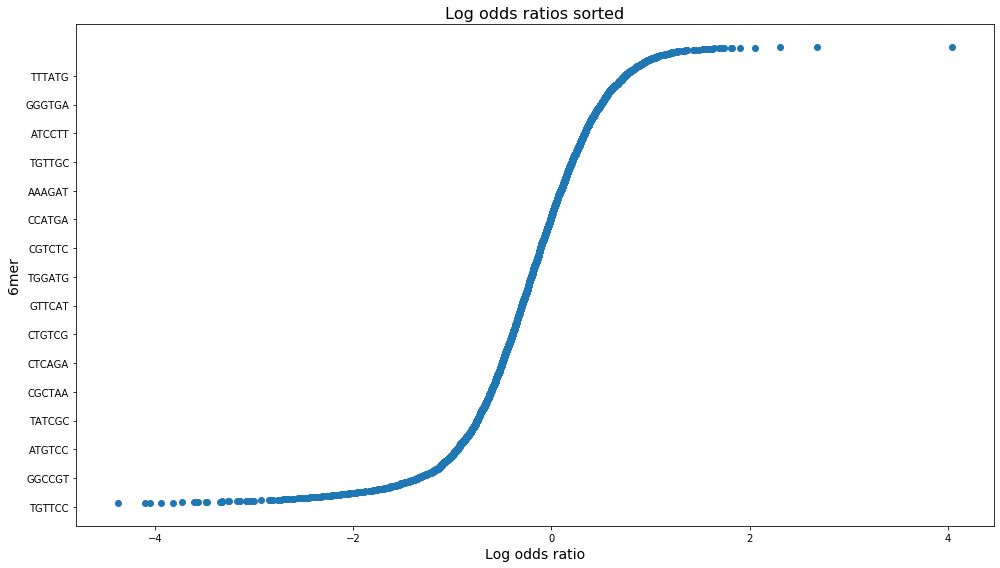
CSE520 Homework 4

Tony Adkins ([anadkins@uw.edu](mailto:anadkins@uw.edu))

# Genomics

## Sorted log odds ratios of 6mers



Maximum log odds ratio: TCTTTC, 4.03856

Minimum log odds ratio: CGACCT, -4.37356

The minimum above was calculated excluding the numerous 6mers which had zero probability, resulting in a negative infinity log odds ratio. The minimum of around -4 means the 6mer CGACCT has less than a 2:100 odds of splice usage, whereas the maximum of around 4 means something closer to 55:1, a nearly 3000-fold increase in odds likelihood.

## KL-divergence Gradient Descent

### ∂L/∂wk derivation

Where yi is a constant from the data and is defined as

So, we can take the derivative as the sum of the derivatives of the two parts of the sum, starting with the first half:

Which can be further split into two, allowing us to find the derivatives of each separately because

So, now to find

We must apply the chain rule F’(w) = f’(g(w))g’(w) where g(w) =

Which shows it will be useful to find the derivative of :

Which we apply the chain rule to get the derivative, e.g.

F’(w) = f’(g(w))g’(w), where f(w) =, and g(w) =

Then, to find f’, we find

And for g’, we find

And then we can put it all back together to get F’(x), or

Now, we can use this result to solve the original derivative applying the chain rule with being our g(w) in the equation F’(w) = f’(g(w))g’(w), so that we say

Similarly, we can find the derivative for

by applying the chain rule F(w) = f’(g(w))\*g’(w), where g(w) is

So, we solve for g’(w) by further applying the chain rule where now h(w) = , so

And we combine that to get

Which we then bring back to our original equation to solve

This just leaves us the last part of the original equation to derive, which is simple

Now, to put that all together, we say