How can molecular phylogenies illuminate morphological evolution?

27 October 2016.

Joe Felsenstein

UNAM

Where this lecture fits in

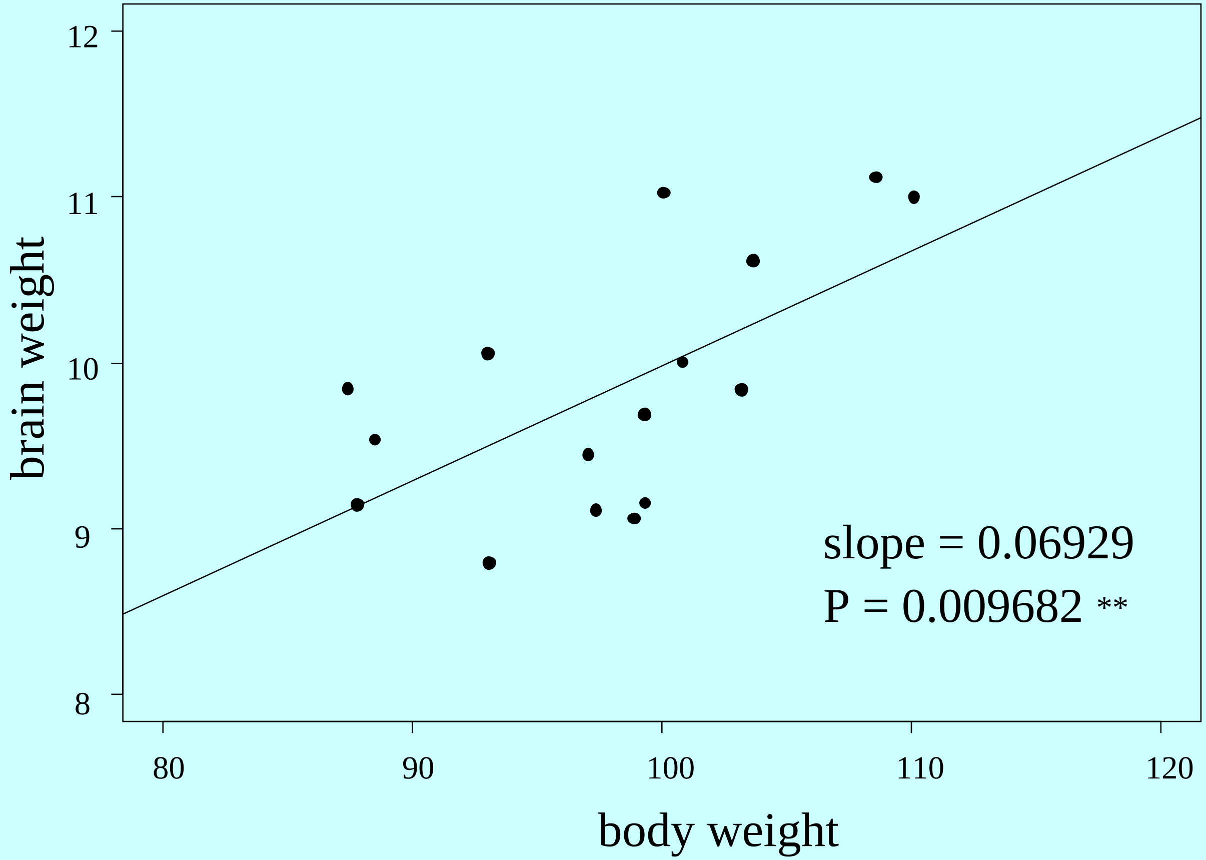
Lately, there has been more integration of

- work on molecular evolution
- work on between-species differences of measurable characters
- work on within-species differences of measurable characters

How can they fit together?

