Correlated Models of Discrete Charcter Evolution in `RevBayes`

Will Freyman

Department of Ecology, Evolution & Behavior
University of Minnesota

willfreyman@gmail.com
http://willfreyman.org

UC Berkeley Workshop, February 26-27 2018

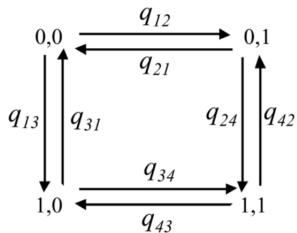# Correlated evolution of discrete characters:

Figure 1: Transitions among the four combinations of states resulting from two binary variables. Subscripts i and j identify the beginning and ending states, respectively, of each transition, where the values 1, 2, 3, and 4 correspond to the state pairs {0,0}, {0,1}, {1,0}, and {1,1}. Thus, q_{12} describes the transition between state {0,0} and state {0,1} over short time interval dt. Dual transitions in which the states of both variables instantaneously change are assumed not to occur.

Correlated evolution of discrete characters:

$$Q_{I,D} = \begin{array}{c} \\ 0,0 \\ 0,1 \\ 1,0 \\ 1,1 \end{array} \begin{array}{cccc} 0,0 & 0,1 & 1,0 & 1,1 \\ \left[\begin{array}{cccc} \ldots & q_{12} & q_{13} & 0 \\ q_{21} & \ldots & 0 & q_{24} \\ q_{31} & 0 & \ldots & q_{34} \\ 0 & q_{42} & q_{43} & \ldots \end{array}\right] \end{array}$$

For a fully independent model: $q_{12} = q_{34}$, $q_{13} = q_{24}$, $q_{21} = q_{43}$, and $q_{31} = q_{42}$.

Image from Pagel and Meade (2006)

Correlated evolution of discrete characters:

We can test the fit of the independent model to a dependent model using either Bayes factors or reversible-jump MCMC.

- ▶ `src/independent_Bayes_factors.Rev`
- ▶ `src/correlated_Bayes_factors.Rev`
- ▶ `src/correlated_reversible_jump.Rev`

Image from Pagel and Meade (2006)

Correlated evolution of discrete characters:

Here we will calculate the marginal likelihood for the most general model of correlated trait evolution described in Pagel and Meade (2006). They notated it as (1,2,3,4,5,6,7,8).

Read in the tree and tip data:

```
tree_obs <- readTrees("data/phylo.tree")[1]
morph_data <- readCharacterData("data/correlated.nex")
```

Correlated evolution of discrete characters:

The character state transition rate matrix

```
q_12 ~ dnExponential(10)
q_13 ~ dnExponential(10)
q_21 ~ dnExponential(10)
q_24 ~ dnExponential(10)
q_31 ~ dnExponential(10)
q_34 ~ dnExponential(10)
q_42 ~ dnExponential(10)
q_43 ~ dnExponential(10)
Q_morph := fnFreeK([q_12, q_13, 0, q_21, 0, q_24, q_31, 0,
    q_34, 0, q_42, q_43], rescaled=FALSE)
```

Correlated evolution of discrete characters:

root frequencies:

```
pi_morph ~ dnDirichlet([1,1,1,1])
```

The phylogenetic CTMC:

```
ctmc_morph ~ dnPhyloCTMC(tree_obs, Q=Q_morph, rootFrequencies=
    pi_morph, nSites=1, type="Standard")
ctmc_morph.clamp(morph_data)
```

Correlated evolution of discrete characters:

Just like before, we need to specify the model, moves, monitors.

```
mymodel = model(ctmc_morph)

moves[1] = mvScale(q_12)
moves[2] = mvScale(q_13)
moves[3] = mvScale(q_21)
moves[4] = mvScale(q_24)
moves[5] = mvScale(q_31)
moves[6] = mvScale(q_34)
moves[7] = mvScale(q_42)
moves[8] = mvScale(q_43)
moves[9] = mvSimplexElementScale(pi_morph)

monitors[1] = mnScreen(printgen=10)
```

Correlated evolution of discrete characters:

Yesterday we ran MCMC analyses. For Bayes factor calculations we need to compute the marginal likelihood of the model so we run a power posterior analysis:

```
pow_p = powerPosterior(mymodel, moves, monitors, "output/
    correlated.out", cats=50)
pow_p.burnin(generations=100, tuningInterval=10)
pow_p.run(generations=100)
```

