Exercises2

Yongqi WANG, Hangjia ZHAO

Problem 1

a)

Since there are 20 unique amino acids, there exists 20^{20} unique amino acid sequences for a sequence of $L = 20$.

b)

Codon consists of 3 nucleotides, so the DNA sequence encoding the amino acids sequence is of length of $20 \times 3 = 60$ and has 4^{60} unique sequences.

Problem 2

a)

Let x, y be two binary sequences of length L, $P(z_i = 0) = P(z_i = 1) = \frac{1}{2}, z \in \{x, y\}, i \in \{0, 1, \dots, L\}$.

Let a, b be two DNA sequences of length L, $P(c_i = A) = P(c_i = T) = P(c_i = C) = P(c_i = G) = \frac{1}{4}, c \in \{a, b\}, i \in \{0, 1, \dots, L\}$

Binary:

$$E(d_H(x,y)) = E(\sum_{i=1}^{L} 1_{x_i \neq y_i}) = \sum_{i=1}^{L} E(1_{x_i \neq y_i}) = L \cdot P(x_i \neq y_i) = \frac{L}{2}$$

Random DNA:

$$E(d_H(a,b)) = E(\sum_{i=1}^{L} 1_{a_i \neq b_i}) = \sum_{i=1}^{L} E(1_{a_i \neq b_i}) = L \cdot P(a_i \neq b_i) = \frac{3L}{4}$$

b)

Binary: $\binom{L}{K}$ of sequences at a Hamming distance K since the size of $\mathcal{A}$ is 2, $\binom{L}{2}$ sequences at a Hamming distance of 2.

Random DNA: $\binom{L}{K} \cdot 3^K$ because of the increase size in the alphabet $\mathcal{A}$ to 4.

Problem 3

a)

$$W = \begin{bmatrix} f_0 q_{00} & f_1 q_{10} \\ f_0 q_{01} & f_1 q_{11} \end{bmatrix} = \begin{bmatrix} f_0 q & 1-q \\ f_0(1-q) & q \end{bmatrix}$$

The eigenvalues of the above matrix by solving $det(W - \lambda I) = 0$:

$$\lambda_{1,2} = \frac{q(f_0+1) \pm \sqrt{q^2(f_0+1)^2 - 4f_0(2q-1)}}{2}$$

b)

Since $f_0 > 1$ and $0 \leq q \leq 1$, the matrix W is non-negative. Besides, the shape of W is $2 \times 2, (I + W)^{n-1} = I + W > 0$, so the matrix W is irreducible. We can apply the Perron–Frobenius theorem.

According to this theorem, λ_{max} is simple and the components of the associated eigenvector are all (strictly) positive. Therefore, the largest eigenvalue corresponds the non-trivial equilibrium point.

c)

Assumed $f_0 = 1.5 > 1, q = 0.9$

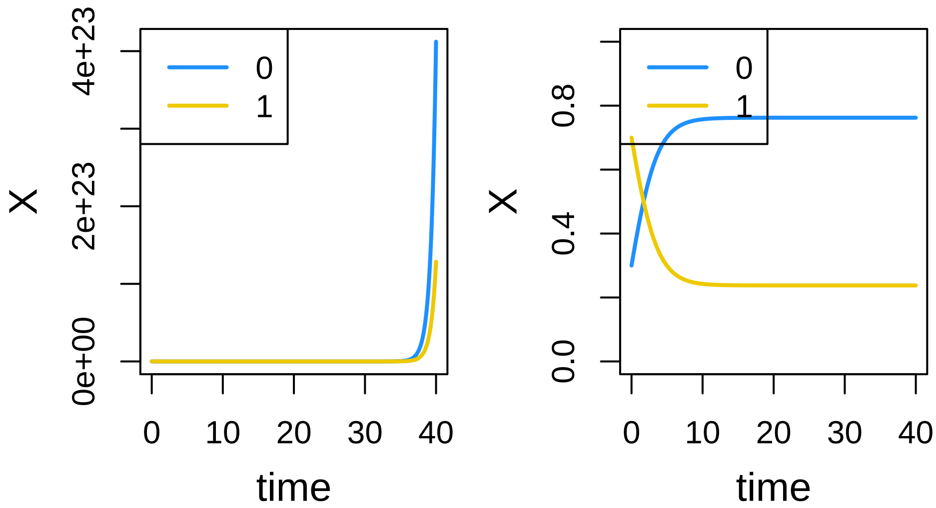
```
library(matlib)
f0 = 1.5
q = 0.9
parms <- c(f0 = f0, f1 = 1, q_00 = q, q_01 = 1-q, q_10 = 1-q, q_11 = q)
times <- c(0:1000)/25
initconds <- c(a = 0.3, b = 0.7)
M = matrix(c(parms["f0"] * parms["q_00"], parms["f0"] * parms["q_01"],
             parms["f1"] * parms["q_10"], parms["f1"] * parms["q_11"]),
           2, 2)
eig = eigen(M)
#M
#eig

x_star = eig$vectors[, 1]/sum(eig$vectors[, 1])
x_star

sum(x_star * c(parms["f0"], parms["f1"]))

ivp = solve(eig$vectors, initconds)
Xa = ivp[1] * eig$vectors[1, 1] * exp(eig$values[1] * times) +
ivp[2] * eig$vectors[1, 2] * exp(eig$values[2] * times)
Xb = ivp[1] * eig$vectors[2, 1] * exp(eig$values[1] * times) +
ivp[2] * eig$vectors[2, 2] * exp(eig$values[2] * times)

par(mar = c(4, 4, 2, 0.5)) # margin size
par(mgp = c(2.5, 1, 0)) # axis localtion
par(cex.lab = 1.25) # size of y axis label
par(mfrow=c(1,2)) # 1x2 fig
plot(times, Xa, xlab = "time", ylab = expression(X), main = "",
     col = "dodgerblue", type = "l", lwd = 2)
lines(times, Xb, col = "gold2", lwd = 2)
legend("topleft", legend=c("0", "1"),
       col=c("dodgerblue", "gold2"), lty=1, lwd = 2)
plot(times, Xa/(Xa + Xb), ylim = c(0, 1), xlab = "time", ylab = expression(X),
     main = "", col = "dodgerblue", type = "l", lwd = 2)
lines(times, Xb/(Xa + Xb), col = "gold2", lwd = 2)
legend("topleft", legend=c("0", "1"),
       col=c("dodgerblue", "gold2"), lty=1, lwd = 2)
```



d)

For $f_0 = f_1 = 1$,

$$\lambda_1 = \frac{2q + \sqrt{4(q-1)^2}}{2} = 2q - 1$$

$$\lambda_2 = \frac{2q - \sqrt{4(q-1)^2}}{2} = 1$$

Since $q \leq 1, \lambda_{max} = \phi = x_0 + x_1 = 1$

Besides,

$$\begin{bmatrix} q & 1-q \\ 1-q & q \end{bmatrix} \times \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

we get $x_0 = x_1 = 0.5$

e)

For $q \approx 1$,

$$\lambda_1 = \frac{f_0 + 1 + \sqrt{(f_0 - 1)^2}}{2} = f_0$$

$$\lambda_2 = \frac{f_0 + 1 - \sqrt{(f_0 - 1)^2}}{2} = 1$$

Since $f_0 > 1, \lambda_{max} = \phi = f_0 x_0 + x_1 = f_0$

Besides, $x_0 + x_1 = 1$

We get $x_0 = 1$ and $x_1 = 0$

Code is available on github repo: (https://github.com/wyq977/evolutionary-dynamics-2019)