

Threshold models



Friday Harbor Laboratories, 9 June 2017

Joe Felsenstein

Evolutionary Quantitative Genetics Workshop

How to use phylogenies for morphology or behavior?

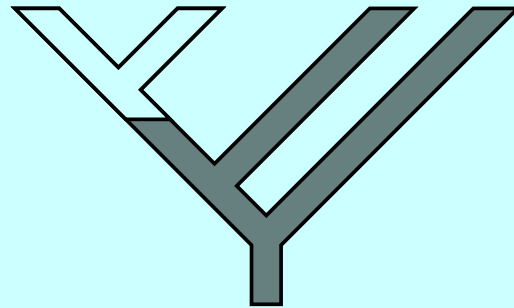
- Currently, most people use parsimony methods for these
- They code the characters discretely, which lead to what used to be called “the character coding problem”
- They have no statistical model and little way of knowing how uncertain are their inferences (except for some bootstrapping or jackknifing which assumes characters change independently)
- But is discrete coding necessary? No.
- Are there statistical models for these discrete phenotypes? Yes.

Current methods for statistical treatment of 0/1 characters

Pagel (1994) and Lewis (2001) treat such data with



Mark Pagel



$$0 \xrightleftharpoons[\beta dt]{\alpha dt} 1$$



Paul Lewis

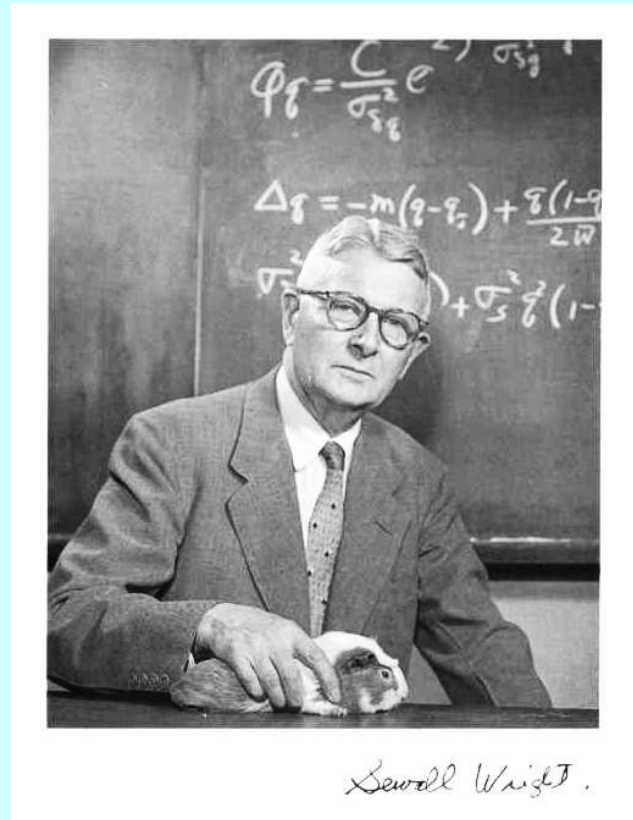
Pagel allows inference of whether changes in two characters are correlated, on a known tree.

Lewis uses a k -state “ M_k ” model to infer the tree, but does not allow for correlations among characters.

Neither takes into account contributions to a 0/1 character from multiple underlying loci.

The threshold model

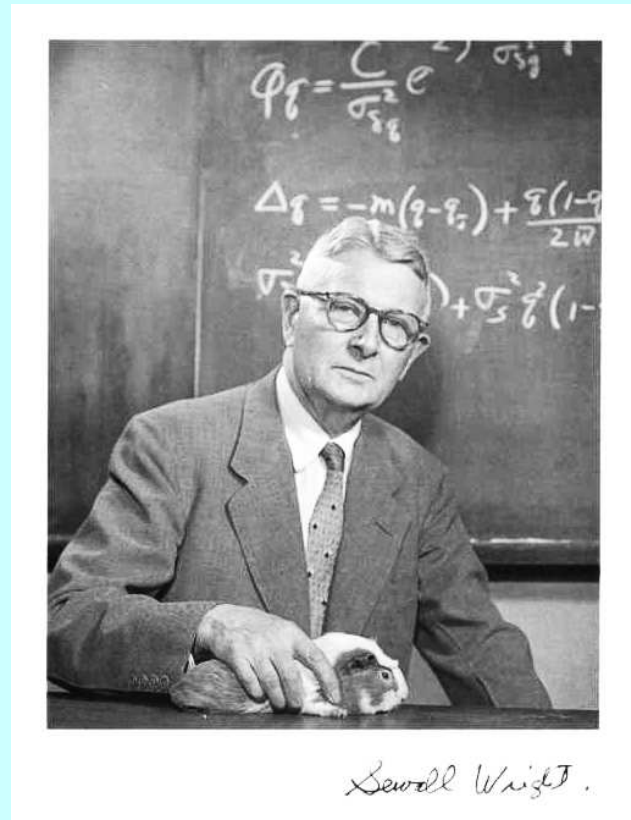
A relevant model was invented in 1934 by



Sewall Wright (1889-1988)
shown here in 1955

The threshold model

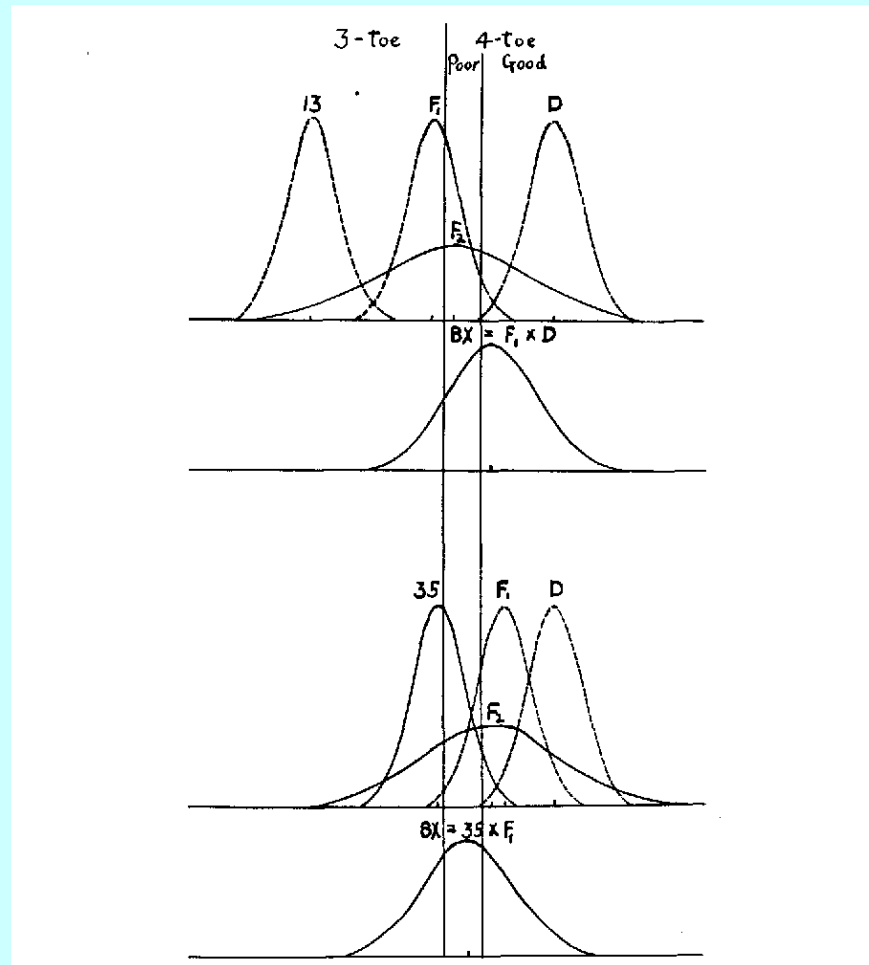
A relevant model was invented in 1934 by



