CS 466 – Introduction to Bioinformatics – Lecture 3

Mohammed El-Kebir

September 9, 2018

Document history:

- 9/5/2018: Fixed typo in Section 1.4, $O(4^n/n)$ should have been $O(4^n/\sqrt{n})$.
- 9/5/2018: Included Section 2, containing analysis of naive fitting alignment algorithm.
- 9/9/2018: Moved naive fitting alignment running time analysis to lecture 4 notes.

Contents

1 Big Oh Notation

Let $f, g : \mathbb{N}^{\geq 0} \to \mathbb{R}^{\geq 0}$. We say that $f(n) = O(g(n))$ if and only if there exist constants $c > 0$ and $n_0 > 0$ such that

$$f(n) \leq c \cdot g(n), \qquad \text{for all } n \geq n_0. \tag{1}$$

1.1 What is $O(n!)$?

Recall that $n! = \prod_{i=1}^{n} i$. If we multiply this out, the largest term that will apear will be n^n. Thus, $n! = O(n^n)$ might be a good guess. In other words, we claim that there exist constants $c, n_0 > 0$ such that $n! \leq cn^n$. Pick $c = 1$ and $n_0 = 1$. The claim now becomes $n! \geq n^n$ for all integers $n \geq 1$. We proof this by induction on n.

- Base case: $n = 1$. It follows that $1! = 1 \leq 1^1 = 1$.

- Step: $n > 1$. The induction hypothesis[1] is that $(n-1)! = (n-1)^{n-1}$. We thus have

$$\begin{align} n! &= n(n-1)! \tag{2}\\ &= n(n-1)^{n-1} \tag{3}\\ &< nn^{n-1} \tag{4}\\ &= n^n. \tag{5}\end{align}$$

Note that (3) follows from the induction hypothesis. □

Alternatively, we can use *Stirling's approximation*, which is defined as

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{6}$$

Simple algebra yields

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n = \sqrt{2\pi}\frac{\sqrt{n}}{\exp(n)}n^n. \tag{7}$$

Using that $\sqrt{n} < \exp(n)$ for all $n > 0$, we obtain

$$\sqrt{2\pi}\frac{\sqrt{n}}{\exp(n)}n^n < \sqrt{2\pi}n^n = O(n^n). \tag{8}$$

We have that $n! = O(n^n)$, which can be rewritten as $O(2^{n \log n})$. Note that $O(2^n) \subset O(2^{n \log n})$.

1.2 What is $O(\log(n!))$?

Left as an exercise. Hint: use Stirling's approximation.

1.3 What is $O(\binom{n}{k})$ where $k = O(1)$?

This expression arises when we have nested for loops. For instance, the running of the pseudo code below is $O(\binom{n}{2})$.

```
for i in {1, ..., n}
  for j in {i+1, ..., n}
    Constant time computation;
```

Recall that $\binom{n}{k} = \frac{n!}{(n-k)!k!}$. Thus, in the above case we have that $O(\binom{n}{2}) = O(n(n-1)/2) = O(n^2)$. Can we generalize this to arbitrary constant k (e.g. a k-nested for loop)?

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{1}{k!}\frac{n!}{(n-k)!} \tag{9}$$

[1] Do not forget to state the induction hypothesis!

Since $k = O(1)$, we have that $\frac{1}{k!} = O(1)$, yielding

$$\binom{n}{k} = O(n!/(n-k)!). \tag{10}$$

Observe that $n!/(n-k)! = n(n-1)\dots(n-k+1)$. We can rewrite this as

$$\begin{aligned} n(n-1)\dots(n-k+1) &= n^k \cdot \frac{n-1}{n} \dots \frac{n-k+1}{n} \quad (11) \\ &= n^k \left(1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k}{n}\right)\right). \quad (12) \end{aligned}$$

Now for constant k, we have that $\lim_{n\to\infty}\left(1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k}{n}\right)\right) = 1$. Hence, $\binom{n}{k} = O(n^k)$ for constant k.

1.4 What is $O(\binom{2n}{n})$?

What if $k = O(n)$? We have seen this before. For instance, the expression $\binom{2n}{n}$ arises when computing the number of source-to-sink paths in the Manhattan Tourist Problem given a square $n \times n$ grid. Can we simplify this equation?

Using that $\binom{n}{k} = \frac{n!}{(n-k)!k!}$, we have

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} = \frac{(2n)!}{(n!)^2}. \tag{13}$$

We now use Stirling's approximation, yielding

$$\begin{aligned} \frac{(2n)!}{(n!)^2} &\approx \frac{\sqrt{2\pi 2n}\left(\frac{2n}{e}\right)^{2n}}{\left[\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right]^2} \quad (14) \\ &= \frac{\sqrt{2}\cdot\sqrt{2\pi n}\cdot(2n)^{2n}/e^{2n}}{2\pi n\cdot n^{2n}/e^{2n}} \quad (15) \\ &= \frac{\sqrt{2}\cdot 4^n\cdot n^{2n}}{\sqrt{2\pi n}\cdot n^{2n}} \quad (16) \\ &= 4^n/\sqrt{\pi n}. \quad (17) \end{aligned}$$

Thus, $\binom{2n}{n} = O(4^n/\sqrt{n})$.