Transition probability of the HKY model

2005-10-04 shimo

■ Some identities and transformations which will be used later

```
In[1]:= Off[General::spell]
```

In[2]:= **id1** = $\{\pi_A + \pi_G + \pi_C + \pi_T == 1,\ \pi_Y + \pi_R == 1\}$;

In[3]:= **tr1** = $\{\pi_C + \pi_T \to \pi_Y,\ \pi_A + \pi_G \to \pi_R\}$;

■ The instantaneous transition matrix

In[4]:= **hkyA** =
$$\begin{pmatrix} -\beta\pi_C - \alpha\pi_G - \beta\pi_T & \alpha\pi_A & \beta\pi_A & \beta\pi_A \\ \alpha\pi_G & -\alpha\pi_A - \beta\pi_C - \beta\pi_T & \beta\pi_G & \beta\pi_G \\ \beta\pi_C & \beta\pi_C & -\beta\pi_A - \beta\pi_G - \alpha\pi_T & \alpha\pi_C \\ \beta\pi_T & \beta\pi_T & \alpha\pi_T & -\beta\pi_A - \alpha\pi_C - \beta\pi_G \end{pmatrix};$$

which satisfies

```
In[5]:= {1, 1, 1, 1}.hkyA

Out[5]= {0, 0, 0, 0}
```

■ The expected rate of change

The expected rate of transitions are obtained as

In[6]:= **hkyE = hkyA.DiagonalMatrix**[$\{\pi_A, \pi_G, \pi_C, \pi_T\}$]; **hkyE // MatrixForm**

Out[6]//MatrixForm=
$$\begin{pmatrix} \pi_A(-\beta\pi_C - \alpha\pi_G - \beta\pi_T) & \alpha\pi_A\pi_G & \beta\pi_A\pi_C & \beta\pi_A\pi_T \\ \alpha\pi_A\pi_G & \pi_G(-\alpha\pi_A - \beta\pi_C - \beta\pi_T) & \beta\pi_C\pi_G & \beta\pi_G\pi_T \\ \beta\pi_A\pi_C & \beta\pi_C\pi_G & \pi_C(-\beta\pi_A - \beta\pi_G - \alpha\pi_T) & \alpha\pi_C\pi_T \\ \beta\pi_A\pi_T & \beta\pi_G\pi_T & \alpha\pi_C\pi_T & (-\beta\pi_A - \alpha\pi_C - \beta\pi_G)\pi_T \end{pmatrix}$$

```
In[7]:= Simplify[{1, 1, 1, 1}.hkyE, id1]

Out[7]= {0, 0, 0, 0}
```

The expected rate of total change is

```
In[8]:= hkyT = Expand[-Sum[hkyE[[i, i]], {i, 1, 4}]]
```

Out[8]= $2\beta\pi_A\pi_C + 2\alpha\pi_A\pi_G + 2\beta\pi_C\pi_G + 2\beta\pi_A\pi_T + 2\alpha\pi_C\pi_T + 2\beta\pi_G\pi_T$

which may be simplifed as

```
In[9]:= hkyT = Collect[hkyT, {α, β}, Simplify] /. tr1
```

Out[9]= $2\,\alpha\,(\pi_A\,\pi_G + \pi_C\,\pi_T) + 2\,\beta\,\pi_R\,\pi_Y$

Let r = α/β. Then, α and β are defined by giving a constaint that hkyT=1.

```
In[10]:= Solve[(hkyT /. {α → r β}) == 1, β]
```

Out[10]= $\left\{\left\{\beta \to \frac{1}{2\,r\,(\pi_A\,\pi_G + \pi_C\,\pi_T) + 2\,\pi_R\,\pi_Y}\right\}\right\}$

■ The eigenvalues

```
In[11]:= foo = Eigensystem[hkyA]
```

Out[11]= $\left\{\{0, -\beta\,\pi_A - \alpha\,\pi_C - \beta\,\pi_G - \alpha\,\pi_T, -\alpha\,\pi_A - \beta\,\pi_C - \alpha\,\pi_G - \beta\,\pi_T, -\beta\,\pi_A - \beta\,\pi_C - \beta\,\pi_G - \beta\,\pi_T\},\right.$
$\left\{\left\{\frac{\pi_A}{\pi_T}, \frac{\pi_G}{\pi_T}, \frac{\pi_C}{\pi_T}, 1\right\}, \{0, 0, -1, 1\},\right.$
$\left.\left.\{-1, 1, 0, 0\}, \left\{-\frac{\pi_A\,\pi_C + \pi_A\,\pi_T}{(\pi_A + \pi_G)\,\pi_T}, -\frac{\pi_C\,\pi_G + \pi_G\,\pi_T}{\pi_A\,\pi_T + \pi_G\,\pi_T}, \frac{\pi_C}{\pi_T}, 1\right\}\right\}\right\}$

