


Background Frequencies

Interested in how different this motif position is from we expect by chance.

Correct for “expect by chance” by dividing by the probability of observing x in a random string:

$$\text{ScoreCorrected}(x) = \frac{\Pr(x \mid M)}{\Pr(x \mid \text{background})} = \prod_{i=1}^L \frac{e_i(x_i)}{b(x_i)}$$


$b(x_i)$:= probability of observing character x_i at random.
Usually computed as (# x_i in entire string) / (length of string)

Often, to avoid multiplying lots of terms, we take the log and then sum:

$$\text{ScoreCorrectedLog}(x) = \log \prod_{i=1}^L \frac{e_i(x_i)}{b(x_i)} = \sum_{i=1}^L \log \left(\frac{e_i(x_i)}{b(x_i)} \right)$$