Sewall Wright (1889-1988)
shown here in 1955

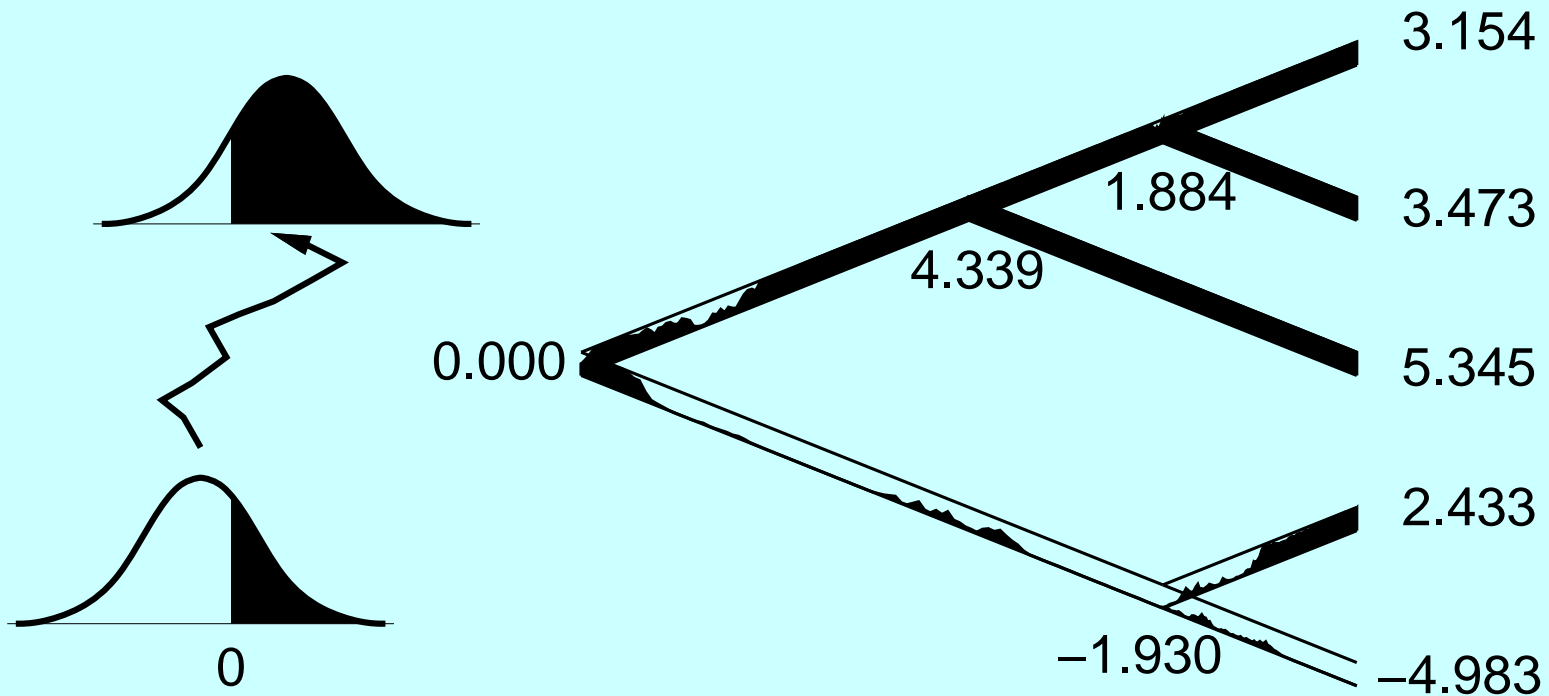
(The story goes that he then absent-mindedly started to erase the board with the guinea pig)

The threshold model, applied

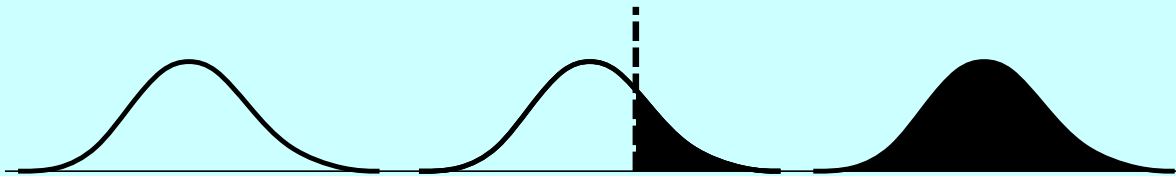


Sewall Wright (1934), guinea pig digit number
(from Wright's follow-up 1934 second paper)

The threshold model on a tree



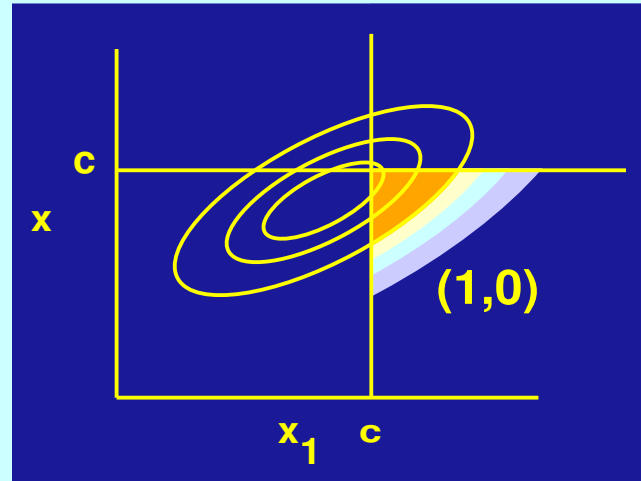
Biologically realistic properties



- Soon after the population crosses the threshold, crossing back is more likely than it is later on (when the population may have gotten further from the threshold). The M_k model does not predict this.
- The model predicts a phase of polymorphism (the M_k model makes no such prediction)
- The model very straightforwardly incorporates covariation between characters. For the M_k model, one needs to introduce vast numbers of new parameters.

Computing the likelihood

With two species, one character:



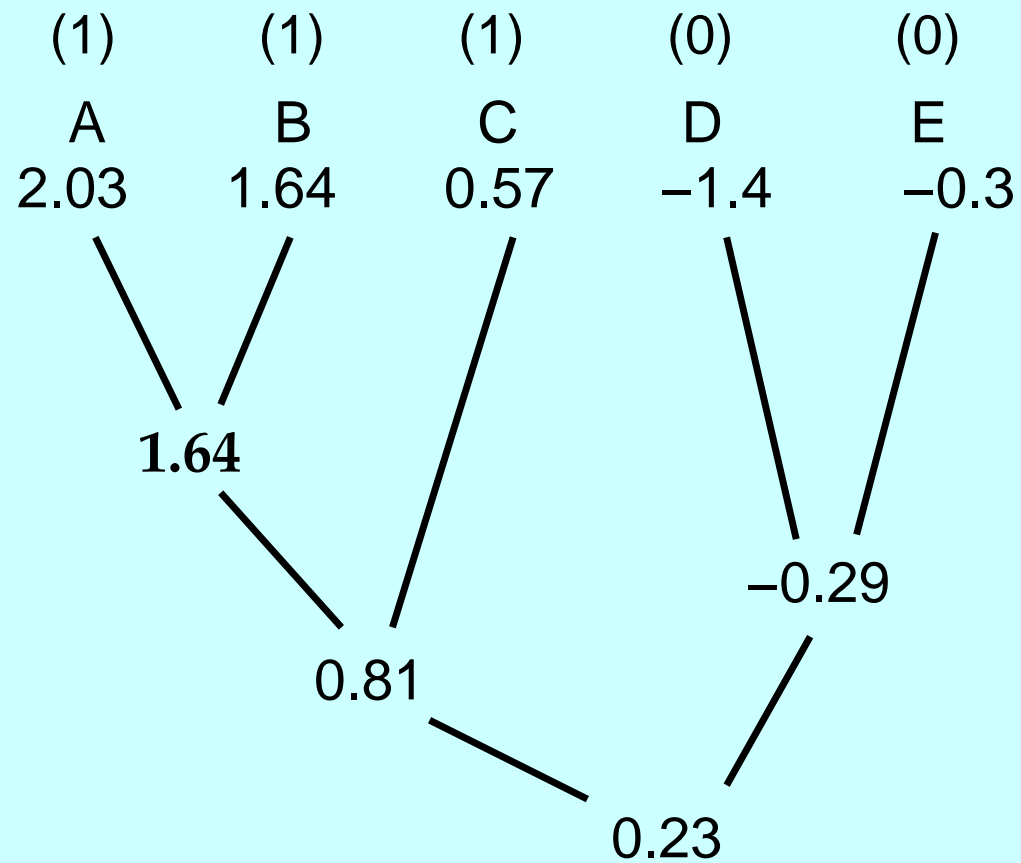
Disadvantages:

Quite hard to compute likelihoods: need to compute area in a corner of a correlated multivariate normal distribution.

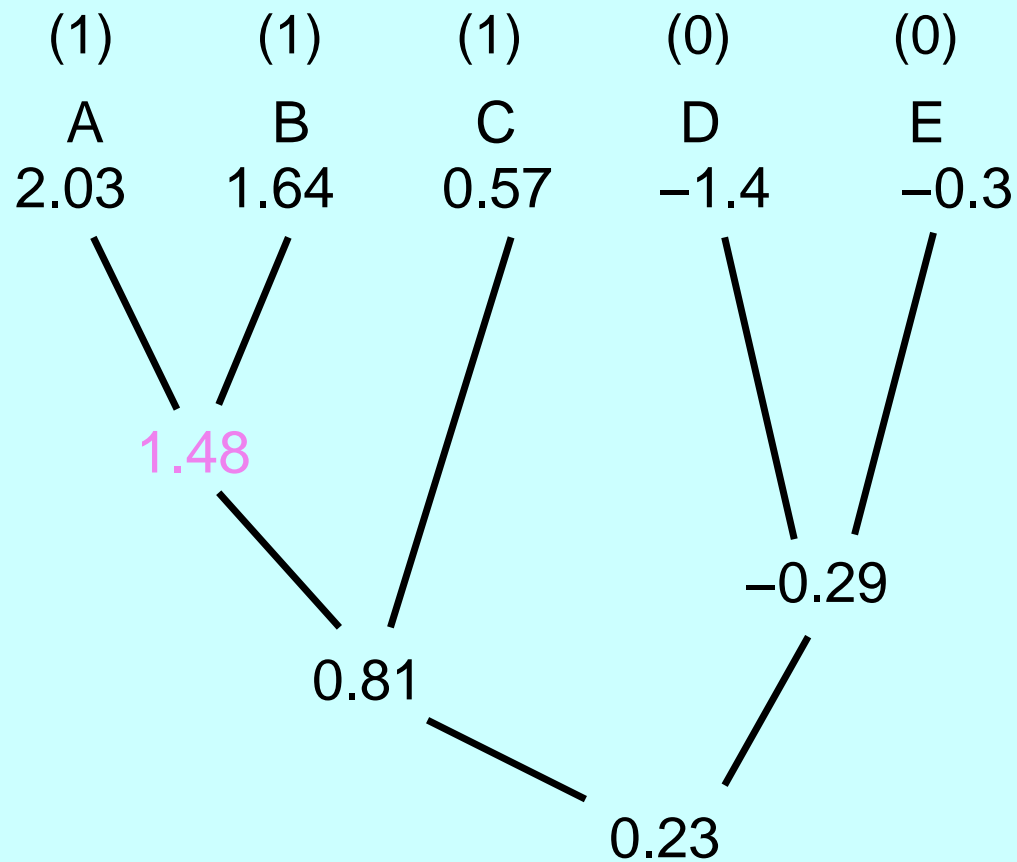
With 5 species, one character:

$$\begin{aligned} L &= \text{Prob}(1, 1, 0, 1, 1) \\ &= \int_0^\infty \int_0^\infty \int_{-\infty}^0 \int_0^\infty \int_0^\infty \varphi(x_1, x_2, x_3, x_4, x_5 \mid \text{Tree}) dx_1 dx_2 dx_3 dx_4 dx_5 \end{aligned}$$

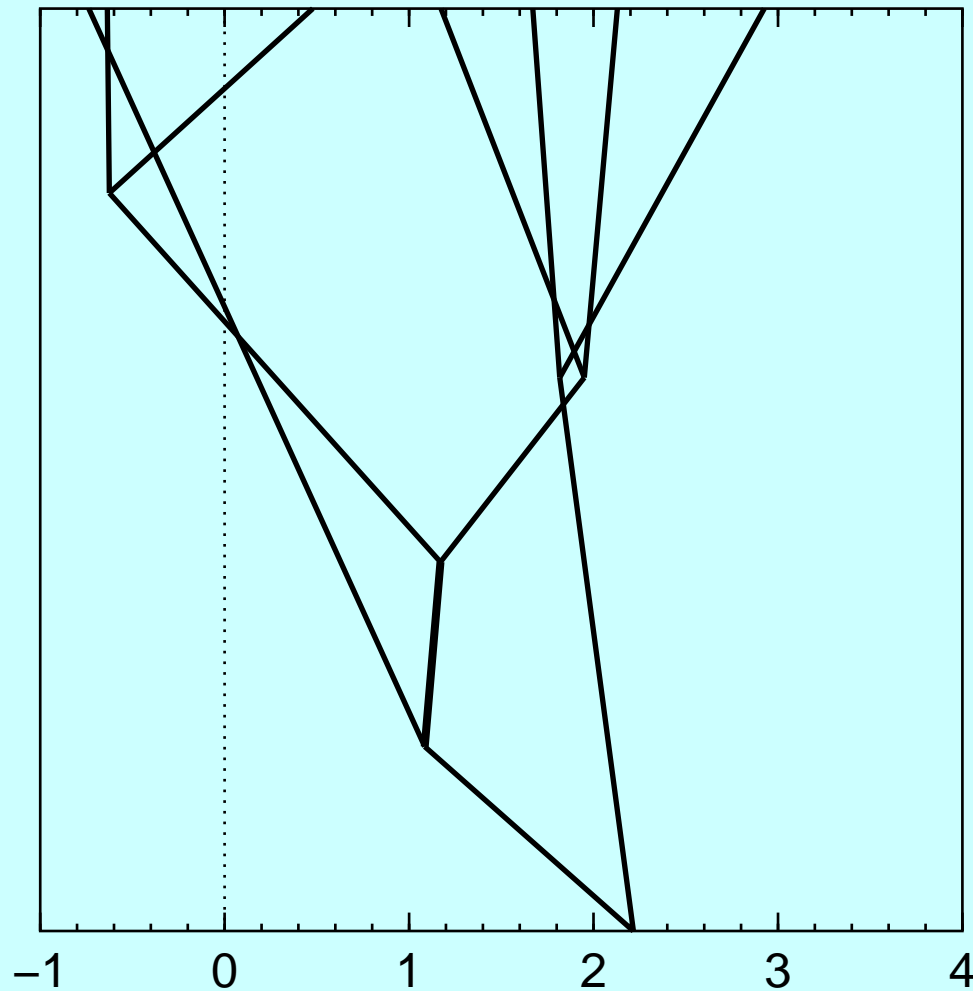
MCMC on liabilities



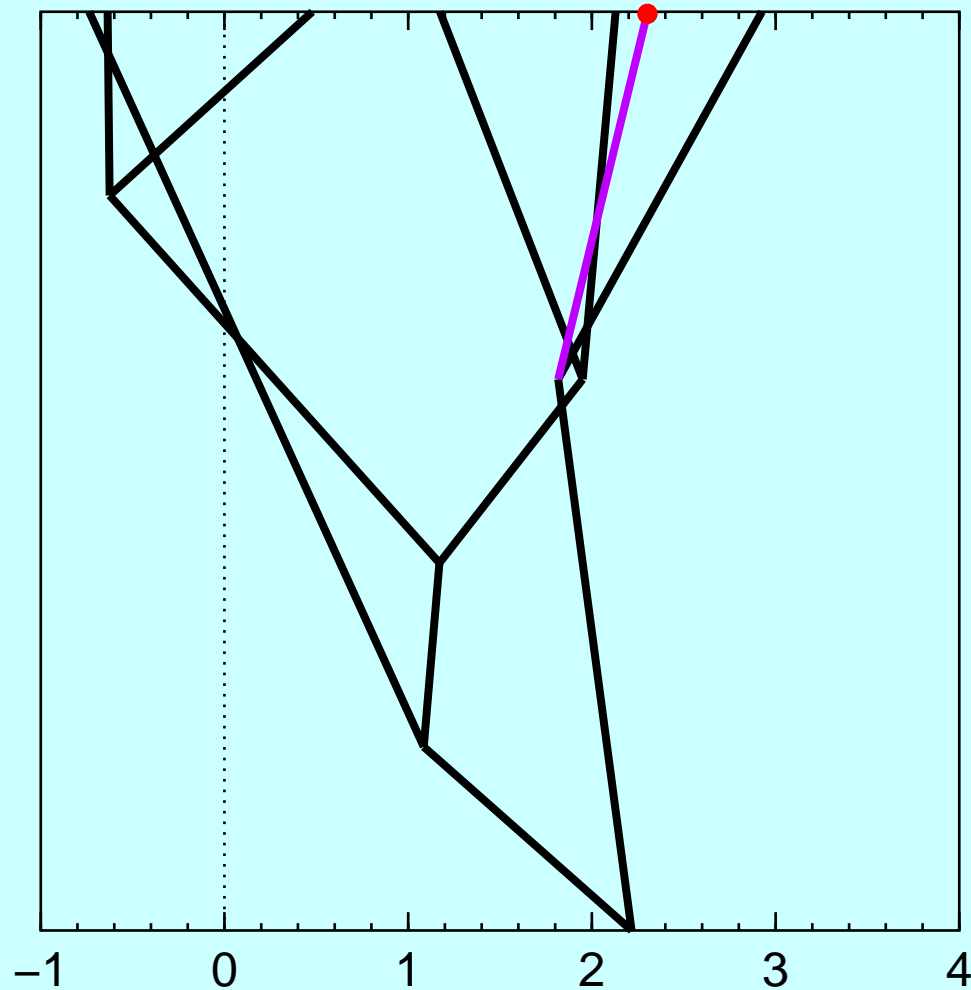
MCMC on liabilities: result of Gibbs sampling