The calculate the marginal likelihood:

```
ss = steppingStoneSampler(file="output/correlated.out",
                    powerColumnName="power",
                    likelihoodColumnName="likelihood")
ss.marginal()
```

Correlated evolution of discrete characters:

We can similarly set up the most general independent
model described in Pagel and Meade (2006) which they
notated ((1,1,2,2,3,3,4,4). It is exactly the same,
except for the rate matrix:

```
q_12 ~ dnExponential(10)
q_13 ~ dnExponential(10)
q_21 ~ dnExponential(10)
q_24 := q_13
q_31 ~ dnExponential(10)
q_34 := q_12
q_42 := q_31
q_43 := q_21
Q_morph := fnFreeK([q_12, q_13, 0, q_21, 0, q_24, q_31, 0,
    q_34, 0, q_42, q_43], rescaled=FALSE)
```

Don't forget to remove the moves for the rates that are
now deterministic.

Correlated evolution of discrete characters:

The marginal likelihoods I got were -19.3 for the independent model and -17.9 for the correlated model.

Since -17.9 + 19.3 = 1.4, this means the 2*ln*BF = 3.8, which is positive support for the correlated model over the independent model (Kass and Raftery 1995).

However, we only examined one of the many possible different correlated models of character evolution. It is possible that some of the character state transitions are dependent on one another and others are not. We can test for these using reversible-jump MCMC.

Correlated evolution of discrete characters:

We can use reversible-jump MCMC over rate multipliers which will allows every pair of rates described in the indepedent model to no longer be set equal.

This is a simple modification to the independent model we previously specified for the Bayes factor calculation.

Correlated evolution of discrete characters:

Each multiplier will have the value 1.0 or a value drawn from a log uniform distribution varying from 0.02 to 55.

```
m_13 ~ dnReversibleJumpMixture(constantValue=0,
                              baseDistribution=dnUniform(-4, 4),
                              p=0.5)
m_12 ~ dnReversibleJumpMixture(constantValue=0,
                              baseDistribution=dnUniform(-4, 4),
                              p=0.5)
m_31 ~ dnReversibleJumpMixture(constantValue=0,
                              baseDistribution=dnUniform(-4, 4),
                              p=0.5)
m_21 ~ dnReversibleJumpMixture(constantValue=0,
                              baseDistribution=dnUniform(-4, 4),
                              p=0.5)
```

Correlated evolution of discrete characters:

We use deterministic nodes to monitor posterior probabilities of each transition being dependent (correlated) as opposed to independent:

```
cor_13_24 := ifelse(m_13 != 0, 1, 0)
cor_12_34 := ifelse(m_12 != 0, 1, 0)
cor_31_42 := ifelse(m_31 != 0, 1, 0)
cor_21_43 := ifelse(m_21 != 0, 1, 0)
```

Correlated evolution of discrete characters:

When we set up a rate matrix we must include the rate multipliers:

```
q_12 ~ dnExponential(10)
q_13 ~ dnExponential(10)
q_21 ~ dnExponential(10)
q_24 := q_13 * exp(m_13)
q_31 ~ dnExponential(10)
q_34 := q_12 * exp(m_12)
q_42 := q_31 * exp(m_31)
q_43 := q_21 * exp(m_21)
Q_morph := fnFreeK([q_12, q_13, 0, q_21, 0, q_24, q_31, 0,
    q_34, 0, q_42, q_43], rescaled=FALSE)
```

Correlated evolution of discrete characters:

And we must also add MCMC moves for the rate multipliers:

```
moves[6]  = mvSlide(m_13)
moves[7]  = mvSlide(m_12)
moves[8]  = mvSlide(m_31)
moves[9]  = mvSlide(m_21)
moves[10] = mvRJSwitch(m_13)
moves[11] = mvRJSwitch(m_12)
moves[12] = mvRJSwitch(m_31)
moves[13] = mvRJSwitch(m_21)
```

Correlated evolution of discrete characters:

After running the MCMC analysis view the results in Tracer.

We see that the posterior probability of transitions being independent are low.

Once one character goes from $0 \to 1$ the other character also goes from $0 \to 1$ with a very high rate.