

Máme 5 zarovnaných sekvencií o dĺžke 5 nukleotidov(NTs).

Vieme, že pravdepodobnosť výskytu jednotlivých NTs pre tento prípad $A = T = C = G = 0.25$

1.	A	T	T	G	C
2.	T	T	T	G	C
3.	T	T	A	G	C
4.	G	T	A	G	C
5.	A	T	T	G	C

1. Vypočítam profilovú tabuľku, t.j. spočítam percentuálne zastúpenie jednotlivých NT na pozícii.

A	0.4	0	0.4	0	0
C	0	0	0	0	1.0
G	0.2	0	0	1.0	0
T	0.4	1.0	0.6	0	0

e.g. Stĺpec pre prvú pozíciu: Adenín(A) sa na prvej pozícii nachádza 2 razy z piatich $\rightarrow 2/5 = 0.4$, Cytosín (C) sa nevyskytol v našich zarovnaných seq, Guanín mám raz $\rightarrow 1/5 = 0.2$, zostáva Tyrosín 0.4. Rovnako sa pokračuje v ďalších stĺpcoch.

2. Z profilovej tabuľky spočítam maticu LLW (Log Likelihood Weight matrix). Za nulové hodnoty dosadím mínus nekonečno, sú nedosiahnuteľné, a nenulové hodnoty postupne prejdem vzťahom $\log_2(Pp/Pv)$, kde Pp je percentuálne zastúpenie nukleotidu na pozícii (jednoducho políčko v profilovej tabuľke) a Pv je pravdepodobnosť výskytu nukleotidu (definovaná na začiatku, všeobecne sa môže líšiť, nemusí byť vždy 25% pre všetky)

A	0.678	-inf	0.678	-inf	-inf
C	-inf	-inf	-inf	-inf	2.0
G	-0.32	-inf	-inf	2.0	-inf
T	0.678	2.0	1.263	-inf	-inf

e.g. Stĺpec pre prvú pozíciu: Pre A: $\log_2(0.4/0.25) = 0.678$,
pre C $-\text{inf}$,
pre G: $\log_2(0.2/0.25) = -0.32$,
pre T: $\log_2(0.4/0.25) = 0.678$, obdobne pokračujem ďalšími stĺpcami

3. Spočítam **relatívnu entropiu** pozícií, mám 5 pozícií, takže:

0.478	2	1.029	2	2
-------	---	-------	---	---

Výpočet pre 1. stĺpec (pre prvú pozíciu):

$0.678 * 0.4 + -0.32 * 0.2 + 0.678 * 0.4 = 0.478$, proste násobím odpovedajúce hodnoty v tabuľkách a v rámci stĺpcu tieto súčiny posčítam.

Výpočet pre 2. stĺpec (pre prvú pozíciu): $2 * 1 = 2$

Výpočet pre 3. stĺpec (pre prvú pozíciu): $0.678 * 0.4 + 1.263 * 0.6 = 1.029$

Máme všetky informácie na nakreslenie DNA sekvenčného loga, a) hodnoty v relatívnej entropii mi určujú výšku do ktorej budú siahať vyobrazenia stĺpcov. Pre DNA je celková výška loga vždy 2. (Pri proteínoch 4,3 + pri proteínoch treba aj LLW tabuľku robiť pomocou vzorca $\log_{4,3}(Pp/Pv)$)

b) proporčne, podľa % zastúpenia v profilovej tabuľke nakreslím do vymedzenej výšky príslušné NT.