An example with 7 species and one character

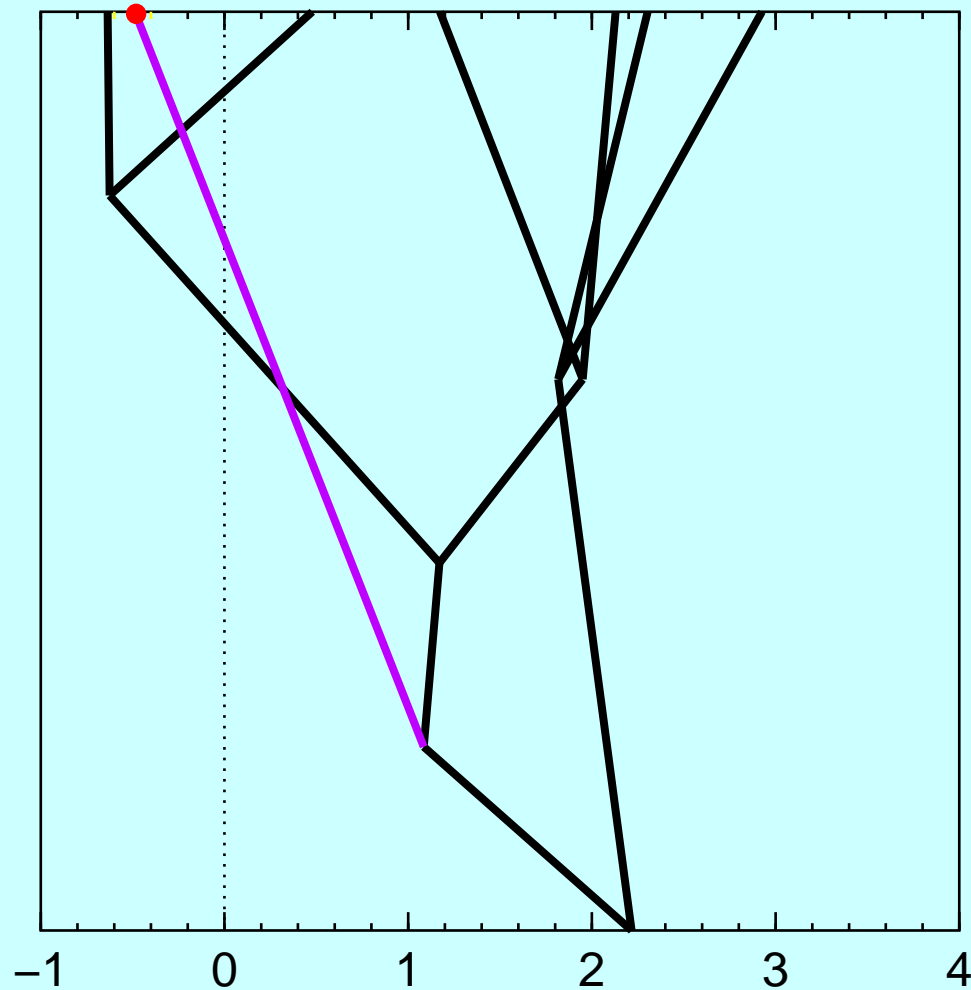


An example with 7 species and one character



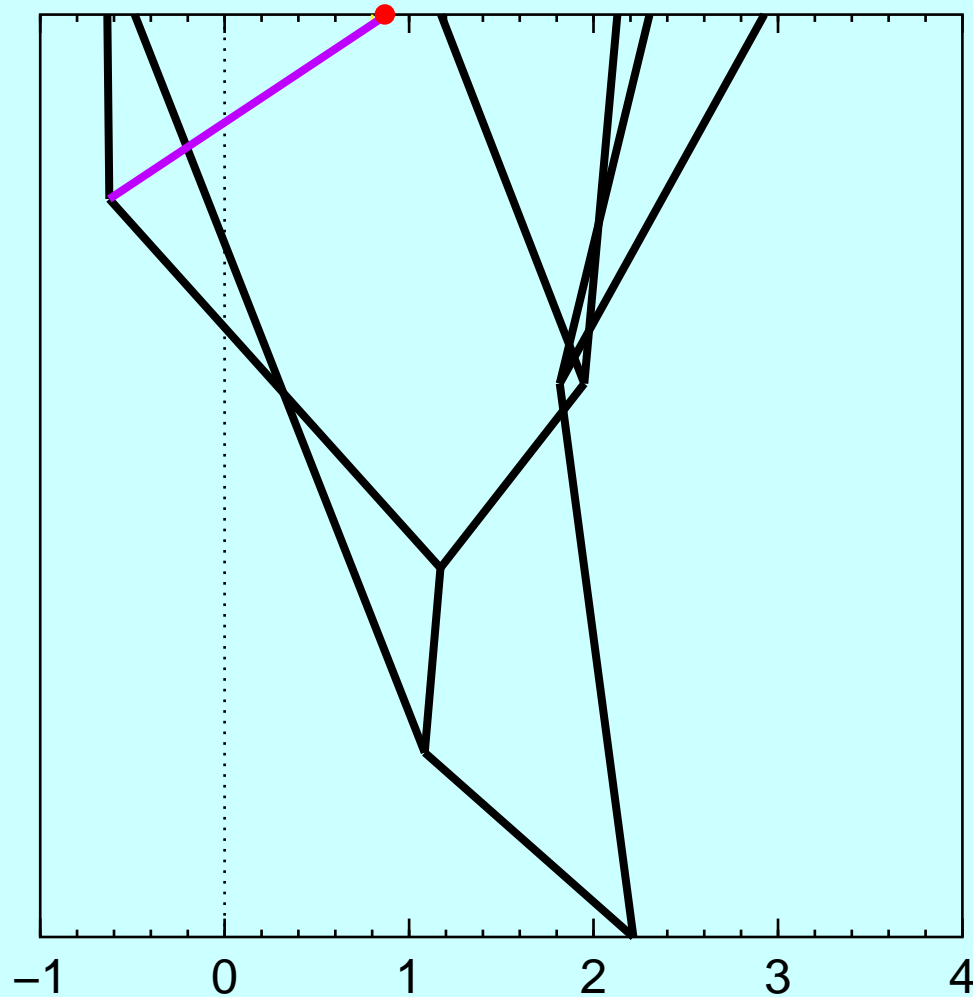
Updating a tip node's liability

An example with 7 species and one character



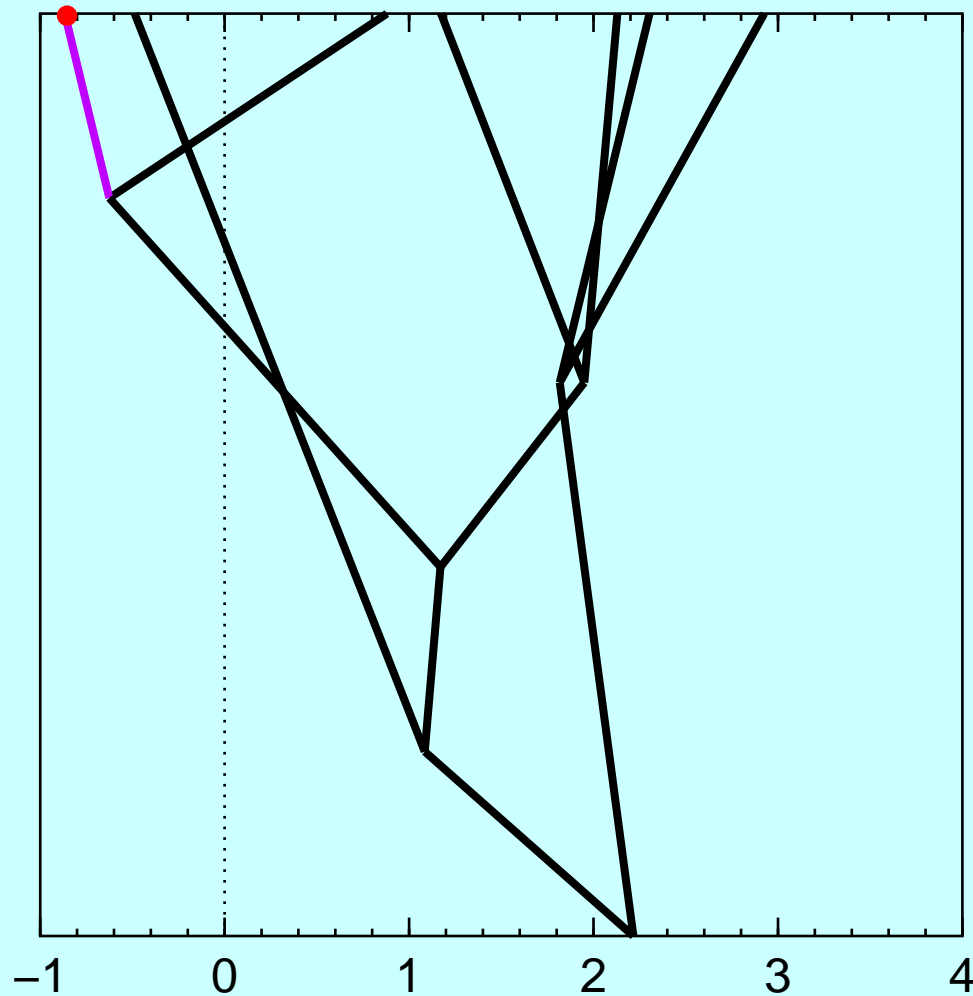
Updating a tip node's liability

An example with 7 species and one character



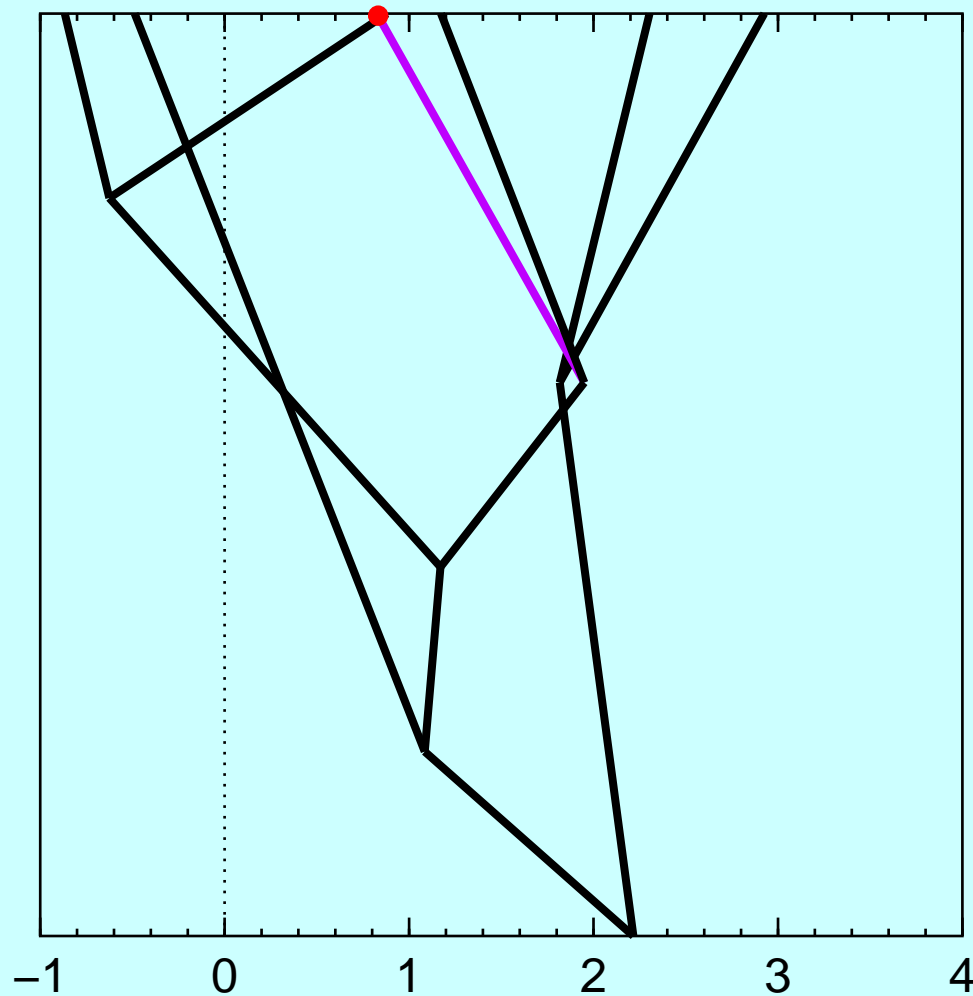
Updating a tip node's liability