A routine study of covariation of characters

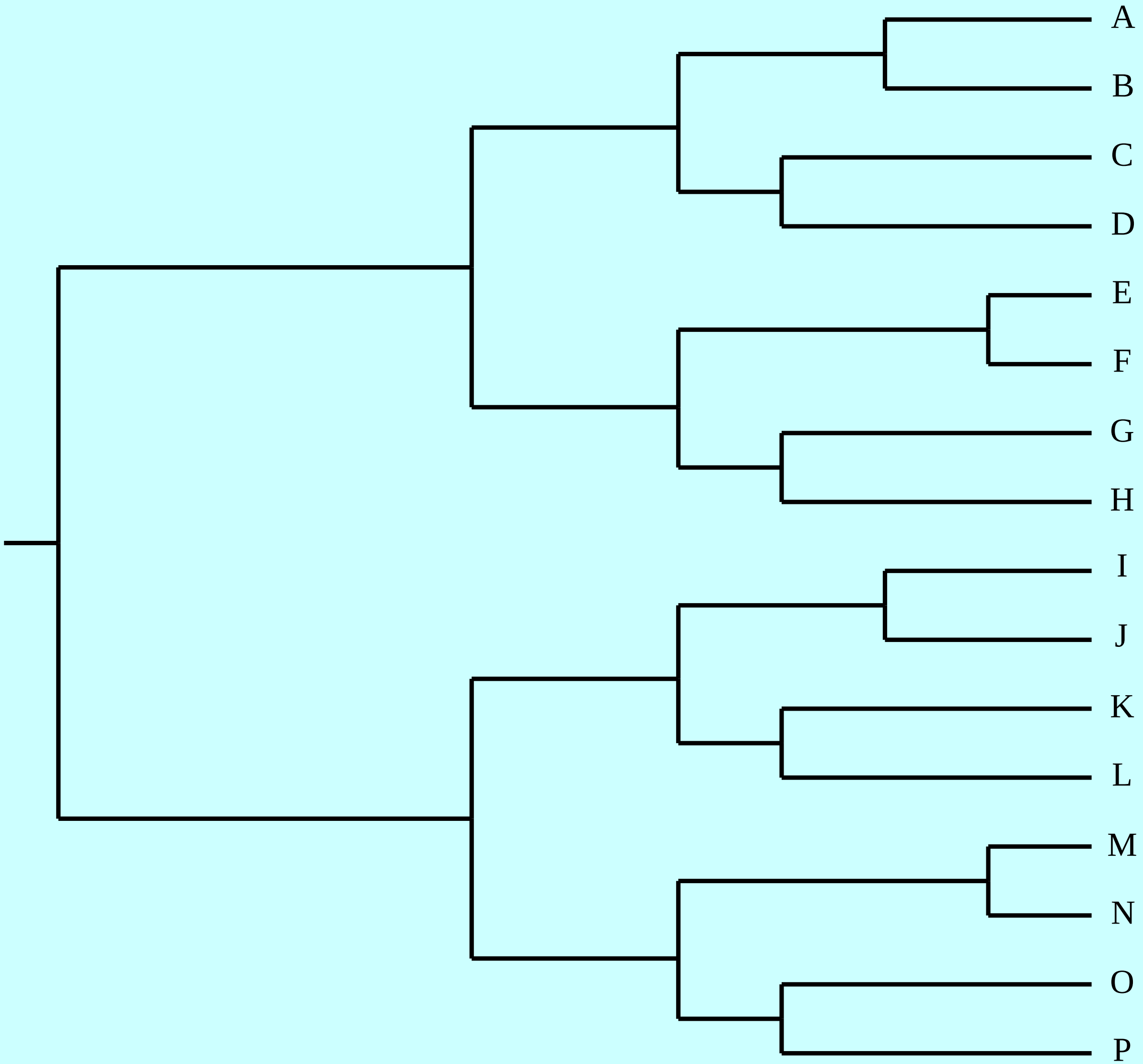
Using an ordinary regression with the species as points, we see a significant relationship between brain weight and body weight:



