האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

אלגוריתמים בביו' חישובית
76558

מטריצות קצב באבולוציה של רצפים

תומי קפלן

11/2/2024

מטריצות מעברים ומטריצת קצב

- קיבלנו משוואה דיפרנציאלית $\frac{\partial P(t)}{\partial t} = P(t) \cdot \frac{\partial P(0)}{\partial t}$
- שפתרונה $P(t) = e^{[tR]}$

[מטריצת הקצב R]

מאפשר לנו לחשב את מטריצת המעברים P[a→b] לכל זמן t

- מיהי R המסתורית? מה הכוונה "מטריצת קצב השינוי ב-P מסביב לזמן אפס"?

מטריצות מעברים ומטריצת קצב

- ננסה להכיר את R

$$R = \lim_{\epsilon \to 0} \frac{P(\epsilon) - I}{\epsilon}$$

- כלומר בקירוב ראשון

$$P(\epsilon) \approx I + \epsilon R$$

P(ε)=

Pε	A	C	G	T
A				
C				
G				
T				

P(0)=

Pε	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

R=

R	A	C	G	T
A	-	+	+	+
C	+	-	+	+
G	+	+	-	+
T	+	+	+	-

$$R_{i,j} = \begin{cases} \leq 0 & i = j \\ \geq 0 & i \neq j \end{cases}$$

$$\forall_i \sum_j R_{i,j} = 0$$

מטריצת הקצב של ג'וקס וקנטור

- Jukes-Cantor [69']
- מטריצת קצב פשטנית וסימטרית

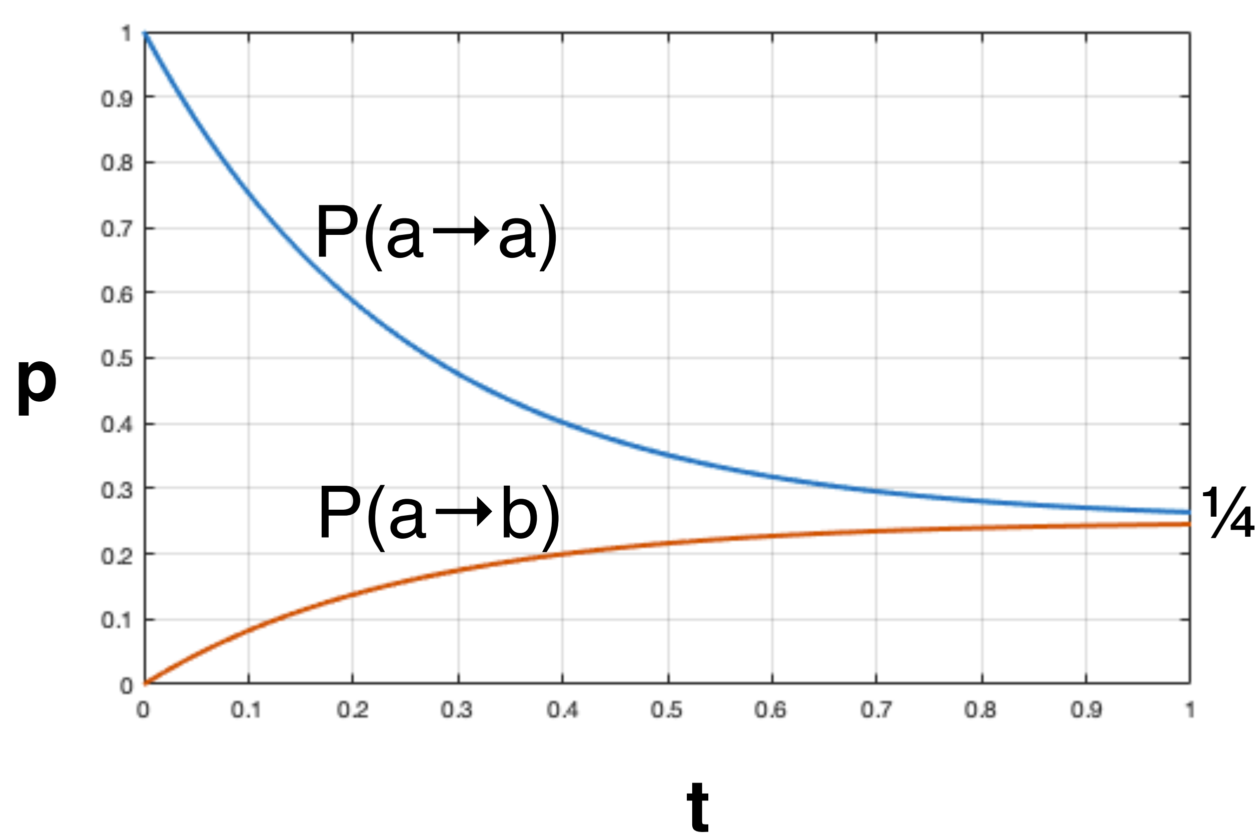
R=

R	A	C	G	T
A	-3α	α	α	α
C	α	-3α	α	α
G	α	α	-3α	α
T	α	α	α	-3α

$$R_{i,j} = \begin{cases} -3\alpha & i = j \\ \alpha & i \neq j \end{cases}$$

- ומטריצות המעברים:

$$P(t)_{i,j} = \begin{cases} \frac{1}{4}(1 + 3e^{-4\alpha t}) & i = j \\ \frac{1}{4}(1 - e^{-4\alpha t}) & i \neq j \end{cases}$$



מטריצת הקצב של ג׳וקס וקנטור

- נפתור יחדיו
- מטעמי סימטריה:

P(t)=

	A	C	G	T
A	r	s	s	s
C	s	r	s	s
G	s	s	r	s
T	s	s	s	r

R=

	A	C	G	T
A	-3α	α	α	α
C	α	-3α	α	α
G	α	α	-3α	α
T	α	α	α	-3α

$$\frac{\partial P(t)}{\partial t} = P(t) \times R$$

$$\begin{cases} r' = & r\cdot(-3\alpha)+3s\alpha \\ s' = & r\alpha - 3\alpha s + 2\alpha s \end{cases}$$

- והרי סכום כל שורה ב P[t] הוא 1 ← $r = 1-3s$

$$\begin{cases} r' = & (1-3s)\cdot(-3\alpha)+3s\alpha = -3\alpha\cdot(1-4s) \\ s' = & (1-3s)\alpha - 3\alpha s + 2\alpha s = \alpha\cdot(1-4s) \end{cases}$$

- נפתור את המשוואה $s' = \alpha\cdot(1-4s)$

מטריצת הקצב של ג'וקס וקנטור

- נפתור את המשוואה $s' = \alpha \cdot (1 - 4s)$
- ונקבל $s = \frac{1}{4}(1 - e^{-4\alpha t})$
- וגם $r = 1 - 3 \cdot \frac{1}{4}(1 - e^{-4\alpha t}) = \frac{1}{4}(1 + 3e^{-4\alpha t})$
- ונשים לב כי $(1 - 4s) = e^{-4\alpha t}$

- נגזור ונבדוק $s' = \left(\frac{1}{4} - \frac{1}{4}e^{-4\alpha t}\right)' = -\frac{1}{4}e^{-4\alpha t} \cdot (-4\alpha) = \alpha e^{-4\alpha t}$

$$r' = \left(\frac{1}{4} + \frac{3}{4}e^{-4\alpha t}\right)' = \frac{3}{4}e^{-4\alpha t} \cdot (-4\alpha) = -3\alpha e^{-4\alpha t}$$

