2] \;&=\; \mathbb{E}\Big[\Big(\sum_i X_i\Big)^2\Big] \;=\; \mathbb{E}\Big[\sum_i X_i^2 + 2\sum_{i<j} X_i X_j\Big] \\ &=\; \sum_i \mathbb{E}[X_i^2] + 2\,\mathbb{E}\Big[\sum_{i<j} X_i X_j\Big]\end{aligned}$$

[5.1] *Pafnuty Lvovich Chebyshev* (Russian: Пафну́тий Льво́вич Чебышёв; May 16, 1821–December 8, 1894) was a Russian mathematician and is considered to be the founding father of Russian mathematics. The Russian letter *ё* in Чебышёв is pronounced "yo", making the native pronunciation "Cheby-SHOV", rhyming with "shove".

$$= \quad n + 2\sum_{i<j} \mathbb{E}\big[X_i X_j\big]$$
$$= \quad n + 2\sum_{i<j} \mathbb{E}[X_i]\,\mathbb{E}\big[X_j\big] \quad = \quad n.$$

Lemma 5.1.1. *Let $X_1, \ldots, X_n \in \{-1, +1\}$ be independent[5.1] random variables, each taking value in $\{-1, +1\}$ with probability $1/2$. For any $t > 0$, we have that*

$$\mathbb{P}\Big[\Big|\sum_i X_i\Big| \geq t\sqrt{n}\Big] \leq 1/t^2.$$

Proof: Let $Y = \sum_i X_i$ and $Z = Y^2$. Notice that $Z \geq 0$ with $\mathbb{E}[Z] = n$. We have

$$\mathbb{P}\Big[\Big|\sum_i X_i\Big| \geq t\sqrt{n}\Big] = \mathbb{P}\Big[\Big(\sum_i X_i\Big)^2 \geq t^2 n\Big] = \mathbb{P}\Big[Y^2 \geq t^2\,\mathbb{E}[Y^2]\Big]$$
$$= \mathbb{P}[Z \geq t^2\,\mathbb{E}[Z]] \leq \frac{1}{t^2},$$

by Markov's inequality. 

5.1.2. Chebyshev's inequality

As a reminder, the variance of a random variable X is

$$\mathbb{V}[X] = \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2] - \mu_X^2,$$

where $\mu_X = \mathbb{E}[X]$.

Theorem 5.1.2 (Chebyshev's inequality). *Let X be a real random variable, with $\mu_X = \mathbb{E}[X]$, and $\sigma_X = \sqrt{\mathbb{V}[X]}$. Then, for any $t > 0$, we have*

$$\mathbb{P}[|X - \mu_X| \geq t\sigma_X] \leq \frac{1}{t^2}.$$

Proof: Set $Y = (X - \mu_X)^2$, and observe that

$$\sigma_X^2 = \mathbb{V}[X] = \mathbb{E}[Y] = \mathbb{E}\Big[(X - \mu_X)^2\Big] = \mathbb{E}[X^2] - \mu_X^2.$$

[5.1] But really pairwise independent, more on that concept later on.

As such, we have that

$$\mathbb{P}[|X - \mu_X| \geq t\sigma_X] = \mathbb{P}\Big[(X - \mu_X)^2 \geq t^2\sigma_X^2\Big] = \mathbb{P}[Y \geq t^2\,\mathbb{E}[Y]] \leq \frac{1}{t^2},$$

by Markov's inequality. 

Remark 5.1.3 (Aside: pairwise independence.). Note, that Chebyshev's inequality only requires pairwise cancellation of cross-terms

$$\mathbb{E}\big[X_i X_j\big] = \mathbb{E}[X_i]\,\mathbb{E}\big[X_j\big],$$

for $i \neq j$, which is provided by *pairwise-independence*. This provides significant algorithmic utility in scenarios where full mutual independence is impossible or computationally costly to construct, such as universal hashing and streaming algorithms. Without going into this here, a group of variables $X_1, \ldots, X_n$ are ***mutually independent*** (i.e., independent) if knowing the value of all variables except one, does not provide any information about the value of this singleton variable. Pairwise independence on the other hand requires this property only hold only for subsets of two two variables.

Exercise 5.1.4 (Not too interesting, skip). Consider the case that the standard deviation and variance are both zero (i.e., $\sigma_X = \mathbb{V}[X] = 0$), what happens then with Chebyshev's inequality?