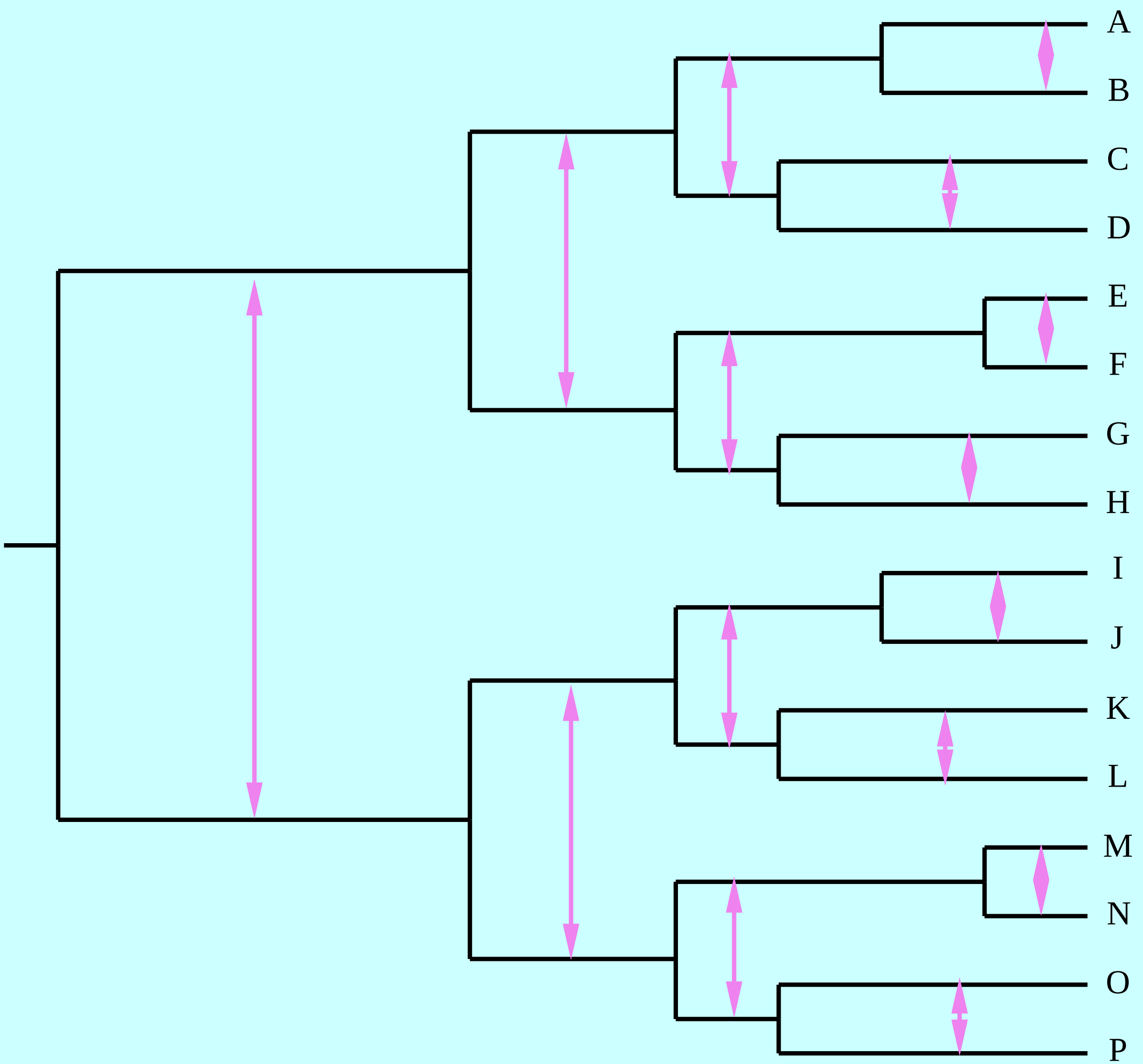
It looks as if we have 16 independent data points and a positive correlation between brain weight and body weight across species.