The eigenvalues

```
In[12]:= foo[[1]] // TableForm
```

Out[12]//TableForm=

0
$-\beta\,\pi_A - \alpha\,\pi_C - \beta\,\pi_G - \alpha\,\pi_T$
$-\alpha\,\pi_A - \beta\,\pi_C - \alpha\,\pi_G - \beta\,\pi_T$
$-\beta\,\pi_A - \beta\,\pi_C - \beta\,\pi_G - \beta\,\pi_T$

```
In[13]:= val = FullSimplify[foo[[1]], id1]
```

Out[13]= $\{0, -\beta - (\alpha - \beta)\,(\pi_C + \pi_T), -\alpha + (\alpha - \beta)\,(\pi_C + \pi_T), -\beta\}$

The eigenvectors

```
In[14]:= Transpose[foo[[2]]] // MatrixForm
```

Out[14]//MatrixForm=

$$\begin{pmatrix} \frac{\pi_A}{\pi_T} & 0 & -1 & -\frac{\pi_A\,\pi_C + \pi_A\,\pi_T}{(\pi_A + \pi_G)\,\pi_T} \\ \frac{\pi_G}{\pi_T} & 0 & 1 & -\frac{\pi_C\,\pi_G + \pi_G\,\pi_T}{\pi_A\,\pi_T + \pi_G\,\pi_T} \\ \frac{\pi_C}{\pi_T} & -1 & 0 & \frac{\pi_C}{\pi_T} \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

```
In[15]:= hkyV = Simplify[Transpose[foo[[2]]].DiagonalMatrix[{πT, 1, 1, (πA + πG) πT}], id1];
```

```
In[16]:= hkyV // MatrixForm
```

Out[16]//MatrixForm=

$$\begin{pmatrix} \pi_A & 0 & -1 & -\pi_A\,(\pi_C + \pi_T) \\ \pi_G & 0 & 1 & -\pi_G\,(\pi_C + \pi_T) \\ \pi_C & -1 & 0 & \pi_C\,(\pi_A + \pi_G) \\ \pi_T & 1 & 0 & (\pi_A + \pi_G)\,\pi_T \end{pmatrix}$$

This matrix consists of columns of eigenvectors

■ Check the eigenvectors

Define a diagonal matrix of eigenvalues

```
In[17]:= hkyL = DiagonalMatrix[val];

In[18]:= hkyL // MatrixForm
```

Out[18]//MatrixForm=

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\beta - (\alpha - \beta)\ (\pi_C + \pi_T) & 0 & 0 \\ 0 & 0 & -\alpha + (\alpha - \beta)\ (\pi_C + \pi_T) & 0 \\ 0 & 0 & 0 & -\beta \end{pmatrix}$$

check if A*V = V*L

```
In[19]:= Simplify[hkyA.hkyV - hkyV.hkyL, id1]

Out[19]= {{0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}}
```

■ Inverse matrix of hkyV

```
In[20]:= hkyIV = Simplify[Inverse[hkyV], id1];

In[21]:= hkyIV // MatrixForm
```

Out[21]//MatrixForm=

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & -\frac{\pi_T}{\pi_C+\pi_T} & \frac{\pi_C}{\pi_C+\pi_T} \\ -\frac{\pi_G}{\pi_A+\pi_G} & \frac{\pi_A}{\pi_A+\pi_G} & 0 & 0 \\ \frac{1}{-1+\pi_C+\pi_T} & \frac{1}{-1+\pi_C+\pi_T} & \frac{1}{\pi_C+\pi_T} & \frac{1}{\pi_C+\pi_T} \end{pmatrix}$$

check if V*IV = Identity matrix

```
In[22]:= Simplify[hkyV.hkyIV, id1]

Out[22]= {{1, 0, 0, 0}, {0, 1, 0, 0}, {0, 0, 1, 0}, {0, 0, 0, 1}}
```

check if V*L*IV=A

```
In[23]:= Simplify[hkyV.hkyL.hkyIV - hkyA, id1]

Out[23]= {{0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}}
```

■ Transition probability matrix P(t)=Exp[A t]

Define Exp[L t]

```
In[24]:= hkyExpLt = DiagonalMatrix[Exp[val t]];
```

In[25]:= **hkyExpLt // MatrixForm**

Out[25]//MatrixForm=

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{t\,(-\beta-(\alpha-\beta)\,(\pi_C+\pi_T))} & 0 & 0 \\ 0 & 0 & e^{t\,(-\alpha+(\alpha-\beta)\,(\pi_C+\pi_T))} & 0 \\ 0 & 0 & 0 & e^{-t\,\beta} \end{pmatrix}$$