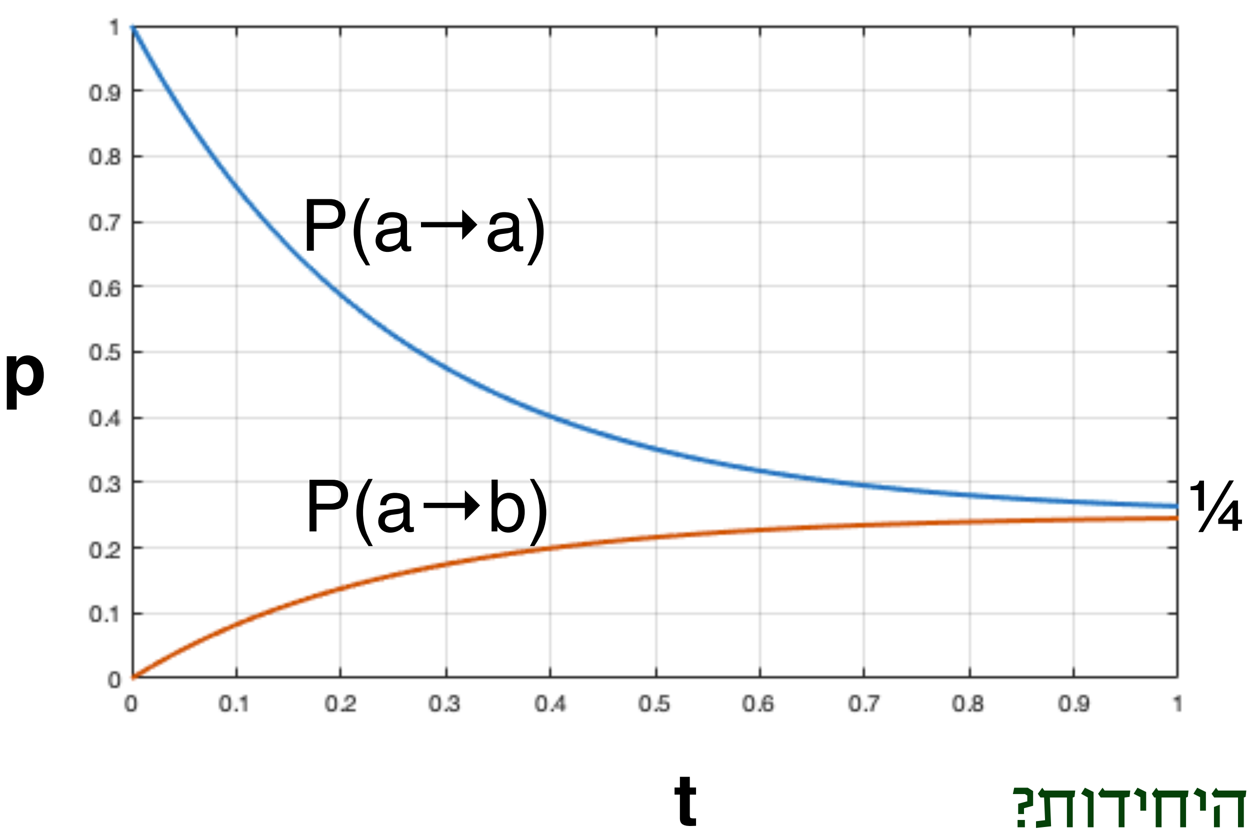
$$s' = \alpha \cdot (1 - 4s)$$
$$r' = -3\alpha \cdot (1 - 4s)$$

מטריצת הקצב של ג'וקס וקנטור

- אז למי שהתבלבלו בינתיים, מה קיבלנו?
- מטריצות המעברים לזמן t:

P(t)=

P(t)	A	C	G	T
A	r	s	s	s
C	s	r	s	s
G	s	s	r	s
T	s	s	s	r



$$P(t)_{i,j} = \begin{cases} r = \frac{1}{4}(1 + 3e^{-4\alpha t}) & i = j \\ s = \frac{1}{4}(1 - e^{-4\alpha t}) & i \neq j \end{cases}$$

מה היחידות?