Example 5.1.5 (Variance of a binomial distribution). Let $Y_1, \ldots, Y_m$ are independent random 0/1 variables that are each one with probability p (i.e., a single Y_i is a ***Bernoulli random variable*** taking value 1 with probability p and 0 with probability $1-p$). Let $Y = \sum_{i=1}^{m} Y_i$, and observe that Y has a binomial distribution with parameters m and p. That is, $Y \sim \mathrm{Bin}(m, p)$. Thus, for all i, we have $\mathbb{E}[Y_i] = p$, and $\mathbb{V}[Y_i] = \mathbb{E}\big[Y_i^2\big] - (\mathbb{E}[Y_i])^2 = p - p^2 = p(1-p)$. Since variance is additive for pairwise independent random variables, we have

$$\mathbb{V}[Y] = \mathbb{V}\Big[\sum_i Y_i\Big] = \sum_{i=1}^{m} \mathbb{V}[Y_i] = mp(1-p). \tag{5.1}$$

This also implies that $\mathbb{E}[Y^2] = \mathbb{V}[Y] + (\mathbb{E}[Y])^2 = mp(1-p) + m^2p^2$.

5.2. Estimation via sampling

One of the main advantages of randomized algorithms is that they sample the world—that is, learn what the input looks like without reading all the input. For example, consider the following problem: We are given a set U of n objects $u_1, \ldots, u_n$, and we want to compute the number of elements of U that have some property. Assume that one can check if this property holds, in constant time, for a single object, and let $\psi(u)$ be the function that returns 1 if the property holds for the element u, and 0 otherwise. Now, let Γ be the number of objects in U that have this property. We want to reliably estimate Γ without computing the property for all the elements of U.

A natural approach is to pick a random sample R of m objects from U: $r_1, \ldots, r_m$ (with replacement), and compute $Y = \sum_{i=1}^{m} \psi(r_i)$. The estimate for Γ is $Z = (n/m)Y$. It is natural to ask how far Z is from the true value Γ.

Lemma 5.2.1. *Let U be a set of elements, where a fraction p of the elements have a certain property ψ. Let R be a uniform random sample from U (with replacement) of size m. Let Y be the number of elements in R that have the property ψ. We have that*

$$\mathbb{P}\Big[\mathbb{E}[Y] - t\sqrt{m}/2 \le Y \le \mathbb{E}[Y] + t\sqrt{m}/2\Big] \;\ge\; 1 - \frac{1}{t^2}.$$

Proof: Let $Y_i = \psi(r_i)$ be an indicator variable that is 1 if the ith sample r_i has the property ψ, for $i = 1, \ldots, m$. Observe that

$$p = \mathbb{E}[Y_i] = \frac{\Gamma}{n} \qquad \text{and} \qquad \mathbb{V}[Y_i] = p(1-p).$$

For the random variable $Y = \sum_{i=1}^{m} Y_i$, by linearity of expectation and Eq. (5.1), we have $\mathbb{E}[Y] = mp$ and $\mathbb{V}[Y] = mp(1-p) \le m/4$, implying that $\sigma_Y = \sqrt{\mathbb{V}[Y]} \le \sqrt{m}/2$. Since $t\sigma_Y \le t\sqrt{m}/2$, by Chebyshev's inequality we have that

$$\mathbb{P}\Big[|Y - \mathbb{E}[Y]| \ge t\sqrt{m}/2\Big] \;\le\; \mathbb{P}\Big[|Y - \mathbb{E}[Y]| \ge t\sigma_Y\Big] \;\le\; \frac{1}{t^2}.$$

Taking the complement yields

$$\mathbb{P}\big[|Y - \mathbb{E}[Y]| \le t\sqrt{m}/2\big] \ge 1 - \frac{1}{t^2},$$

which is equivalent to the stated bound. 