But the points are not independent

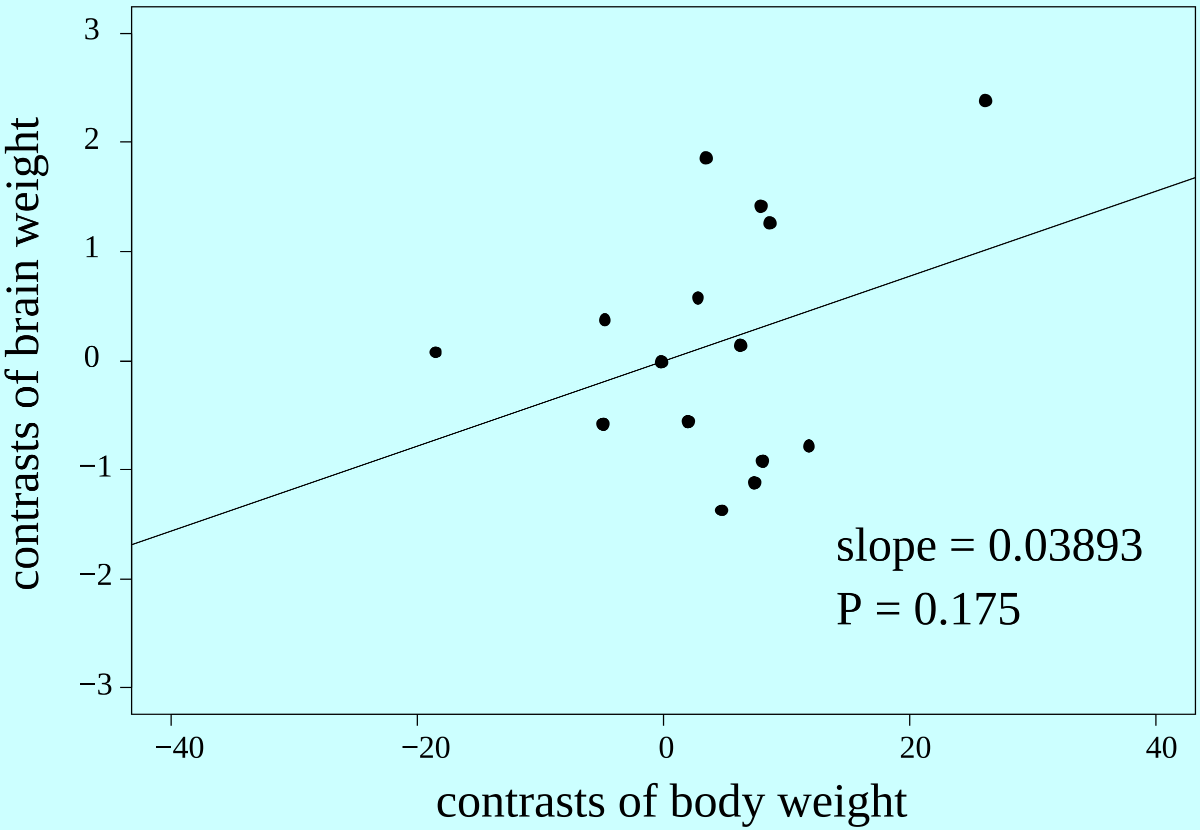
They evolved on a phylogeny. More closely related points are similar.



Using contrasts on the phylogeny ...



Is evolution of brain and body weight correlated?

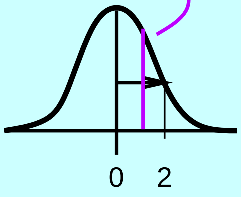


Using the contrasts method we see no significant relationship.