מטריצת הקצב של קימורה

- Kimura [80']
- אותו רעיון אבל עם הפרדה לפורינים ופירמידינים

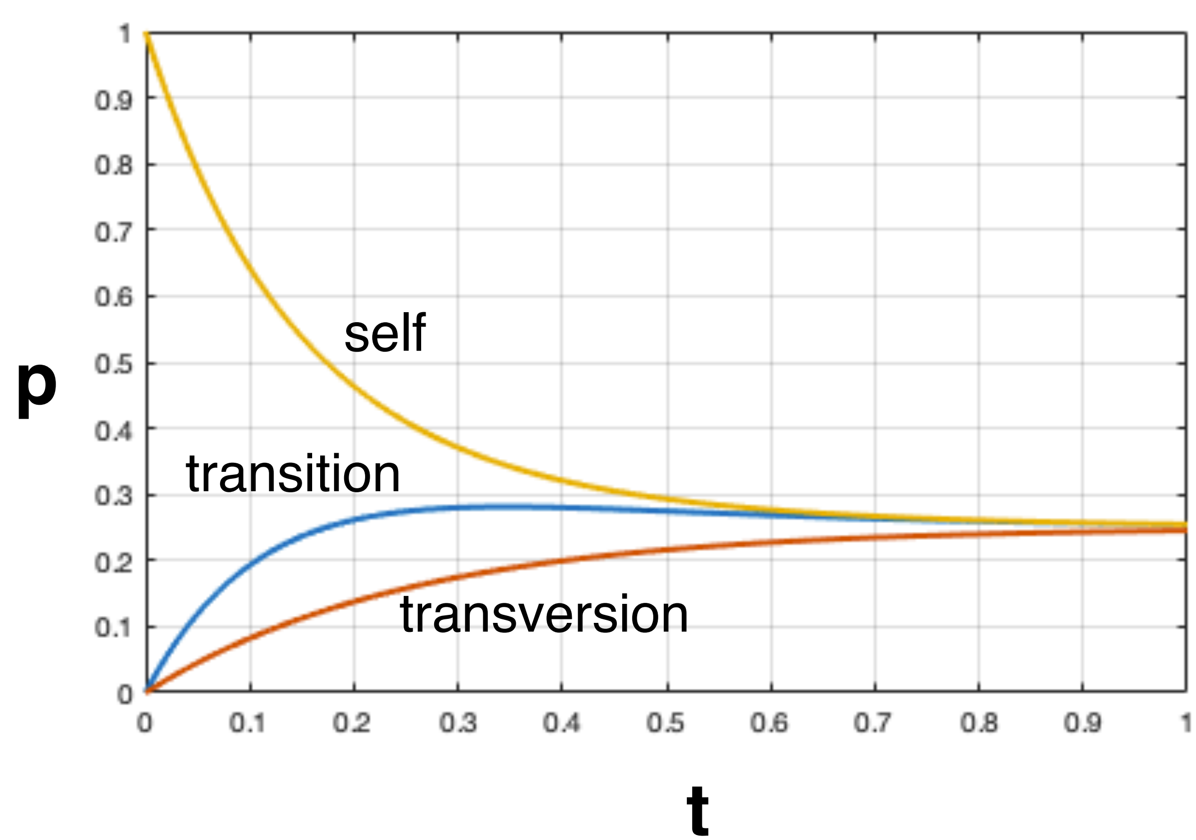
R=

	A	C	G	T
A	-2β-α	β	α	β
C	β	-2β-α	β	α
G	α	β	-2β-α	β
T	β	α	β	-2β-α

Purines: A $\Longleftrightarrow$ G

Pyrimidine C $\Longleftrightarrow$ T

Transitions **α** > Transversions **β**



- מטריצות המעברים לזמן t:

$$P(t)_{i,j} = \begin{cases} 1-2s-u & i=j \\ u=\frac{1}{4}(1+e^{-4\beta t}-2e^{-2(\alpha+\beta)t}) & \text{transition} \\ s=\frac{1}{4}(1-e^{-4\beta t}) & \text{transversion} \end{cases}$$

דוגמא

```
>> R=-4*eye(4)+ones(4);
```

R	A	C	G	T
A	-3	1	1	1
C	1	-3	1	1
G	1	1	-3	1
T	1	1	1	-3

דוגמא

```
>> R=-4*eye(4)+ones(4);
>> T=[0:.01:1];
>> for i=1:length(T), P(:,:,i)=expm(T(i)*R); end;
```

R	A	C	G	T
A	-3	1	1	1
C	1	-3	1	1
G	1	1	-3	1
T	1	1	1	-3

0	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

0.1	A	C	G	T
A	0.75	0.08	0.08	0.08
C	0.08	0.75	0.08	0.08
G	0.08	0.08	0.75	0.08
T	0.08	0.08	0.08	0.75

0.2	A	C	G	T
A	0.59	0.14	0.14	0.14
C	0.14	0.59	0.14	0.14
G	0.14	0.14	0.59	0.14
T	0.14	0.14	0.14	0.59

1	A	C	G	T
A	0.26	0.25	0.25	0.25
C	0.25	0.26	0.25	0.25
G	0.25	0.25	0.26	0.25
T	0.25	0.25	0.25	0.26

דוגמא

```
>> R=-4*eye(4)+ones(4);
>> T=[0:.01:1];
>> for i=1:length(T), P(:,:,i)=expm(T(i)*R); end;
>> plot(T,squeeze(P(1,1,:)),T,squeeze(P(1,2,:)))
```