Example 5.2.2 (Numerical example—skip if you hate numbers.). Consider a setting where $U = \{0, 1\}$ and only 1 has the desired property and as such $p = 1/2$. Let us sample $m = 4 \cdot 10^6$ elements—this is completely equivalent to flipping a coin m times. The above says that the number of 1s (i.e., heads) we get is concentrated. Indeed, $\sigma_Y = \sqrt{m}/2 = 1000$ and $\mathbb{E}[Y] = m/2$, and by the above lemma, we have that

$$\mathbb{P}\left[\left|Y - 2 \cdot 10^6\right| \leq t1000\right] \geq 1 - \frac{1}{t^2}.$$

For $t = 10$, this implies that the probability that the total number of heads in $m = 4 \cdot 10^6$ coin flips is in the range

$$[2,000,000 - 10,000, 2,000,000 + 10,000]$$

with probability at least 99%. That is pretty crazy—flipping a coin is a totally random process for each coin flip, but in the aggregate it is pretty predictable.

Example 5.2.3. If the size of U is known, and you do not know p, then a natural estimate for p is to compute how many of the m samples have this property. That is, $\hat{p} = Y/m$ is an estimate of p (i.e., the fraction of the sample that has the desired property). In particular, we can now estimate how many elements in U have the property—that is, $\hat{p}n = \frac{n}{m}Y$.

Lemma 5.2.4. *In the setting of Lemma 5.2.1, let $n = |U|$ and assume that Γ of them have the property (i.e., $p = \frac{\Gamma}{n}$). Let $Z = \frac{n}{m}Y$ be the estimate for Γ. Then, for any $t \geq 1$, we have that*

$$\mathbb{P}\left[\Gamma - t\frac{n}{2\sqrt{m}} \leq Z \leq \Gamma + t\frac{n}{2\sqrt{m}}\right] \geq 1 - \frac{1}{t^2}.$$

Proof: By Lemma 5.2.1, we have

$$\begin{aligned}
\mathbb{P}\left[|Z - \Gamma| \geq t\frac{n}{2\sqrt{m}}\right] &= \mathbb{P}\left[|Z - \Gamma| \geq \frac{n}{m}t \cdot \frac{\sqrt{m}}{2}\right] \\
&\leq \mathbb{P}\left[|Z - \Gamma| \geq \frac{n}{m}t\sigma_Y\right] \\
&= \mathbb{P}\left[\left|\frac{n}{m}Y - \frac{n}{m}\mathbb{E}[Y]\right| \geq \frac{n}{m}t\sigma_Y\right]
\end{aligned}$$

$$= \mathbb{P}\left[|Y - \mathbb{E}[Y]| \geq t\sigma_Y\right] \leq \frac{1}{t^2}.$$

Taking the complement completes the proof. 

5.3. Randomized selection via sampling

5.3.1. Inverse estimation

We are given a set $U = \{u_1, \ldots, u_n\}$ of n distinct numbers. Let $U_{\langle i \rangle}$ denote the ith smallest number in U—that is, $U_{\langle i \rangle}$ is the element of ***rank*** i in U.

Lemma 5.3.1. *Given a set U of n numbers, a rank $k \in \{1, \ldots, n\}$, and parameters $t \geq 1$ and $m \geq 1$, one can compute, in $O(m \log m)$ time, two numbers $r_-, r_+ \in U$, such that:*

(A) The element of rank k in U is in the interval $\mathcal{I} = [r_-, r_+]$.

(B) There are at most $8tn/\sqrt{m}$ elements of U in $\mathcal{I}$.

The above two properties hold with probability $\geq 1 - 3/t^2$.

Proof: **(A)** Compute a random sample R of U of size m in $O(m)$ time (assuming the input numbers are given in an array, say). Next sort the numbers of R in $O(m \log m)$ time. Let

$$\ell_- = \left\lfloor m\frac{k}{n} - t\sqrt{m}/2 \right\rfloor - 1 \qquad \text{and} \qquad \ell_+ = \left\lceil m\frac{k}{n} + t\sqrt{m}/2 \right\rceil + 1.$$

Set $r_- = \mathsf{R}[\ell_-]$ and $r_+ = \mathsf{R}[\ell_+]$ (one needs to take care of boundary conditions[5.1]).

Let Y be the number of elements in the sample R that are $\leq U_{\langle k \rangle}$. By Lemma 5.2.1, we have

$$\mathbb{P}\left[\mathbb{E}[Y] - t\sqrt{m}/2 \leq Y \leq \mathbb{E}[Y] + t\sqrt{m}/2\right] \geq 1 - 1/t^2.$$

In particular, if this happens, then since R is sorted and $\ell_- < Y < \ell_+$, we have $r_- \leq U_{\langle k \rangle} \leq r_+$.