A standard quantitative genetics model

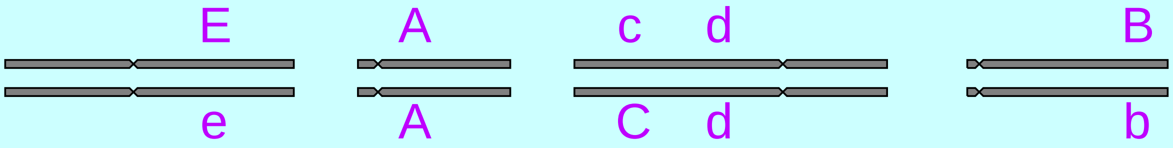
$$P = \mu + \left\{\begin{array}{ll} AA & 2 \\ Aa & 4 \\ aa & 7 \end{array}\right\} + \left\{\begin{array}{ll} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{array}\right\} + \left\{\begin{array}{ll} CC & -1 \\ Cc & 6 \\ cc & 6 \end{array}\right\} + \left\{\begin{array}{ll} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{array}\right\} + \left\{\begin{array}{ll} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{array}\right\} + \text{environmental effect}$$

arbitrary starting point

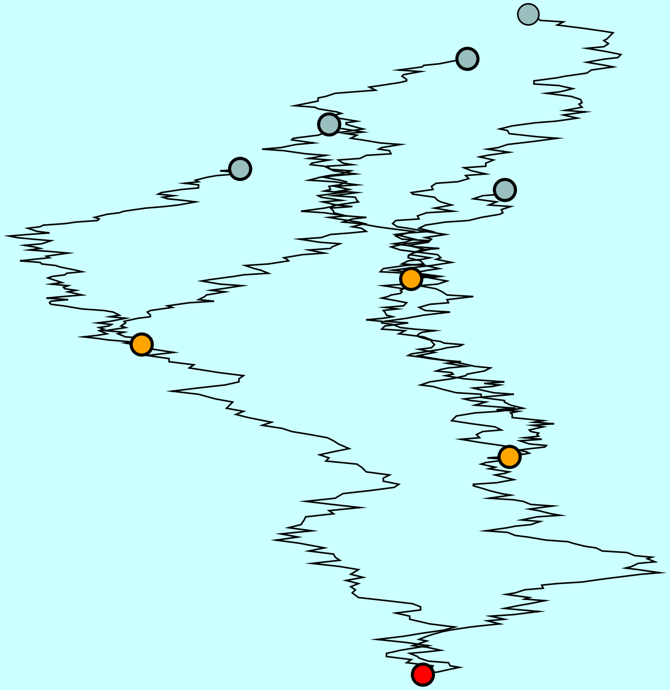


AA	**Bb**	**Cc**	**dd**	**Ee**	**10**
Aa	**bb**	**cc**	**DD**	**ee**	
aa	**bb**	**CC**	**DD**	**Ee**	
aa	**bb**	**Cc**	**DD**	**EE**	
Aa	**Bb**	**Cc**	**DD**	**Ee**	