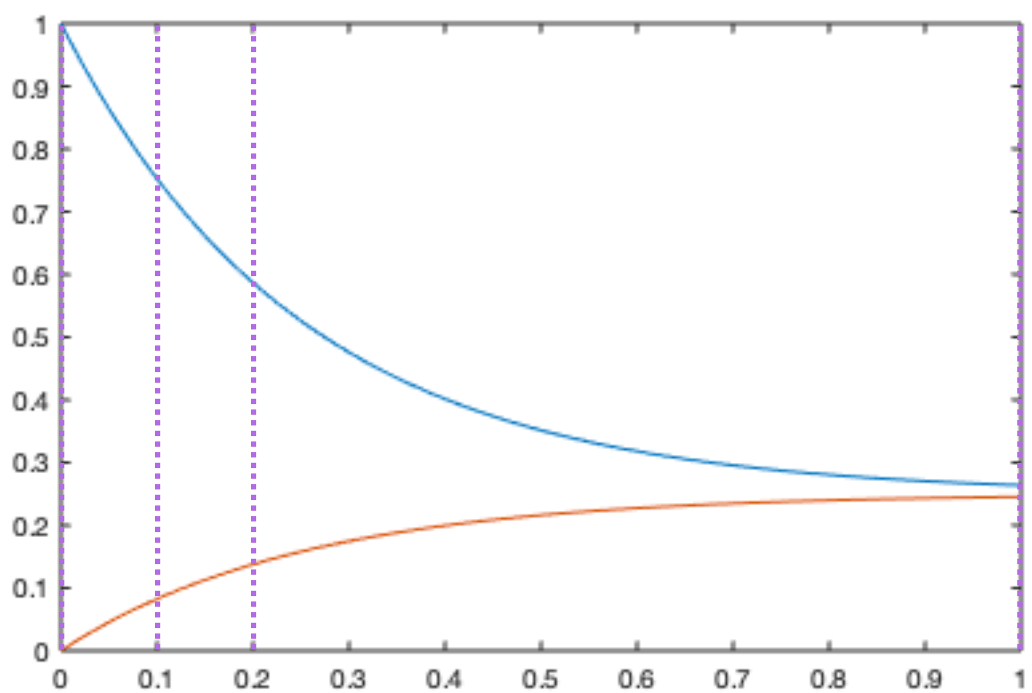
R	A	C	G	T
A	-3	1	1	1
C	1	-3	1	1
G	1	1	-3	1
T	1	1	1	-3

0	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

0.1	A	C	G	T
A	0.75	0.08	0.08	0.08
C	0.08	0.75	0.08	0.08
G	0.08	0.08	0.75	0.08
T	0.08	0.08	0.08	0.75

0.2	A	C	G	T
A	0.59	0.14	0.14	0.14
C	0.14	0.59	0.14	0.14
G	0.14	0.14	0.59	0.14
T	0.14	0.14	0.14	0.59

1	A	C	G	T
A	0.26	0.25	0.25	0.25
C	0.25	0.26	0.25	0.25
G	0.25	0.25	0.26	0.25
T	0.25	0.25	0.25	0.26



דוגמא

R	A	C	G	T
A	-3	1	1	1
C	1	-3	1	1
G	1	1	-3	1
T	1	1	1	-3

```
>> R=-4*eye(4)+ones(4);
>> T=[0:.01:1];
>> for i=1:length(T), P(:,:,i)=expm(T(i)*R); end;
>> plot(T,squeeze(P(1,1,:)),T,squeeze(P(1,2,:)))

>> p=ones(1,4)/4;
```