[5.1] Boundary condition if you must: If $\ell_- < 1$, we set $r_- = -\infty$ (or the minimum element of U); similarly, if $\ell_+ > m$, we set $r_+ = +\infty$ (or the maximum element of U).

(B) Let $g = k - t\frac{n}{\sqrt{m}} - 3\frac{n}{m}$, and let g_{R} be the number of elements in R that are at most $U_{\langle g \rangle}$ (i.e., $\leq U_{\langle g \rangle}$). Arguing as above, we have that

$$\mathbb{P}\left[g_{\mathsf{R}} \leq \frac{g}{n}m + t\sqrt{m}/2\right] \geq 1 - 1/t^2.$$

Now

$$\begin{aligned}\frac{g}{n}m + t\sqrt{m}/2 &= \frac{m}{n}\left(k - t\frac{n}{\sqrt{m}} - 3\frac{n}{m}\right) + t\sqrt{m}/2 \\ &= k\frac{m}{n} - t\sqrt{m} - 3 + t\sqrt{m}/2 \\ &= k\frac{m}{n} - t\sqrt{m}/2 - 3 < \ell_-.\end{aligned}$$

This implies that the g smallest numbers in U are outside the interval $[r_-, r_+]$ with probability $\geq 1 - 1/t^2$.

Next, let $h = k + t\frac{n}{\sqrt{m}} + 3\frac{n}{m}$. A similar argument shows that all the $n - h$ largest elements in U are strictly larger than r_+ (and hence outside $[r_-, r_+]$). This implies that

$$|[r_-, r_+] \cap U| \leq h - g = 6\frac{n}{m} + 2t\frac{n}{\sqrt{m}} \leq 8\frac{tn}{\sqrt{m}}.$$

In words, since ranks $1, \ldots, g$ and $h+1, \ldots, n$ are excluded, at most $h - g$ candidate elements remain in $\{g+1, \ldots, h\}$.

5.3.1.1. Inverse estimation – intuition Here we are trying to give some intuition to the proof of the previous lemma. Feel free to skip this part if you feel you already understand what is going on.

Given k, we are interested in estimating $s_k = U_{\langle k \rangle}$ quickly. So, let us take a sample R of size m. Let $\mathsf{R}_{\leq s_k}$ be the set of all the numbers in R that are $\leq s_k$. For $Y = \left|\mathsf{R}_{\leq s_k}\right|$, we have that $\mu = \mathbb{E}[Y] = m\frac{k}{n}$. Furthermore, for any $t \geq 1$, Lemma 5.2.1 implies that $\mathbb{P}[\mu - t\sqrt{m}/2 \leq Y \leq \mu + t\sqrt{m}/2] \geq 1 - 1/t^2$. In particular, with probability $\geq 1 - 1/t^2$ the number $r_- = \mathsf{R}_{\langle \ell_- \rangle}$, for $\ell_- = \lfloor \mu - t\sqrt{m}/2 \rfloor - 1$, is smaller than s_k, and similarly, the number $r_+ = \mathsf{R}_{\langle \ell_+ \rangle}$ of rank $\ell_+ = \lceil \mu + t\sqrt{m}/2 \rceil + 1$ in R is larger than s_k.

One can conceptually think about the interval $\mathcal{I}(k) = [r_-, r_+]$ as a *confidence interval*—we know that $s_k \in \mathcal{I}(k)$ with probability $\geq 1 - 1/t^2$.