An example with 7 species and one character



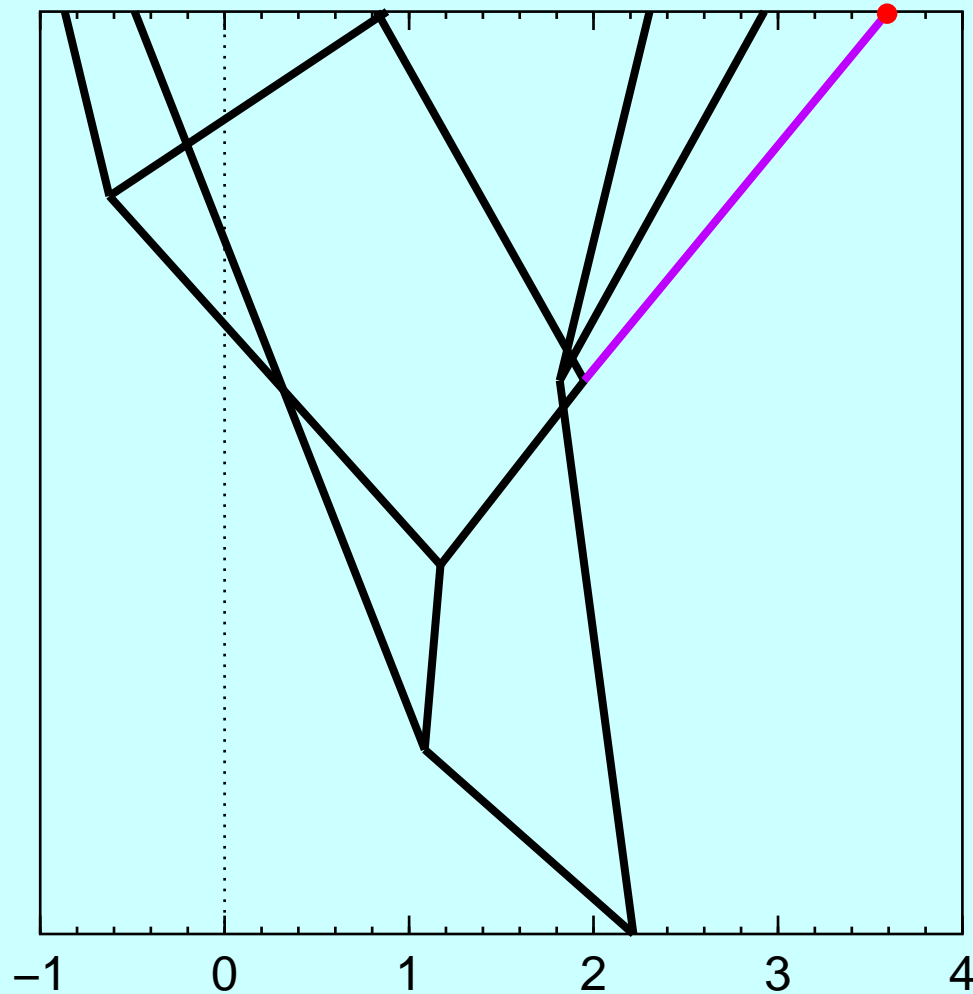
Updating a tip node's liability

An example with 7 species and one character



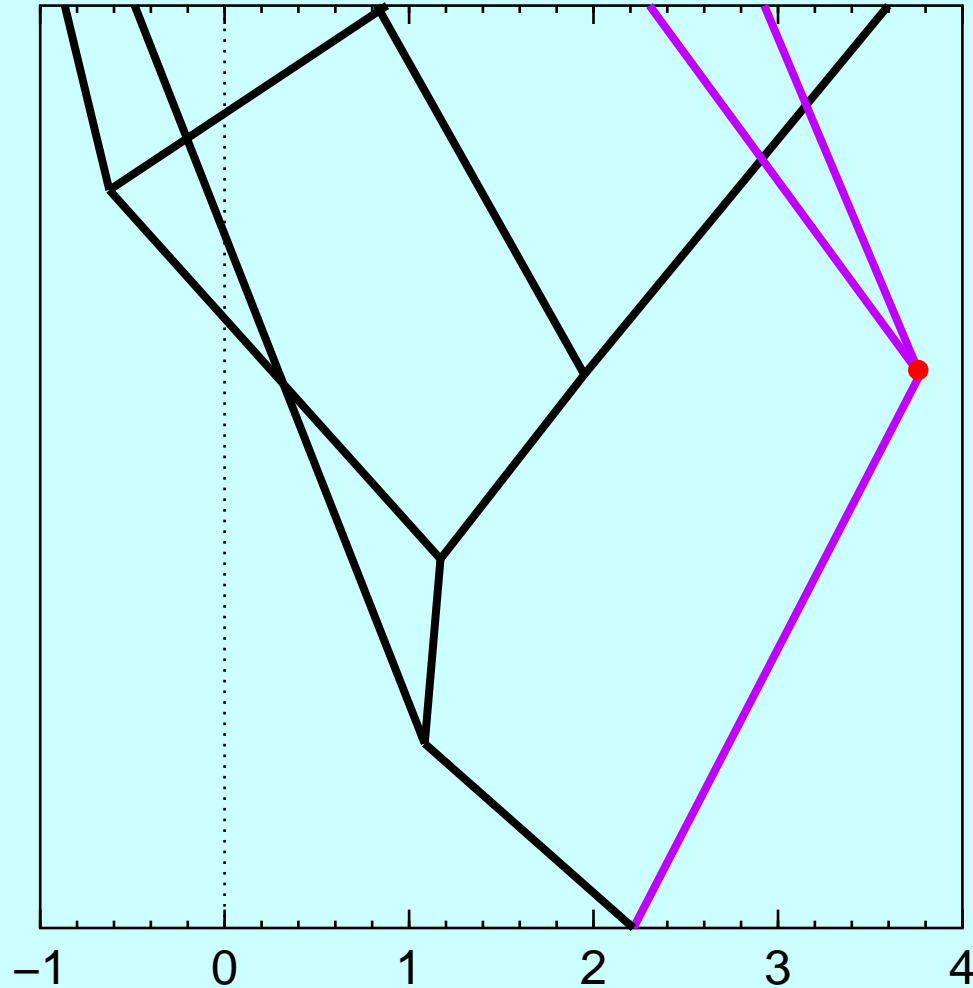
Updating a tip node's liability

An example with 7 species and one character



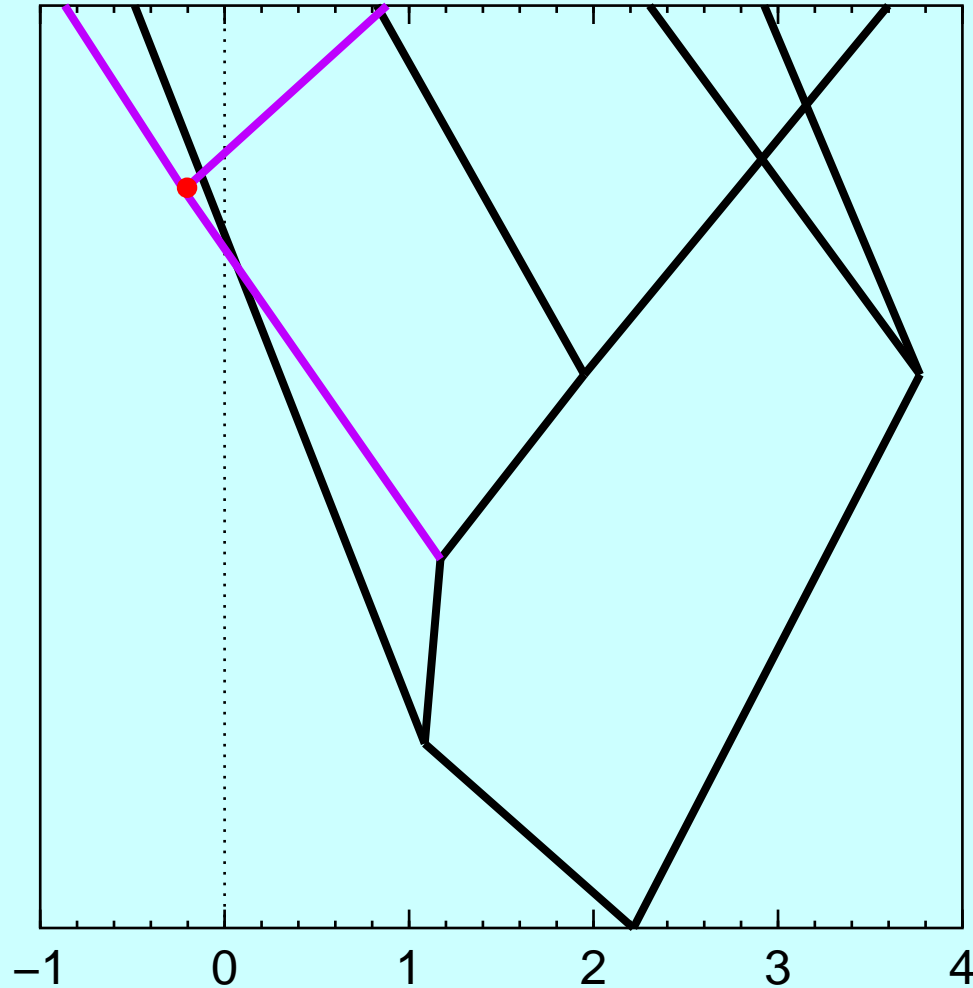
Updating a tip node's liability

An example with 7 species and one character



