

Class 4: Agenda, Questions, and Links

1 Warm-Up

Go to <http://PollEv.com/cs265> and answer the following question.

Suppose you flip a p -biased coin n times. Use Markov's inequality and Chebyshev's inequality to compute the probability that you see more than $2pn$ heads.

2 Announcements

- HW1 is due on Friday!

3 Questions?

Any questions from the minilectures or warmup? (Markov and Chebyshev's inequalities).

- Go into small groups and ask each other your questions.
- Ask any questions that the group can't resolve in the chat, or DM one of the course staff and we'll come to your room.

4 Sampling-Based Median

Sampling-Based Median Algorithm

Median:(A list S of n distinct numbers, where n is odd):

1. Let $t = n^{3/4}$. Sample $R = \{r_1, \dots, r_t\} \subseteq S$ by drawing r_i uniformly at random, independently.
2. Sort R in time $O(t \log t)$. Henceforth, assume that $r_1 \leq r_2 \leq \dots \leq r_t$.
3. Let $a = r_{t/2 - \sqrt{n}}$, $b = r_{t/2 + \sqrt{n}}$.
4. Let $N_{<a}$ and $N_{>b}$ denote the number of elements in S less than a and greater than b respectively.
5. Let $T = \{x \in S : a \leq x \leq b\}$. Construct T , and compute $N_{<a}$ and $N_{>b}$, in time $O(n)$.
6. If $|T| < 4t$, sort T in time $O(t \log t)$; otherwise output FAIL.
7. If $N_{<a}, N_{>b} \leq n/2$ (aka, $\text{median}(S) \in T$):

- Return the i 'th smallest element of T , where $i = (n + 1)/2 - N_{<a}$.
8. Otherwise, output FAIL.

[We'll see an example on a slide.]

[**Note:** Above there should be some floors or ceilings or something. Don't worry about it, and ignore off-by-one errors throughout this class.]

5 Analyzing the sampling-based median algorithm

You will analyze this algorithm in group work.

Group Work

Important: as you make progress on the question(s), one person in each room should record your progress on <http://PollEv.com/cs265>.

Note: Throughout this group work, don't worry about \leq vs $<$, or whether or not something is true up to ± 1 , or anything small like that.

1. Make sure that you all understand the algorithm. Pseudo-code is above, and Mary's one-slide example is available on the course website (cs265.stanford.edu), in the class-by-class resources for Class 4. Ask/answer any questions that you have amongst yourselves, and ask in chat or flag down a member of the course staff if you still have questions.
2. Suppose that you could show that:
 - with probability ≥ 0.9 , the median of S is in the list T ; and
 - with probability ≥ 0.9 , $|T| < 4t$.

Explain (to each other) why these two things would imply that the algorithm returns the correct answer with probability 0.8. And if it does not return the median then it returns FAIL.

3. Convince yourself that this algorithm uses at most $O(n)$ operations. (Assuming that "sample a random element of S " is a single operation).

At this point, please register your progress on PollEverywhere

4. In the following parts, you will show that the median of S is in T , with probability at least 0.9. Let m be the median of S . Consider two events:
 - $|\{r_i \in R : r_i < m\}| < \frac{t}{2} + \sqrt{n}$
 - $|\{r_i \in R : r_i > m\}| < \frac{t}{2} + \sqrt{n}$(a) Explain why, if both of these events hold, then $\text{median}(S) \in T$.

- (b) Use Chebyshev's inequality to bound the probability that the first event does not hold. (Hint: let X_i be the indicator random variable that is 1 iff $r_i \leq m$, and consider $\sum_i X_i$).
- (c) Convince yourself that the same argument will work for the second event, and write a statement of the form:

$$\Pr[\text{median}(S) \in T] \geq 1 - \text{----}.$$

At this point, please register your progress on PollEverywhere

5. Now, we turn our attention to the probability that $|T| < 4t$.
- (a) Explain why it is sufficient to show that a is *not* one of the smallest $n/2 - 2t$ elements of S , and b is *not* one of the largest $n/2 + 2t$ elements of S .
 - (b) Use Chebyshev's inequality to bound the probability that a is not one of the smallest $n/2 - 2t$ elements of S . (Hint: Consider the indicator random variable Y_i that is 1 if r_i is in the smallest $n/2 - 2t$ elements of S . Argue that a is one of the smallest $n/2 - 2t$ elements of S iff $\sum_i Y_i \geq t/2 - \sqrt{n}$ (why?) and apply Chebyshev's inequality.)
 - (c) Convince yourself that the analogous statement for b , and write a statement of the form:

$$\Pr[|T| < 4t] \geq 1 - \text{-----}.$$

At this point, please register your progress on PollEverywhere