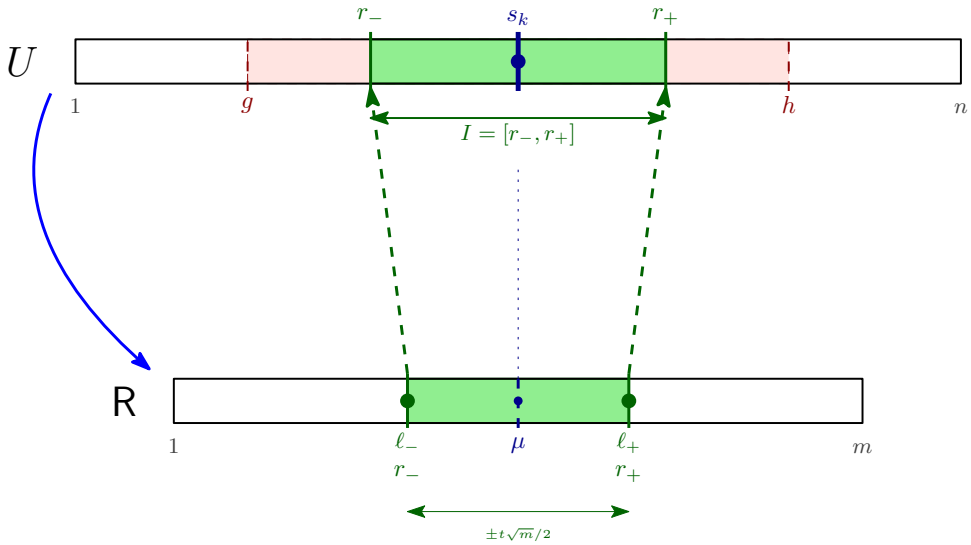

Figure 5.1: Inverse estimation via sampling: Sample pivots $r_-, r_+ \in \mathsf{R}$ are chosen around the expected rank μ, bracketing the rank-k element s_k in the interval $I = [r_-, r_+]$ in U. See Lemma 5.3.1 for details.

But how heavy is this interval? Namely, how many elements are there in $\mathcal{I}(k) \cap U$?

To this end, consider the interval of ranks, *in the sample*, that might contain the kth element. By the above, this is

$$\mathcal{I}(k,t) = k\frac{m}{n} + \left[-t\sqrt{m}/2 - 1, t\sqrt{m}/2 + 1\right].$$

In particular, consider the maximum $\nu \leq k$ such that $\mathcal{I}(\nu,t)$ and $\mathcal{I}(k,t)$ are disjoint. We have the condition that

$$\nu\frac{m}{n} + t\sqrt{m}/2 + 1 \leq k\frac{m}{n} - t\sqrt{m}/2 - 1 \implies \nu \leq k - t\frac{n}{\sqrt{m}} - 2\frac{n}{m}.$$

Let

$$g = k - t\frac{n}{\sqrt{m}} - 2\frac{n}{m} \qquad \text{and} \qquad h = k + t\frac{n}{\sqrt{m}} + 2\frac{n}{m}.$$

By construction, the sample rank intervals $\mathcal{I}(g,t)$, $\mathcal{I}(k,t)$, and $\mathcal{I}(h,t)$ are deterministically pairwise disjoint.

It is easy to verify (using the same argumentation as above) that with probability at least $1 - 3/t^2$, the three confidence intervals $\mathcal{I}(g), \mathcal{I}(k)$ and

$\mathcal{I}(h)$ do not intersect. As such, we have $\left|\mathcal{I}(k) \cap U\right| \leq h - g \leq 2t\frac{n}{\sqrt{m}} + 4\frac{n}{m} \leq 6t\frac{n}{\sqrt{m}}$.

5.3.2. Randomized selection

5.3.2.1. The algorithm. Given an array S of n numbers and a rank k, the algorithm needs to compute $S_{\langle k \rangle}$. To this end, it sets

$$t = \lceil n^{1/8} \rceil \qquad \text{and} \qquad m = \lceil n^{3/4} \rceil .$$

Using the algorithm of Lemma 5.3.1, in $O(m \log m)$ time, the algorithm gets two numbers r_- and r_+.

Remark 5.3.2. For the sake of self containment, here is this part described in detail. The algorithm now samples m elements into a set R with replacement from S. The algorithm then sorts the sample, let R_s be the sorted array containing the sample, and it compute the two numbers:

$$\ell_- = \left\lfloor \frac{k}{n^{1/4}} - \frac{t}{2}\sqrt{n} \right\rfloor - 1 \qquad \text{and} \qquad \ell_+ = \left\lfloor \frac{k}{n^{1/4}} + \frac{t}{2}\sqrt{n} \right\rfloor + 1.$$

We have that property that that $S_{\langle k \rangle} \in [r_-, r_+]$, and

$$\underbrace{|S \cap (r_-, r_+)|}_{S_m} = O(tn/\sqrt{m}) = O\big(n^{1/8+1}/n^{3/8}\big) = O(n^{3/4}),$$

and this happens with good probability (more on that later). To this end, the algorithm breaks S into three sets:

(i) $S_< = \{s \in S \mid s \leq r_-\}$,
(ii) $S_m = \{s \in S \mid r_- < s < r_+\}$, and
(iii) $S_> = \{s \in S \mid r_+ \leq s\}$.

