

Übung 3

Dienstag, 17. Mai 2022

15:10

Aufgabe 3.1 Evolutionärer Distanzen

Teilaufgabe 3.1.1 Berechnung evolutionärer Distanzen (20%)

Sie haben die folgenden vier DNA Sequenzen verschiedener Spezies.

Spezies	1	G	A	T	A	G
"	2	G	A	T	A	A
"	3	A	T	G	G	A
"	4	G	A	T	A	G

(1)

Bestimmen Sie deren

- (a) p -Distanz,
- (b) Poisson-korrigierte Distanz, und
- (c) Jukes-Cantor Distanz.

Diskutieren Sie die Ergebnisse.

Teilaufgabe 3.1.2 Prozentuale Übereinstimmung (10%)

Die Prozentuale Übereinstimmung zwischen zwei Sequenzen sei I und somit die Übereinstimmung $I' = 100 \times I$. Zeigen Sie, dass die Poisson-korrigierte Distanz zwischen den beiden Sequenzen

$$d_p = -\ln(I')$$
(2)

ergibt.

a) p - Distanz

	1	2	3	4
1	0	0,2	1	0
2	0,2	0	0,8	0,2
3	1	0,8	0	1
4	0	0,2	1	0

$$p = \frac{D}{L}$$

mit L = Länge der 2 Sequenzen
und D = Anzahl der Pos. an denen
sich die Seq. unterscheiden

$$p_{12} = \frac{1}{5} = 0,2$$

$$p_{24} = \frac{1}{5} = 0,2$$

$$p_{13} = \frac{5}{5} = 1$$

$$p_{34} = \frac{5}{5} = 1$$

$$p_{14} = \frac{0}{5} = 0$$

$$p_{23} = \frac{4}{5} = 0,8$$

b) Poisson - korrigierte Distanz

→ berücksichtigt das wiederholte Auftauchen von Mutationen an derselben Position

$$d_p = -\ln(1-p)$$

$$d_{p_{12}} = -\ln(1-0,2) = 0,22$$

$$d_{p_{24}} = 0,22$$

$$d_{p_{13}} = \text{Math error}$$

$$d_{p_{34}} = \text{Math error}$$

$$d_{p_{14}} = 0$$

$$d_{p_{23}} = 1,61$$

c) Jukes - Cantor Distanz

→ alternative Korrektur, die aussagt, mit welcher Rate α Mutationen zwischen Nukleotiden auftreten

(nicht für hohe p)

$$d_{jc} = -\frac{3}{4} \ln\left(1 - \frac{4}{3}p\right)$$

$$d_{p_{12}} = -\frac{3}{4} \ln\left(1 - \frac{4}{3} \cdot 0,2\right)$$

$$= 0,23$$

$$d_{p_{24}} = 0,23$$

$$d_{p_{13}} = \text{Math Error}$$

$$d_{p_{34}} = \text{Math error}$$

$$d_{p_{14}} = 0$$

$$d_{p_{23}} = \text{Math error}$$

zu hohe p

→ Mit der p -Distanz lässt sich am einfachsten die Distanz

zweier Sequenzen bestimmen, allerdings wird hierbei die

reale evolutionäre Distanz unterschätzt, da die Möglichkeit,

dass an einer Stelle wiederholt Mutationen auftreten,

komplett unbeachtet bleibt. Dies wird mit der Poisson-Korrektur

ausgeglichen, die diese mitberechnet. Für kleine p ändert sich nicht viel, für große p ist sie allerdings sehr wichtig.

Wenn aber $p=1$ ist, geht die Kurve ins Unendliche

(Math error), die Distanz damit auch unendlich groß.

Umgekehrt ist es mit der Jukes-Cantor Distanz, die für

kleine p verlässlich aussagt, mit welcher Rate α Mutationen

zwischen Nukleotiden auftreten. Für große p ist diese

Berechnung aber nicht zuverlässig. Daher ergänzen sich die

Poisson-Korrektur und die Jukes-Cantor Distanz.

Teilaufgabe 3.1.2

→ ich denke, es sollte $I' = I/100$ sein,
da es sonst nicht passt...

I = prozentuale Übereinstimmung 2er Sequenzen

I' = Übereinstimmung

Bsp.

$$I = 90\%$$

$$I' = 90\% / 100 = 0,9$$

• zu zeigen:

$$d_p = -\ln(I')$$

$$I = 100 \cdot \frac{M}{L-M}$$

$L-M$

M

mit M = alle Positionen, die
übereinstimmen

übereinstimmen

$$p = \frac{L-M}{L} = 1 - \frac{M}{L}$$

$$I = 100 \cdot \frac{M}{L}$$

$$I' = \frac{100 \cdot \frac{M}{L}}{100} = \frac{M}{L}$$

$$\begin{aligned} d_p &= -\ln(1-p) \\ &= -\ln\left(1 - \left(1 - \frac{M}{L}\right)\right) \\ &= -\ln\left(1 - 1 + \frac{M}{L}\right) \\ &= -\ln\left(0 + \frac{M}{L}\right) \\ &= -\ln\left(\frac{M}{L}\right) \\ &= -\ln(I') \end{aligned}$$

Aufgabe 3.2 Konstruktion phylogenetischer Bäume (40 %)

Sie haben durch den Vergleich der DNA-Sequenzen die folgende Distanzmatrix erhalten

$$D = \begin{pmatrix} \textcolor{red}{A} & \textcolor{red}{B} & \textcolor{red}{C} & \textcolor{red}{D} & \textcolor{red}{E} \\ 0 & 5 & 9 & 9 & 8 \\ 5 & 0 & 10 & 10 & 9 \\ 9 & 10 & 0 & 8 & 7 \\ 9 & 10 & 8 & 0 & 3 \\ 8 & 9 & 7 & 3 & 0 \end{pmatrix} \begin{matrix} \textcolor{red}{A} \\ \textcolor{red}{B} \\ \textcolor{red}{C} \\ \textcolor{red}{D} \\ \textcolor{red}{E} \end{matrix} \quad (3)$$

Teilaufgabe 3.2.1 Überprüfung von Distanzmatrizen

Handelt es sich dabei um eine valide Distanzmatrix für evolutionäre Distanzen?
Überprüfen sie hierzu die Identität des Ununterscheidbaren und die Symmetrie.