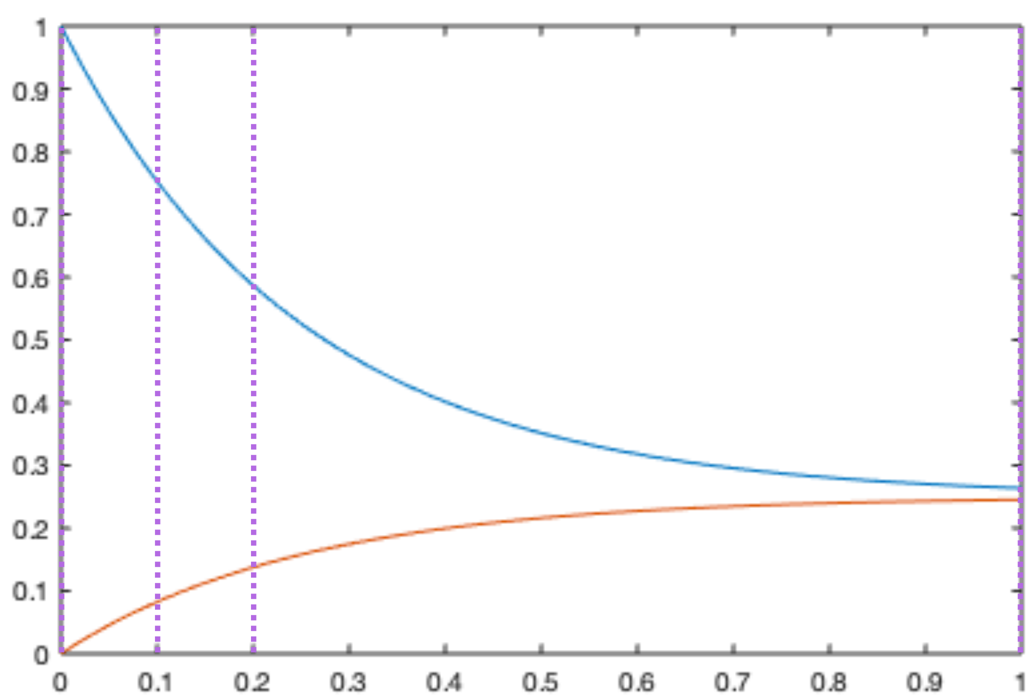
$$\pi = \lim_{t \to \infty} P(t)$$

0	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

0.1	A	C	G	T
A	0.75	0.08	0.08	0.08
C	0.08	0.75	0.08	0.08
G	0.08	0.08	0.75	0.08
T	0.08	0.08	0.08	0.75

0.2	A	C	G	T
A	0.59	0.14	0.14	0.14
C	0.14	0.59	0.14	0.14
G	0.14	0.14	0.59	0.14
T	0.14	0.14	0.14	0.59

1	A	C	G	T
A	0.26	0.25	0.25	0.25
C	0.25	0.26	0.25	0.25
G	0.25	0.25	0.26	0.25
T	0.25	0.25	0.25	0.26



π	A	C	G	T
	0.25	0.25	0.25	0.25

דוגמא

```
>> A=‘AAAAAAAAAA';
>> B=‘AAATAACGAA’;
>> sum(log2(p(nt2int(A))))
```

$$a \xrightarrow{t} b$$

$$\log_2 P(A) = \sum \log_2 P(A_i)$$

$$\log_2 P(B\,|\,A) = \sum \log_2 P_t(B_i\,|\,A_i)$$

R	A	C	G	T
A	-3	1	1	1
C	1	-3	1	1
G	1	1	-3	1
T	1	1	1	-3

דוגמא

```
>> A='AAAAAAAAAA';
>> B='AAATAACGAA';
>> sum(log2(p(nt2int(A))))
```

$a \xrightarrow{t} b$

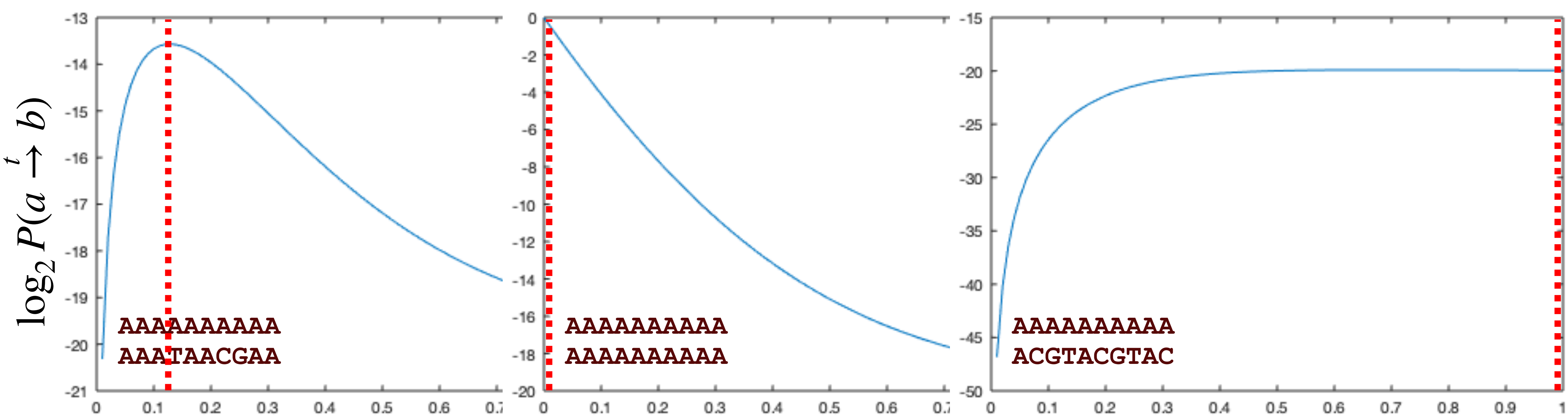
R	A	C	G	T
A	-3	1	1	1
C	1	-3	1	1
G	1	1	-3	1
T	1	1	1	-3