This three-way partition can be done using $2n$ comparisons and in linear time. The algorithm computes the rank of r_- in S (it is $|S_<|$) and the rank of r_+ in S (it is $|S_<| + |S_m| + 1$). There are several cases:

- If $r_{-\langle S \rangle} > k$ or $r_{+\langle S \rangle} < k$ then the algorithm fails.
- If $|S_m| > 8tn/\sqrt{m} = O(n^{3/4})$, then the set S_m is too large, and the algorithm fails.

If either failure occurs, the algorithm restarts from scratch.

- If $k = |S_<|$, the algorithm returns r_- as the desired answer.

- If $k = |S_<| + |S_m| + 1$, then the algorithm returns r_+.
- Otherwise, $|S_<| < k \leq |S_<| + |S_m|$, and the algorithm needs to compute the element of rank $\Delta = k - |S_<|$ in the set S_m. This is done by sorting S_m and returning the element of rank Δ in the sorted subarray.

Remark 5.3.3. Crucially for the analysis, this algorithm is not recursive—it performs some rounds of sampling followed by partition, and then a single sorting of the interval of interest and extracting the desired element from this sorted subset. It is thus fundamentally different from QuickSelect.

5.3.2.2. Analysis. For the sake of simplicity of exposition, we ignore any floors/ceilings in the calculations here[5.1]. The correctness is intuitively easy—the algorithm clearly returns the desired element (the analysis below shows that the probability of running an infinite number of rounds is zero, which implies that the algorithm terminates almost surely[5.2]).

Running time. By Lemma 5.3.1, with probability $\geq 1 - 3/n^{1/4}$, the algorithm succeeds on the first try. The running time per round is $O(n + |S_m| \log |S_m|) = O(n)$, as $|S_m| = O(n^{3/4})$.

Not too many rounds. More generally, the probability that the algorithm fails in the first α tries to get a good interval $[r_-, r_+]$ is at most $(3/n^{1/4})^\alpha = O(3^\alpha/n^{\alpha/4})$. Let Y be the number of rounds the algorithm performs. The running time of the algorithm is $O(Yn)$. The variable Y is a geometric variable and has

$$\mu = \mathbb{E}[Y] \leq \frac{1}{1 - 3/n^{1/4}} \leq 1 + 4/n^{1/4}, \tag{5.2}$$

for n sufficiently large, since Y is a geometric random variable with success probability

$$p \geq 1 - 3/n^{1/4}. \tag{5.3}$$

Thus, the expected running time of the algorithm is $O(n)$.

[5.1] It is easy to verify the floor function only introduces tedium and anxiety in carrying out (essentially) the same calculations.

[5.2] In probability, ***almost surely*** means it happens with probability 1, so there is really no almost about it. It is probability's way of politely stating there is no escape—for example, almost surely we are all going to regret writing/reading this footnote.

5.3.2.3. Doing better. One can slightly improve the number of comparisons performed by the algorithm using the following modifications.

Lemma 5.3.4. *Given the numbers r_-, r_+ computed by Lemma 5.3.1, one can compute the sets $S_<$, S_m, and $S_>$ using in expectation (strikingly only) $1.5n + |S_m|/2$ comparisons.*

Proof: We need to compute the sets $S_<, S_m, S_>$. Namely, we need to compare all the numbers of S to r_- and r_+. Since only $O(n^{3/4})$ numbers fall in S_m, almost all of the numbers are in either $S_<$ or $S_>$. If a number is in $S_<$ (resp. $S_>$), then comparing it to r_- (resp. r_+) is enough to verify that this is indeed the case. Otherwise, perform the other comparison and put the element in its proper set (in this case we had to perform two comparisons to handle the element).

