Discrete & continuous characters: The threshold model

Discrete & continuous characters: the threshold model

- So far we have discussed continuous & discrete character models separately for estimating ancestral state; and for estimating the evolutionary correlation between characters.
- In recent years a new model has been proposed (or, more accurately, an old model has been revisited) to model the evolutionary covariance between discrete & continuous character on a phylogeny (Felsenstein 2005, 2012; Revell 2013).
- This model is called the *threshold model*.

Review: the Mk model

- The most commonly used model for discrete character evolution on trees is a model called the *Mk* model.
- *M* stands for Markov because the modeled process is a continuous-time Markov chain; and *k* because the model is generalized to include an arbitrary number (*k*) states.
- The central attribute of the *Mk* model is a transition matrix, *Q*, giving the instantaneous transition rates between states.

$$\mathbf{Q} = \begin{bmatrix} -(\alpha + \beta) & \alpha & \beta \\ \gamma & -(\gamma + \delta) & \delta \\ \varepsilon & \zeta & -(\varepsilon + \zeta) \end{bmatrix}$$

$$\mathbf{p}_t = \mathbf{p}_0 \exp(\mathbf{Q}t)$$

Properties of the Mk model

- Because the process is (by definition) memoryless, a character that changes state from 0 -> 1 (or A -> B, etc.) has an indefinitely equal probability of reverting back, B -> A.
- That probability can be large or small (even 0), but it is indefinitely constant.
- In addition, for multistate data a character that has recently changed from A -> B immediately assumes the probability P_{bj} of subsequently changing to state j.
- (Assuming fixation is rapid relative to the scale of time being studied) this could be a reasonable assumption for nucleotide data and some types of morphological characters.
- However, for complex morphological & ecological characters, it may be time to consider another model.

The threshold model

- Wright (1934) proposed a model for discrete characters in which the value of the discrete phenotype is determined by an underlying , unobserved continuous character called 'liability.'
- If liability crossed a fixed threshold value, the character changed state.



Sewall Wright (1889-1988)

What is liability?

- Liability is by definition unobserved or unmeasured.
- It could be a superficially invisible (but theoretically measurable) trait such as circulating blood hormone – for instance.
- I also argue that liability could be a proxy for the complex, multilocus genetic changes that are likely to underlie a shift in a discretely measured ecological trait (Revell 2013).

The threshold model

- In spite of the long history of this model in quantitative genetics, Felsenstein (2005, 2012) was the first to apply it to comparative biology.
- He developed an approach to estimate the evolutionary correlation between discrete characters, or between discrete and continuous traits, using the threshold model.



Joseph Felsenstein

- In that case, the correlation is merely the correlation of liabilities.
- Subsequently I (Revell 2013) proposed using the threshold model for ancestral state reconstruction.

Properties of the threshold model

- The threshold model is inherently ordered.
- Although we can use a Markov process to model the evolution of liability (e.g., Brownian motion), discrete character evolution under the threshold model is not memoryless.
- This is because if a character changed state recently from A -> B, it is much more likely to change back immediately (when near the threshold) than far in the future.
- The model also provides a natural framework for withinspecies polymorphism (although this is not implemented so far).

Properties of the threshold model



- Simulating under the threshold model is trivial.
- We just simulate liability up the tree under our continuous character model (say, Brownian motion); and then we translate our simulated liabilities to the discrete threshold character.



1. Simulate liability up the branches of the tree under our continuous character model.



2. Apply the thresholds to translate tip & node states to the discrete character.



2. Apply the thresholds to translate tip & node states to the discrete character.



3. Project the implied states back onto the nodes of the tree.

Estimating ancestral states under the threshold model

- Fitting the threshold model to discrete character data is distinctly more difficult.
- This is because computing the probability of a character pattern would involve calculating a bunch of integrals of the multivariate normal distribution that we can't compute (and this ignores the positions of the thresholds).
- My solution is to sample the tip & node liabilities, and the relative positions of the thresholds, from their joint posterior probability distribution using Bayesian MCMC under the Metropolis-Hastings algorithm.

Estimating ancestral states under the threshold model

 While computing the likelihood of our discrete character data would be difficult; computing the likelihood (& thus posterior odds ratio) of a set of tip & node liabilities given the tip data & tree is easy:

$$likelihood \begin{pmatrix} \text{tip liabilities,} \\ \text{ancestral states,} \\ \& \text{thresholds} \end{pmatrix} = P \begin{pmatrix} \text{liabilities} \\ \text{tree & model} \end{pmatrix} \times \begin{cases} 1.0 \text{ if liabilities correct} \\ 0.0 \text{ otherwise} \end{cases}$$

$$l(\mathbf{x}, a_0, \mathbf{a}, \mathbf{\tau} \mid \mathbf{y}, \mathbf{C}) = \frac{\exp\left[-\frac{1}{2}\left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)'\mathbf{C}^{-1}\left([\mathbf{x}, \mathbf{a}] - a_0\mathbf{1}\right)\right]}{(2\pi)^{(n+i-1)/2} |\mathbf{C}|^{1/2}} \times \begin{cases} 1 & \text{if } f(\mathbf{x}, \mathbf{\tau}) = \mathbf{y} \\ 0 & \text{if } f(\mathbf{x}, \mathbf{\tau}) \neq \mathbf{y} \end{cases}$$

Estimating ancestral states under the threshold model

- Something that you might observe about this expression is that there is no rate of liability evolution, σ^2 .
- This is because liability is scaleless, thus we can fix the position of the threshold(s) and estimate σ^2 ; or fix σ^2 , and estimate the positions of the thresholds but not both.
- What value we fix σ^2 to is inconsequential, because σ^2 cancels from the numerator & denominator of the posterior odds ratio during MCMC.

Estimating the evolutionary correlation under the threshold model

- In addition to ancestral states, we can also use the threshold model to estimate the evolutionary correlation between discrete traits or between discrete & continuous characters.
- In this case the likelihood expression is as follows:

$$likelihood \begin{pmatrix} \text{tip liabilities,} \\ \text{tip trait values,} \\ \text{covariances,} \\ & \text{thresholds *} \end{pmatrix} = P \begin{pmatrix} \text{liabilities, tip values,} \\ & \text{correlation} | \text{tree } \& \\ & \text{model} \end{pmatrix} \times \begin{cases} 1.0 \text{ if correct} \\ 0.0 \text{ otherwise} \end{cases}$$

• We can't maximize the likelihood, but we can sample liabilities & covariances from their joint posterior probability distribution.