Define Exp[A t] = Exp[V*(L t)*IV] = V*Exp[L t]*IV

In[26]:= **hkyExpAt = Simplify[hkyV.hkyExpLt.hkyIV , id1];;**

In[27]:= **hkyExpAt // MatrixForm**

Out[27]//MatrixForm=

$$\left(\begin{array}{ccc} \pi_A + \frac{e^{t\,(-\alpha+(\alpha-\beta)\,(\pi_C+\pi_T))}\,\pi_G}{\pi_A+\pi_G} - \frac{e^{-t\,\beta}\,\pi_A\,(\pi_C+\pi_T)}{-1+\pi_C+\pi_T} & \pi_A\left(1 - \frac{e^{t\,(-\alpha+(\alpha-\beta)\,(\pi_C+\pi_T))}}{\pi_A+\pi_G} - \frac{e^{-t\,\beta}\,(\pi_C+\pi_T)}{-1+\pi_C+\pi_T}\right) & (1 - e^{-} \\ \pi_G\left(1 - \frac{e^{t\,(-\alpha+(\alpha-\beta)\,(\pi_C+\pi_T))}}{\pi_A+\pi_G} - \frac{e^{-t\,\beta}\,(\pi_C+\pi_T)}{-1+\pi_C+\pi_T}\right) & \pi_G + \frac{e^{t\,(-\alpha+(\alpha-\beta)\,(\pi_C+\pi_T))}\,\pi_A}{\pi_A+\pi_G} - \frac{e^{-t\,\beta}\,\pi_G\,(\pi_C+\pi_T)}{-1+\pi_C+\pi_T} & (1 - e^{-} \\ (1 - e^{-t\,\beta})\,\pi_C & (1 - e^{-t\,\beta})\,\pi_C & \pi_C + \frac{e^{-t\,\beta}\,\pi_C\,(\pi_A+\pi_G)}{\pi_C+\pi_T} + \\ (1 - e^{-t\,\beta})\,\pi_T & (1 - e^{-t\,\beta})\,\pi_T & -\frac{e^{-t\,\beta}\,\pi_T\,(-1+e^{t\,(-\alpha+\beta)\,(\pi_C+\pi_T)}}{\pi_C} \end{array} \right.$$

■ Look at the each term of the transition probability matrix

■ Prob(C|A,t)

Transition probability from A to C (transversion) is

In[28]:= **Pcat = hkyExpAt[[3, 1]]**

Out[28]= $(1 - e^{-t\,\beta})\,\pi_C$

■ Prob(G|A,t)

Transition probability from A to G (transition) is

In[29]:= **Pgat = hkyExpAt[[2, 1]]**

Out[29]= $\pi_G\left(1 - \frac{e^{t\,(-\alpha+(\alpha-\beta)\,(\pi_C+\pi_T))}}{\pi_A + \pi_G} - \frac{e^{-t\,\beta}\,(\pi_C + \pi_T)}{-1 + \pi_C + \pi_T}\right)$

Try simplification of the above

In[30]:= **Pgat = Collect[Expand[hkyExpAt[[2, 1]] /. tr1 /. $\pi_Y \to 1 - \pi_R$], {π_G, $e^{-t\,\beta}$}, Simplify[#] &]**

Out[30]= $\pi_G\left(1 + e^{-t\,\beta}\left(-1 + \frac{1}{\pi_R}\right) - \frac{e^{-t\,(\beta+(\alpha-\beta)\,\pi_R)}}{\pi_R}\right)$

The transition probability Prob(G|A,t) given in eq.(13.10) at page 203 of Felsenstein (2003) is

In[31]:= **joePgat = Exp[$-\beta$ t] (1 - Exp[$-\alpha_R$ t]) $\frac{\pi_G}{\pi_R}$ + (1 - Exp[$-\beta$ t]) π_G;**

in which α_R for HKY model is defined as

In[32]:= **hkytr1 = {$\alpha_R \to (\alpha - \beta)\,\pi_R$, $\alpha_Y \to (\alpha - \beta)\,\pi_Y$};**

So, the Pgat for HKY model should be

```
In[33]:= joePgat /. hkytr1
```

Out[33]= $\left(1-e^{-t\beta}\right)\pi_G+\frac{e^{-t\beta}\left(1-e^{-t(\alpha-\beta)\pi_R}\right)\pi_G}{\pi_R}$

Check if Pgat = joePgat

```
In[34]:= Simplify[Pgat - (joePgat /. hkytr1)]
```

Out[34]= 0

■ Prob(T|C,t)

Transition probability from C to T (transition) is

```
In[35]:= Ptct = hkyExpAt[[4, 3]]
```

Out[35]= $-\frac{e^{-t\beta}\pi_T\left(-1+e^{t(-\alpha+\beta)(\pi_C+\pi_T)}-\left(-1+e^{t\beta}\right)\pi_C-\left(-1+e^{t\beta}\right)\pi_T\right)}{\pi_C+\pi_T}$

Try simplification of the above

In[36]:= Ptct = Collect[Expand[hkyExpAt[[4, 3]] /. tr1 /. $\pi_C\to\pi_Y-\pi_T$], {π_T, $e^{-t\beta}$}]

Out[36]= $\pi_T\left(1+e^{-t\beta}\left(-1+\frac{1}{\pi_Y}-\frac{e^{t(-\alpha+\beta)\pi_Y}}{\pi_Y}\right)\right)$

Check if Ptct is obtained from Pgat by replacing $\pi_G\to\pi_T$, $\pi_R\to\pi_Y$

In[37]:= Simplify[Ptct - (Pgat /. {$\pi_G\to\pi_T$, $\pi_R\to\pi_Y$})]

Out[37]= 0