So let us guess, by a coin flip, for each element of S whether it is in $S_<$ or $S_>$. If we are right, then the algorithm would require only one comparison to put it into the right set. Otherwise, it would need two comparisons. Let X_s be the random variable that is the number of comparisons used by this algorithm for an element $s \in S$. We have that if $s \in S_< \cup S_>$ then $\mathbb{E}[X_s] = 1(1/2) + 2(1/2) = 3/2$. If $s \in S_m$ then both comparisons will be performed, and thus $\mathbb{E}[X_s] = 2$ in this case.

Thus, the expected number of comparisons for all the elements of S, by linearity of expectation, is $\frac{3}{2}(n - |S_m|) + 2|S_m| = (3/2)n + |S_m|/2$. 

Expected number of comparisons in a round. A round is successful with probability $\geq 1 - 3/n^{1/4}$, see Eq. (5.3). Thus, in expectation the algorithm performs

$$\begin{aligned}
\xi &= p \cdot \Big(1.5n + O(n^{3/4}) \log n\Big) + (1-p)2n \\
&\leq 1.5n + O(n^{3/4} \log n) + (3/n^{1/4}) \cdot 2n \\
&= 1.5n + O(n^{3/4} \log n)
\end{aligned}$$

comparisons per round.

Theorem 5.3.5. *Given an array S with n numbers and a rank k, one can compute the element of rank k in S in expected linear time. Formally, the resulting algorithm performs in expectation $1.5n + O(n^{3/4} \log n)$ comparisons.*

Proof: Intuitively, the algorithm performs Y rounds, and each round takes in expectation ξ comparisons. As the expectation of Y is $\mathbb{E}[Y] = 1/p \leq 1 + O(1/n^{1/4})$, it follows the expected number of comparisons is $\mathbb{E}[Y]\,\xi$ which is the desired quantity as an easy calculation shows. This argument is correct, but formalizing it requires some tedious work.

Let F be the expected number of comparisons performed by the algorithm. Consider the first round, with probability p it performed $1.5n + O(n^{3/4})$ comparisons, and with probability $1 - p$ it is a lost round, and the algorithm has to restart. In case of such a failure it had to perform (at most) $2n$ comparisons in this round, and whatever comparisons the algorithm performs later on. However, since the process is memoryless, the expected number of remaining comparisons the algorithm has to perform is F. Namely, we have that

$$\begin{aligned} F &\leq 1.5n + O(n^{3/4}\log n) + (1-p)(2n + F) \\ &\leq 1.5n + O(n^{3/4}\log n) + (1-p)F. \end{aligned}$$

This implies that

$$\begin{aligned} F &\leq \frac{1.5n + O(n^{3/4}\log n)}{p} \\ &\leq \left(1 + \frac{4}{n^{1/4}}\right) \cdot \left(1.5n + O(n^{3/4}\log n)\right) \\ &\leq 1.5n + O(n^{3/4}\log n), \end{aligned}$$

The recurrence above cleanly accounts for the expected $1.5n + O(n^{3/4}\log n)$ comparisons performed in partitioning across *all* rounds. In addition, sorting the sample R takes $O(m \log m) = O(n^{3/4}\log n)$ comparisons per round, and sorting S_m in the *final* successful round takes

$$O(|S_m| \log |S_m|) = O(n^{3/4}\log n)$$

comparisons. Since the expected number of rounds is

$$\mathbb{E}[Y] = 1 + O(1/n^{1/4}) \leq 2,$$

summing the partitioning and sorting costs yields the claimed total of $1.5n + O(n^{3/4}\log n)$ expected comparisons. 

The above implicitly proves a special case of the following—this more general result is stated here for the record, but is outside the scope of the material covered here (i.e., no need to read it).

Theorem 5.3.6 (Wald's Equation). *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) real-valued random variables with $\mathbb{E}[|X_1|] < \infty$. Let Y be an integer-valued stopping time with respect to the filtration generated by $(X_n)_{n\geq 1}$ such that $\mathbb{E}[Y] < \infty$. Then $\mathbb{E}\left[\sum_{i=1}^{Y} X_i\right] = \mathbb{E}[Y]\,\mathbb{E}[X_1]$.*

5.4. Tools from previous lecture

Theorem 5.4.1 (Markov's Inequality). *Let Y be a random variable assuming only non-negative values. Then for all $t > 0$, we have*

$$\mathbb{P}[Y \geq t] \leq \frac{\mathbb{E}[Y]}{t}.$$