Interview Cake

← course home (/table-of-contents#section_combinatorics-probability-math_question_which-appears-twice)

I have a list of $n + 1$ numbers. Every number in the range $1..n$ appears once except for one number that appears twice.

Write a function for finding the number that appears twice.

Gotchas

We can do this with $O(1)$ additional memory.

Breakdown

To avoid using up extra memory space, lets use some math!

Solution

First, we sum all numbers $1..n$. We can do this using the equation:

$$\frac{n^2+n}{2}$$

because the numbers in $1..n$ are a triangular series.

Second, we sum all numbers in our input list, which should be the same as our other sum but with our repeat number added in twice. So the difference between these two sums is the repeated number!

Python 2.7

```
def find_repeat(numbers_list):
    if len(numbers_list) < 2:
        raise ValueError('Finding duplicate requires at least two numbers')

    n = len(numbers_list) - 1
    sum_without_duplicate = (n * n + n) / 2

    actual_sum = sum(numbers_list)

    return actual_sum - sum_without_duplicate
```

Complexity

$O(n)$ time. We can sum all the numbers $1..n$ in $O(1)$ time using the fancy formula, but it still takes $O(n)$ time to sum all the numbers in our input list.

$O(1)$ additional space—the only additional space we use is for numbers to hold the sums with and without the repeated value.

Bonus

If our list contains huge numbers or is really long, our sum might be so big it causes an integer overflow. What are some ways to protect against this?

← course home (/table-of-contents)

Next up: Find in Ordered Set → (/question/find-in-ordered-set?course=fc1&section=combinatorics-probability-math)