$$\log_2 P(A) = \sum \log_2 P(A_i)$$

$$\log_2 P(B\,|\,A) = \sum \log_2 P_t(B_i\,|\,A_i)$$

```
>> for i=1:length(T),
      Pt=P(:,:,i);
      LL(i)=sum(log2(Pt(sub2ind([4,4],nt2int(A)',nt2int(B)'))));
   end;
>> plot(T,LL);
```

איזה זמן ממקסם את הנראות?



התפלגות סטציונרית

$$\pi = \lim_{t\to\infty} P(t)$$

	A	C	G	T
A	— π —			
C	— π —			
G	— π —			
T	— π —			

- האם תמיד קיימת?
- **תכונת הארגודיות** $\forall_{a,b,t>0}\ P(a \xrightarrow{t} b) > 0$

[רכיב קשירות אחד, ללא בורות או מחזוריות]

- אם R מגדירה תהליך מרקובי ארגודי, אזי

$$\lim_{t\to\infty} P(t) = \begin{bmatrix} -\pi- \\ -\pi- \\ \cdots \\ -\pi- \end{bmatrix} \qquad \forall_{a,b}\ \lim_{t\to\infty} P(a \xrightarrow{t} b) = \pi_b$$

- ואז π מוגדרת בתור ההתפלגות הסטציונרית של P[t]

$$\vec{\pi} \cdot \left[P(t)\right] = \vec{\pi}$$

- כלומר π הוא ו"ע עם ע"ע 1 עבור $P[t]^T$

וקטור ה-$\mathbb{1}$ הוא ו"ע של P[t].

משפט פרון-פרובניוס.

התפלגות סטציונרית

- אם R מגדירה תהליך מרקובי ארגודי, אזי

$$\lim_{t\to\infty} P(t) = \begin{bmatrix} -\pi- \\ -\pi- \\ \cdots \\ -\pi- \end{bmatrix} \qquad \forall_{a,b} \lim_{t\to\infty} P(a \xrightarrow{t} b) = \pi_b$$

- ואז π מוגדרת בתור ההתפלגות הסטציונרית של P[t]

$$\vec{\pi} \cdot \left[P(t)\right] = \vec{\pi}$$

$$\pi_b \triangleq P(\infty)_{a,b} = P(\infty + t)_{a,b} = \sum_c P(\infty)_{a,c} \cdot P(t)_{c,b} = \sum_c \pi_c \cdot P(t)_{c,b}$$

π	A	C	G	T

X

	A	C	G	T
A				
C				
G				
T				

=

π	A	C	G	T

- כלומר π הוא ו"ע עם ע"ע 1 עבור $P[t]^T$

התפלגות סטציונרית

- ועבור ε יתקיים גם:

$$\vec{\pi} = \left[P(t)\right]^T \vec{\pi} \quad = [I + \epsilon R]^T \vec{\pi} \;= \vec{\pi} + \epsilon R^T \vec{\pi} \qquad\qquad \vec{0} = R^T \vec{\pi}$$

- כלומר π הוא נק' שבת עבור P, ונק' אפס עבור R

הבה נגרילה מטריצות קצב

```
>> R=rand(4,4); I=find(eye(4)); R(I)=0;
>> R(I)=-sum(R,2);
>> T=[0:.01:4];
>> for i=1:length(T), P(:,:,i)=expm(T(i)*R); end;
>> plot(T,reshape(P,16,[])'),; P(:,:,end)
```