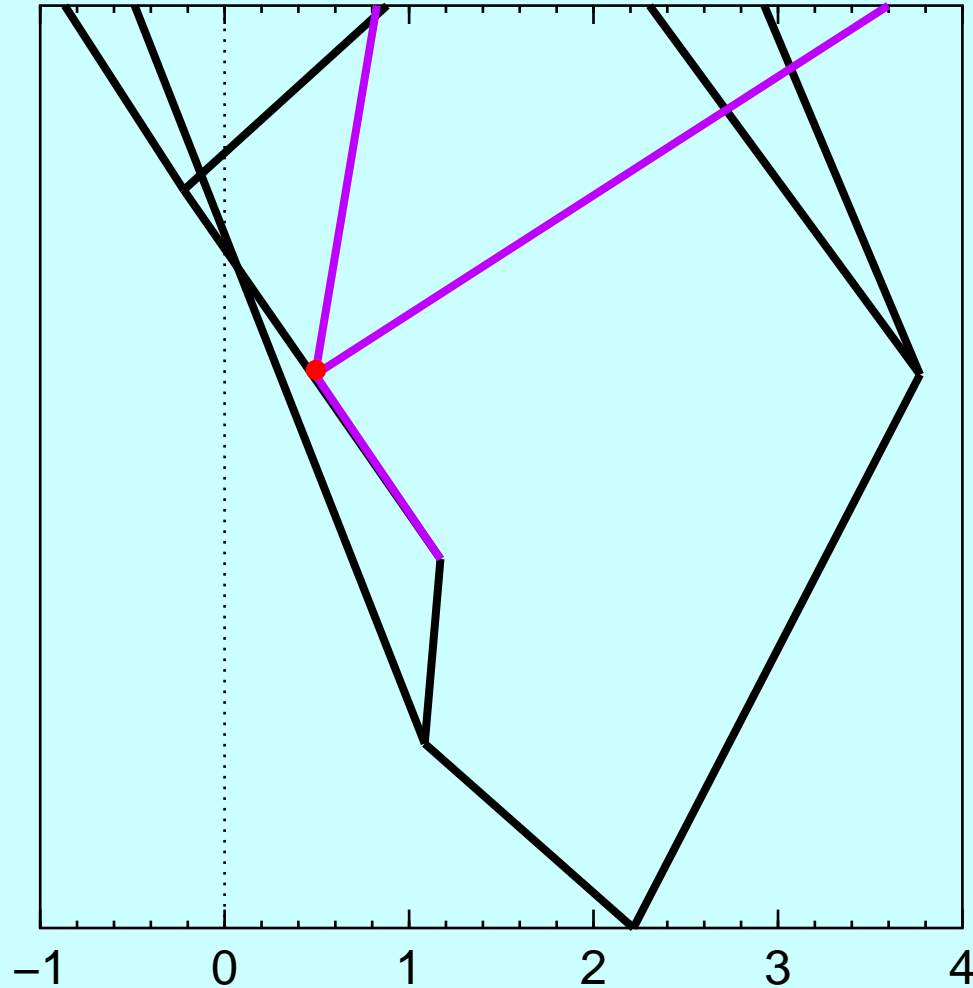
Updating an interior node's liability

An example with 7 species and one character



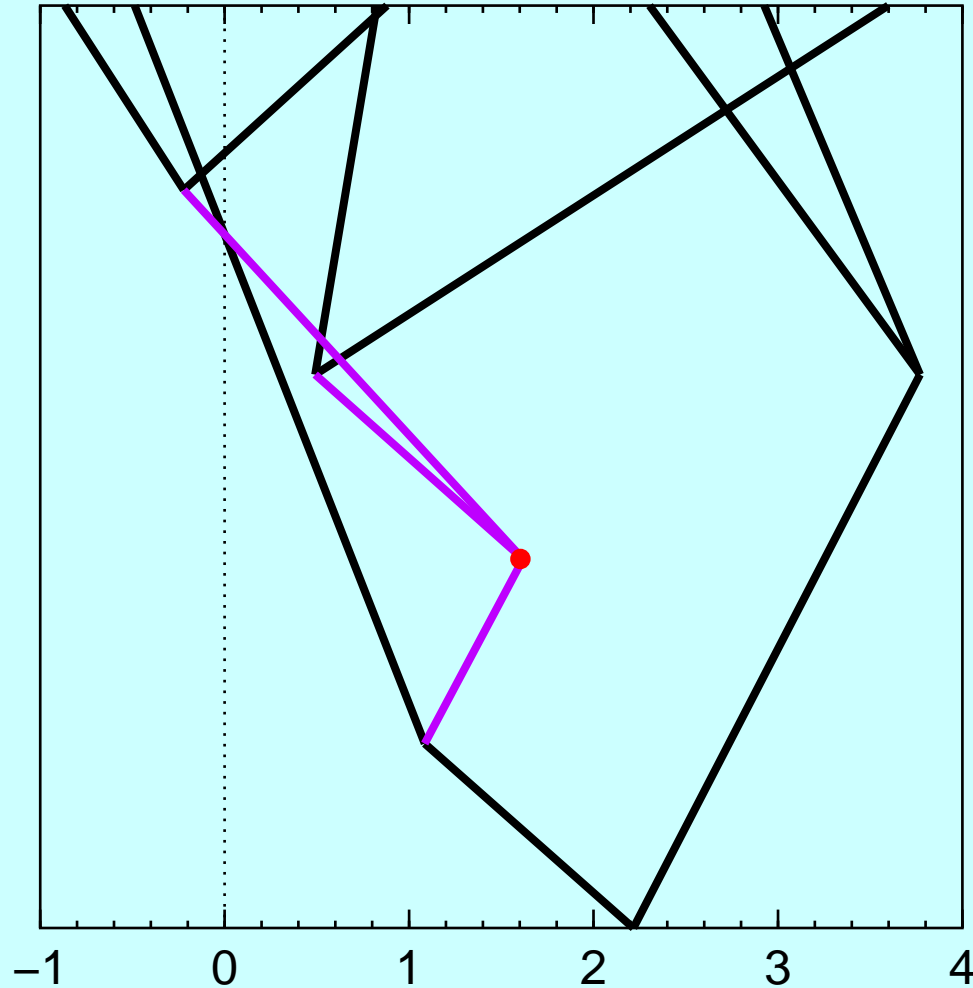
Updating an interior node's liability

An example with 7 species and one character



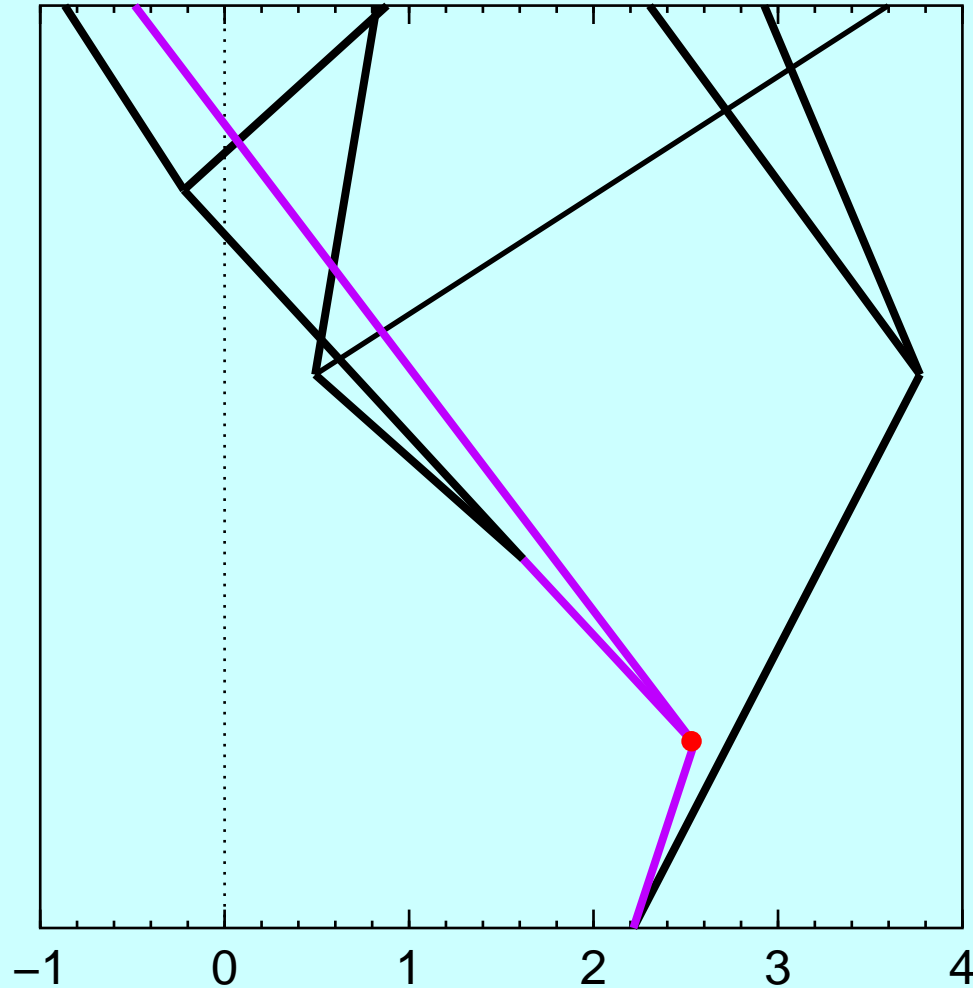
Updating an interior node's liability

An example with 7 species and one character



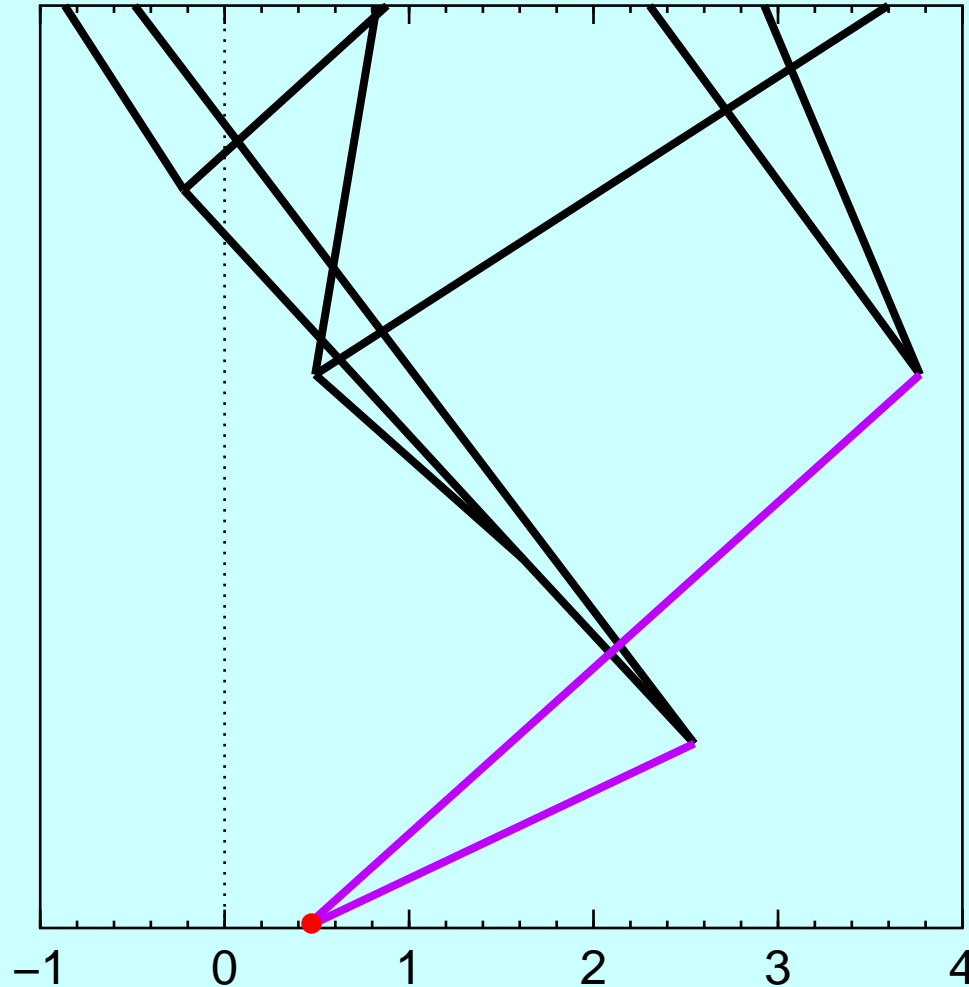
Updating an interior node's liability

