BCB 567 Homework 5

Vu Thi-Hong-Ha; NetID: 851924086

December 16, 2019

1 Exercise 1:

s1.a+ $\rightarrow$ s1.b-
s4.b+ $\rightarrow$ s4.a-
s6.b+ $\rightarrow$ s6.a-

2 Exercise 2:

PAM_{120} is the closest estimate to the evolutionary distance between A and B.

3 Exercise 3:

Let $A[1]$, $A[2]$, ..., $A[n]$ denote a DNA sequence of length n;
Let p be a C+G content cutoff;
$max = 0$; $s = 0$; $start = -1$; $end = -1$;
for $(i = 1; i <= n; i = i + 1)$ {
    if $(A[i] == A$ or $T)$ {
        $w = -p$;
    } else {
        $w = 1 - p$;
    }
    $s = s + w$;
    if $(s <= 0)$ {
        $s = 0$;
        $b = i$;
    } else {
        if $(s > max)$ {
            $max = s$;
            $end = i$;
            if $(s == 1 - p)$ { // in case the region starts at i=1 and b is not initialized yet
                $start = i$;
            } else {
                $start = b + 1$;
            }
        }

```
    }
}
Print("Score of a highest-scoring region: ", max);
Print("Starting position of the region: ", start);
Print("Ending position of the region: ", end);
```

4 Exercise 4:

(a) The longest evolutionary distance in the tree is $t_1 + t_3 + t_2 + t_2 + t_3$.
(b)
$Prob(D^{(i)}|T) = \sum_x \sum_y \sum_z \sum_w Prob(G, G, G, C, G, y, z, w, x|T)$
$= \sum_x \sum_y \sum_z \sum_w Prob(x) \times Prob(w|x, t_3) \times Prob(z|x, t_3) \times Prob(y|z, t_2) \times Prob(G|z, t_2) \times Prob(G|z, t_1) \times Prob(G|y, t_1) \times Prob(G|y, t_1) \times Prob(C|w, t_2) \times Prob(G|w, t_2)$
$= \sum_x Prob(x) \times (\sum_w Prob(w|x, t_3) \times Prob(C|w, t_2) \times Prob(G|w, t_2)) \times (\sum_z Prob(G|z, t_2) \times Prob(z|x, t_3) \times \sum_y Prob(y|z, t_2) \times Prob(G|y, t_1) \times Prob(G|y, t_1))$
$= \sum_x Prob(x) \times (\sum_w (\sum_x \pi_x Prob_{xw}(t_3)) \sum_w \pi_w Prob_{wC}(t_2) \sum_w \pi_w Prob_{wG}(t_2))$
$\times (\sum_z (\sum_x \pi_x Prob_{xz}(t_3))(\sum_z \pi_z Prob_{zG}(t_2)) \times \sum_y (\sum_z \pi_z Prob_{zy}(t_2)) \sum_y \pi_y Prob_{yG}(t_1) \sum_y \pi_y Prob_{yG}(t_1)$