1 Random Variables and Expectation

Plucky has an $n$-sided die that will generate numbers in $\{1, 2, \ldots, n\}$ uniformly at random. She is bored and she has decided to start and keep rolling her die until she has seen all the numbers in $\{1, 2, \ldots, n\}$ at least once.

We want to calculate how many die rolls it takes in expectation for Plucky to stop.

For each $i \in \{1, 2, \ldots, n\}$, we define a random variable $X_i$: its value is equal to the number of additional die rolls we need to see the $i$-th unique value after having already seen $i-1$ unique values.

What is the type of probability distribution that the random variable $X_i$ follows?

- Binomial
- Bernoulli
- Poisson
- Geometric

Correct

Assume Plucky has started rolling her die and she has seen $i-1$ unique values so far. What is the probability of seeing a new number that she has not seen before in her next die roll?

- $\frac{1}{n}$
- $\frac{i-1}{n}$
- $\frac{i}{n}$

Correct

What is $E[X_i]$?

- $\frac{i}{2}$
- $i$ (asymptotically)
- $i - 1$

Correct

What is the expected total number of die rolls, until Plucky sees all the $n$ values at least once?

- $\Theta(n)$
- $\Theta(n \log n)$
- $\Theta(n^2)$
- $\Omega(n^2 \log n)$

Correct

2 Randomized Algorithms

Can we use the random pivot selection idea in QuickSort for the selection problem?

Assume we modify the $k$-select algorithm that we saw in previous lectures; instead of picking the pivot cleverly, we just pick a uniformly random element as the pivot each time. We call the resulting algorithm QuickSelect.

What is the worst case runtime of QuickSelect?

- $\Theta(n)$
- $\Theta(n \log n)$
- $\Theta(n^2)$
- $\Omega(n^2 \log n)$

Correct

What is the probability that our random pivot partitions the array into two parts, each of size at most $\frac{3}{4}n$?

- $\frac{1}{2} \pm O(1/n)$
- $\frac{3}{4} \pm O(1/n)$
- $\frac{1}{2} \pm O(1/n)$
- $\frac{1}{2} \pm O(1/n)$

Correct

Assume we group QuickSelect’s recursive calls into multiple phases. Phase $i$ is when the size of the array is in the interval $\left(\left(\frac{3}{4}\right)^i n, \left(\frac{3}{4}\right)^{i+1} n\right]$. Note that we start at phase 0 with an array of size $n$.

For each phase we define a random variable $X_i$, whose value is the number of recursive calls in that phase. Using the answers to previous questions, calculate an upper bound for $E[X_i]$. Which of the following is the (asymptotically) smallest upper bound on $E[X_i]$?

- $3 + O(1/n)$
- $2 + O(1/n)$
- $n$
- $O(\log n)$

Correct

What is the expected (average case) runtime of QuickSelect?

- $\Theta(n)$
- $\Theta(n \log n)$
- $O(n^2)$
- $O(n^2 \log n)$

Correct