An example with 7 species and one character



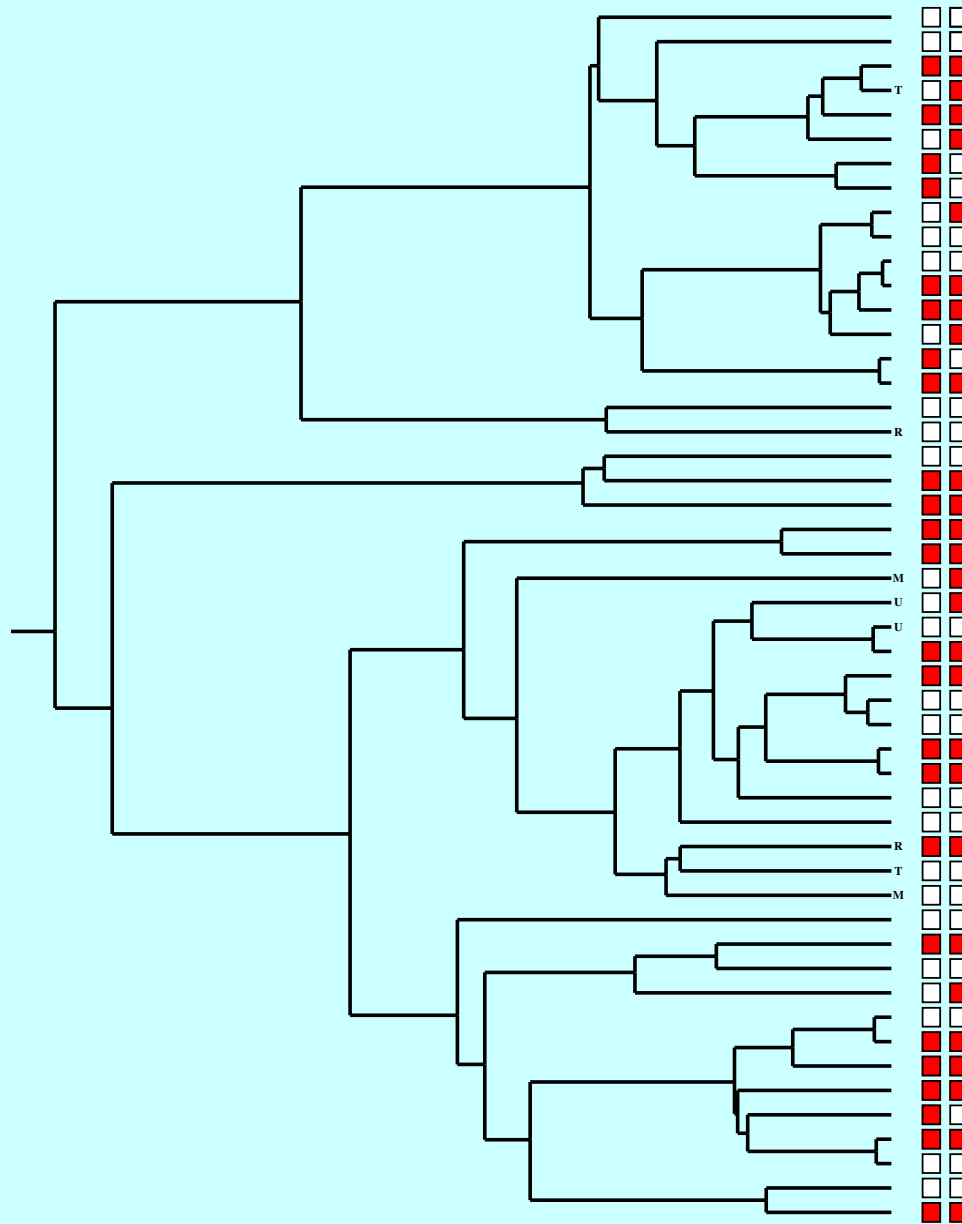
Updating an interior node's liability

An example with 7 species and one character



... and one keeps doing this for tip and interior nodes, many times

A simulated example, with its true tree



A (very) small-scale simulation

True covariance matrix

1	0.8	0
0.8	1.64	-0.6
0	-0.6	1.36

One data set with 50 species on a the same tree:

For the same simulated data set

Run 1			Run 2		
1.70072	1.39356	0.39289	1.68167	1.40056	0.40495
1.39356	1.65104	0.20923	1.40056	1.67836	0.23021
0.39289	0.20923	1.08066	0.40495	0.23021	1.09550



A simulation: inferred correlation coefficients

With 50 trees, each with 2 simulated data sets:

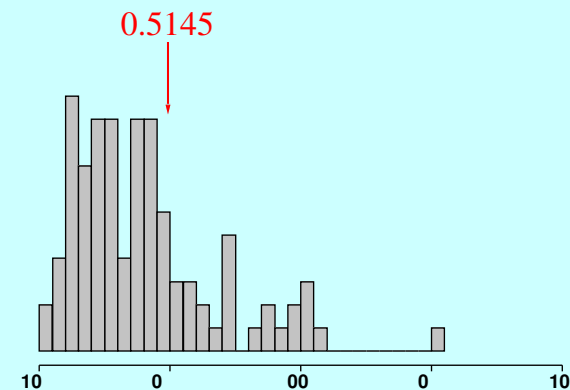
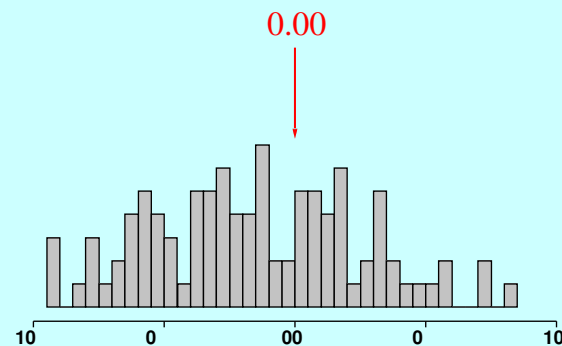
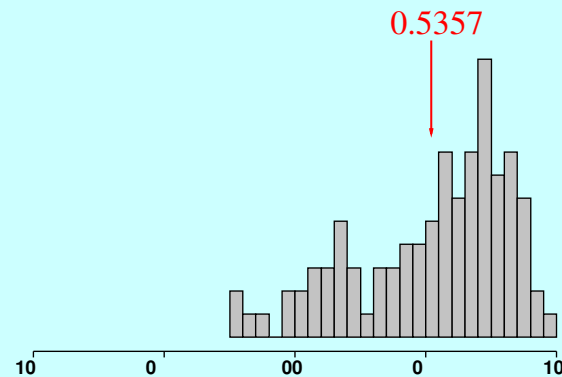
character 1

character 2

character 1

character 2

character 3



Continuous characters too? Both?

- We could do this for continuous characters too by assuming the tip liabilities are observed and don't change

Continuous characters too? Both?

- We could do this for continuous characters too by assuming the tip liabilities are observed and don't change
- Then we could estimate covariances.

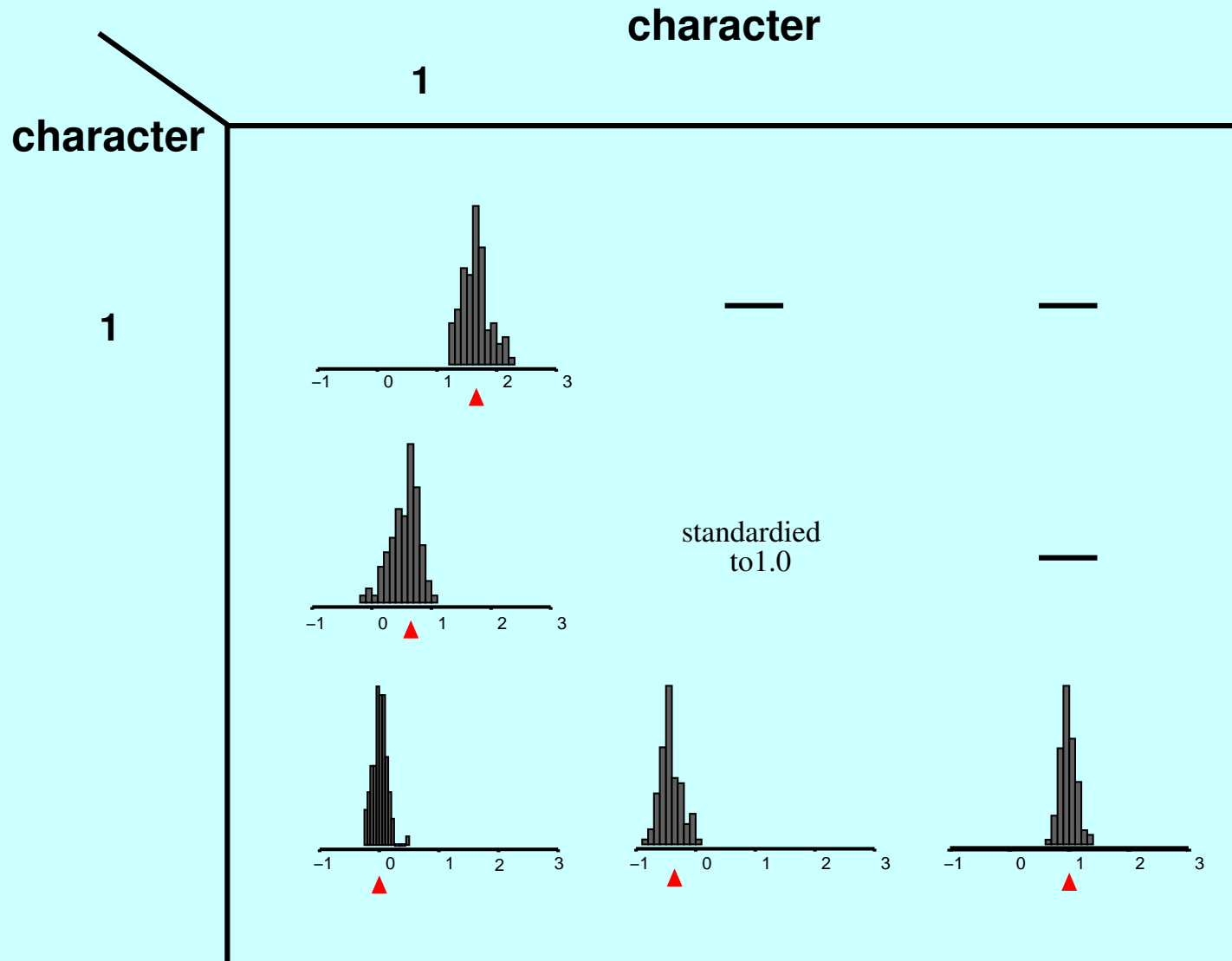
Continuous characters too? Both?

- We could do this for continuous characters too by assuming the tip liabilities are observed and don't change
- Then we could estimate covariances.
- This would be a noisier version of what we can do exactly already, so there is no point to it, right?

Continuous characters too? Both?

- We could do this for continuous characters too by assuming the tip liabilities are observed and don't change
- Then we could estimate covariances.
- This would be a noisier version of what we can do exactly already, so there is no point to it, right?
- But ... we could do both discrete and continuous characters together in this way, with almost no extra effort.

Simulations with both discrete and continuous characters



Characters 1 and 3 are continuous, character 2 is discrete. The inferred covariances are shown for the 100 simulated data sets.

References for threshold methods

- Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* **19**: 506-536. [The threshold model for discrete traits]
- Falconer, D. S. 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics* **29**: 51-76. [Threshold model applied to human diseases]
- Gianola, D. 1982. Theory and analysis of threshold characters. *Journal of Animal Science* **54**: 1079-1096. [Use in animal and plant breeding]
- Sorensen, D. and D. Gianola. 2002. *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer, New York. [More ...]
- Korsgaard, I. R., M. S. Lund, D. Sorensen, D. Gianola, P. Madsen, and J. Jensen. 2003. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genetics Selection Evolution* **35** (2): 159-183. [... and more ...]
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913-925. [Uses 0/1 stochastic process to infer morphological phylogenies]

References, continued

Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London Series B Biological Sciences*. **255**: 37-45. [0/1

stochastic model for discrete characters]

Felsenstein, J. 2002. Quantitative characters, phylogenies, and morphometrics.pp. 27-44 in “Morphology, Shape, and Phylogenetics”, ed. N. MacLeod. Systematics Association Special Volume Series 64. Taylor and Francis, London. [Review repeating 1988 material and going into some more detail on the question of threshold models.]

Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society of London, series B* **360**: 1427-1434. [This project in a slightly earlier version]

Felsenstein, J. 2012. A comparative method for both discrete and continuous characters using the threshold model. *American Naturalist* **179**: 145-156. [The final paper for discrete and continuous analyses]

Felsenstein, J. 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts. [Mentions this model and method.]