Xiheng He

Lisanne Friedrich

Exercises for Algorithmic Bioinformatics II

Assignment 4

Xiheng He

November 2021

Exercise 4 (BLAST E-value and P-value, 10P):

Given the following two alignments:

Length sequence 1	Length sequence 2	Score
100	200	80
100	120	160

and the parameters $K = 0.3$ and $\lambda = 0.1$.

(a) Calculate the E-values for both alignments and determine using the poisson distribution the probabilities, of 0, 1, 2 HSPs matching with at least the observed score.

E-value for alignment 1, 2:

$$
\begin{aligned}
E_1 &= K m_1 n_1 e^{-\lambda S_1} \\
&= 0.3 \cdot m_1 \cdot n_1 \cdot e^{-0.1 \cdot S_1} \\
&= 0.3 \cdot 100 \cdot 200 \cdot e^{-0.180} \\
&= 2.01277576742 \approx 2.01 \\
E_2 &= K m_2 n_2 e^{-\lambda S_2} \\
&= 0.3 \cdot m_2 \cdot n_2 \cdot e^{-0.1 \cdot S_2} \\
&= 0.3 \cdot 100 \cdot 120 \cdot e^{-0.1160} \\
&= 0.00040512662 \approx 0.0004
\end{aligned}
$$

Bit score:

$$S_1' = \frac{\lambda \cdot S_1 - \ln K}{\ln 2} = \frac{0.1 \cdot 80 - \ln 0.3}{\ln 2} = 13.2785259213 \approx 13.28$$
$$S_2' = \frac{\lambda \cdot S_2 - \ln K}{\ln 2} = \frac{0.1 \cdot 160 - \ln 0.3}{\ln 2} = 24.8200862484 \approx 24.82$$

Poisson probabilities of 0, 1, 2 HSPs matching with at least the observed score:

$$P_1(S_1', 0) = e^{-E_1} = 0.13398867466 \approx 0.13$$
$$P_1(S_1', 1) = \frac{e^{-E_1} \cdot E_1^1}{1!} = 0.26931723608 \approx 0.27$$
$$P_1(S_1', 2) = \frac{e^{-E_1} \cdot E_1^2}{2!} = 0.27066382226 \approx 0.27$$
$$P_2(S_2', 0) = e^{-E_2} = 0.99960007998 \approx 0.9996$$
$$P_2(S_2', 1) = \frac{e^{-E_2} \cdot E_2^1}{1!} = 0.00039984003 \approx 0.0004$$
$$P_2(S_2', 2) = \frac{e^{-E_2} \cdot E_2^2}{2!} = 7.99680063991 \times 10^8 \approx 0$$

(b) Calculate the P-values for both alignments. Are both alignments statistically significant?

$$P_1 = 1 - e^{-E_1} = 0.86601132533 \approx 0.8660$$
$$P_2 = 1 - e^{-E_2} = 0.00039992001 \approx 0.0004$$

According to the P-value, the alignment 2 is statistically significant.

(c) Consider some method which calculates normal-distributed scores ($\mu = \pi$ and $\sigma = e$). How probably is a score of at least 11.3? How do you calculate this probability?

PDF of normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$
$$f(11.3) = \frac{1}{e\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{11.3-\pi}{e}\right)^2\right) = 0.001623988046 \approx 0.0016$$