Teilaufgabe 3.2.2 UPGMA Methode

Konstruieren Sie einen phylogenetischen Baum mit der UPGMA Methode.

3.2.1

• 2 mathematische Eigenschaften für eine valide Distanzmatrix

→ die Identität des Ununterscheidbaren

$$d(\text{Sequenz 1}, \text{Sequenz 1}) = 0$$

d.h. die Distanz einer Sequenz zu sich selbst ist 0

→ hier in der Matrix ist ersichtlich

z.B. $d(A, A) = 0$

→ die Symmetrie

$$d(\text{Sequenz 1}, \text{Sequenz 2}) = d(\text{Sequenz 2}, \text{Sequenz 1})$$

→ die Matrix ist gespiegelt

z.B. $d(A, B) = 5 = d(B, A)$

3.2.2

	A	B	C	D	E
A	0	5	9	9	8
B	5	0	10	10	9
D = C	9	10	0	8	7
D	9	10	8	0	3
E	8	9	7	3	0

$$d_{xy} = \frac{1}{N_x N_y} \sum_{i \in x; j \in y} d_{ij}$$

a)

die kleinste Distanz ist zwischen D und E

$$d(DE) = \frac{1}{1 \cdot 1} (3) = 3$$

$$\rightarrow DE = X$$

· um die Matrix zu aktualisieren, wird der Durchschnitt der Distanzen von D und E genommen

	A	B	C	X	
A	0	5	9	8,5	$d(XA) = \frac{1}{2} (DA, EA) = \frac{1}{2} (9 + 8) = 8,5$
B	5	0	10	9,5	$d(XB) = \frac{1}{2} (DB, EB) = \frac{1}{2} (10 + 9) = 9,5$
					$d(XC) = \frac{1}{2} (DC, EC) = \frac{1}{2} (8 + 7) = 7,5$

C	9	10	0	7,5
X	8,5	9,5	7,5	0

b)

die kleinste Distanz ist zwischen B und A

$$d(BA) = \frac{1}{1+1} (5) = 5$$

$$\rightarrow d(BA) = 5$$

	C	X	Y
C	0	7,5	9,5
X	7,5	0	4,5
Y	9,5	4,5	0

$$d(XY) = \frac{1}{2+2} (X_A, X_B) = \frac{1}{4} \cdot 18 = 4,5$$

$$d(CY) = \frac{1}{2+1} (A_C, B_C) = \frac{1}{3} \cdot 14 = 4,67$$

c)

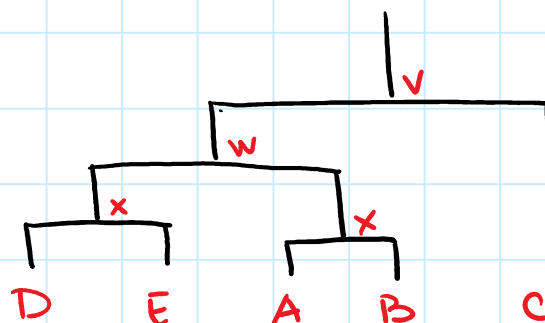
die kleinste Distanz ist zwischen X und Y

$$d(XY) = 4,5$$

$$\rightarrow d(XY) = W$$

	C	W
C	0	4,25
W	4,25	0

$$d(CW) = \frac{1}{4+1} (C_X, C_Y) = \frac{1}{5} (17) = 3,4$$



Aufgabe 3.3 Konstruktion phylogenetischer Bäume mit R (30 %)

Nutzen Sie die R library PHANGORN um phylogenetische Bäume zu erstellen. Führen Sie dazu die folgenden Befehle aus um einen Teil der DNA Sequenz von vierzehn Primaten zu laden.

```
library(phangorn)
fdir <- system.file("extdata/trees", package = "phangorn")
primates <- read.phyDat(file.path(fdir, "primates.dna"),
                        format = "interleaved")
```

Erstellen Sie phylogenetische Bäume mit der UPGMA Methode und der Neighbour-joining Methode. Vergleichen Sie und Interpretieren Sie die Ergebnisse. Beantworten Sie insbesondere die folgenden Fragen

1. Befindet sich unter den Species ein Nicht-Primat? Wie ist dessen Position in den Bäumen zu beschreiben?
2. Mit welchen anderen zwei Primaten formt 'Lemur' ein Cluster?
3. Welcher Primat formt mit dem Menschen ein Cluster

1. Mouse und Bovine

- Mouse bildet ein eigenes Cluster
- Bovine bildet ein Cluster mit Lemur und Tarsier

2. Lemur formt mit Bovine und Tarsier ein Cluster

3. Human formt mit Chimp ein Cluster (Menschen & Schimpansen)