$$0.3 + 2 + 6 + 0.7 + 0.1 + 0.9$$

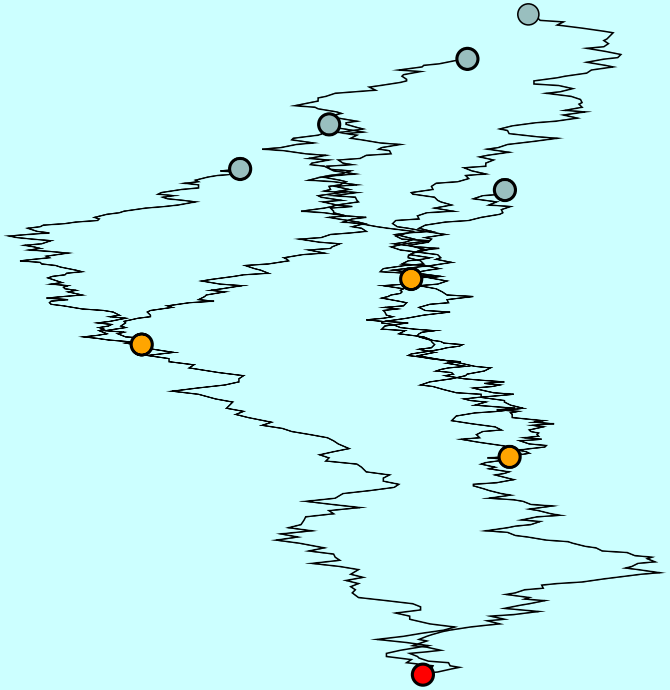


A model of quantitative characters on a phylogeny



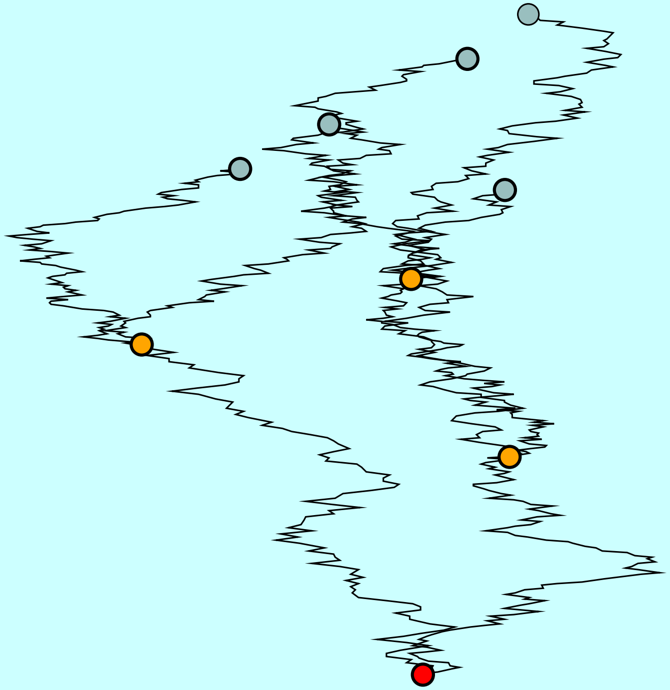
- Brownian motion with multiple characters with different variances and with covariation as well.

A model of quantitative characters on a phylogeny



- Brownian motion with multiple characters with different variances and with covariation as well.
- This started with approximating gene frequencies in the 1960s by Anthony Edwards and Luca Cavalli-Sforza.

A model of quantitative characters on a phylogeny



- Brownian motion with multiple characters with different variances and with covariation as well.
- This started with approximating gene frequencies in the 1960s by Anthony Edwards and Luca Cavalli-Sforza.
- I expanded it to model quantitative characters determined by these geness (1973, 1981, 1988).

Models for long-term evolution

The use of quantative genetics approximations to model long-term evolution in lineages was largely introduced by Russ Lande in the 1980s.

Russell Lande, from his website at Imperial College, U.K., where he has been in recent years.

Where do the covariances come from?

- **Genetic covariances** (the same loci affect two or more traits). Genetic drift or natural selection can change the gene frequencies at these loci, and thus make correlated changes in the two traits.

Where do the covariances come from?

- **Genetic covariances** (the same loci affect two or more traits). Genetic drift or natural selection can change the gene frequencies at these loci, and thus make correlated changes in the two traits.
- **Selective covariances** (Olof Tedin, 1926; G. Ledyard Stebbins 1950) The same environmental conditions can select changes in two or more traits – even though they may have no genetic covariance. This source of evolutionary covariance is widely ignored.

An example: morphometrics and phylogenies

Fred Bookstein is a co-author on this part of the talk

Fred Bookstein

me

“J. F. L. Bookenstein”

(Our reconstructed common ancestor)

How to use morphometric coordinates on phylogenies?

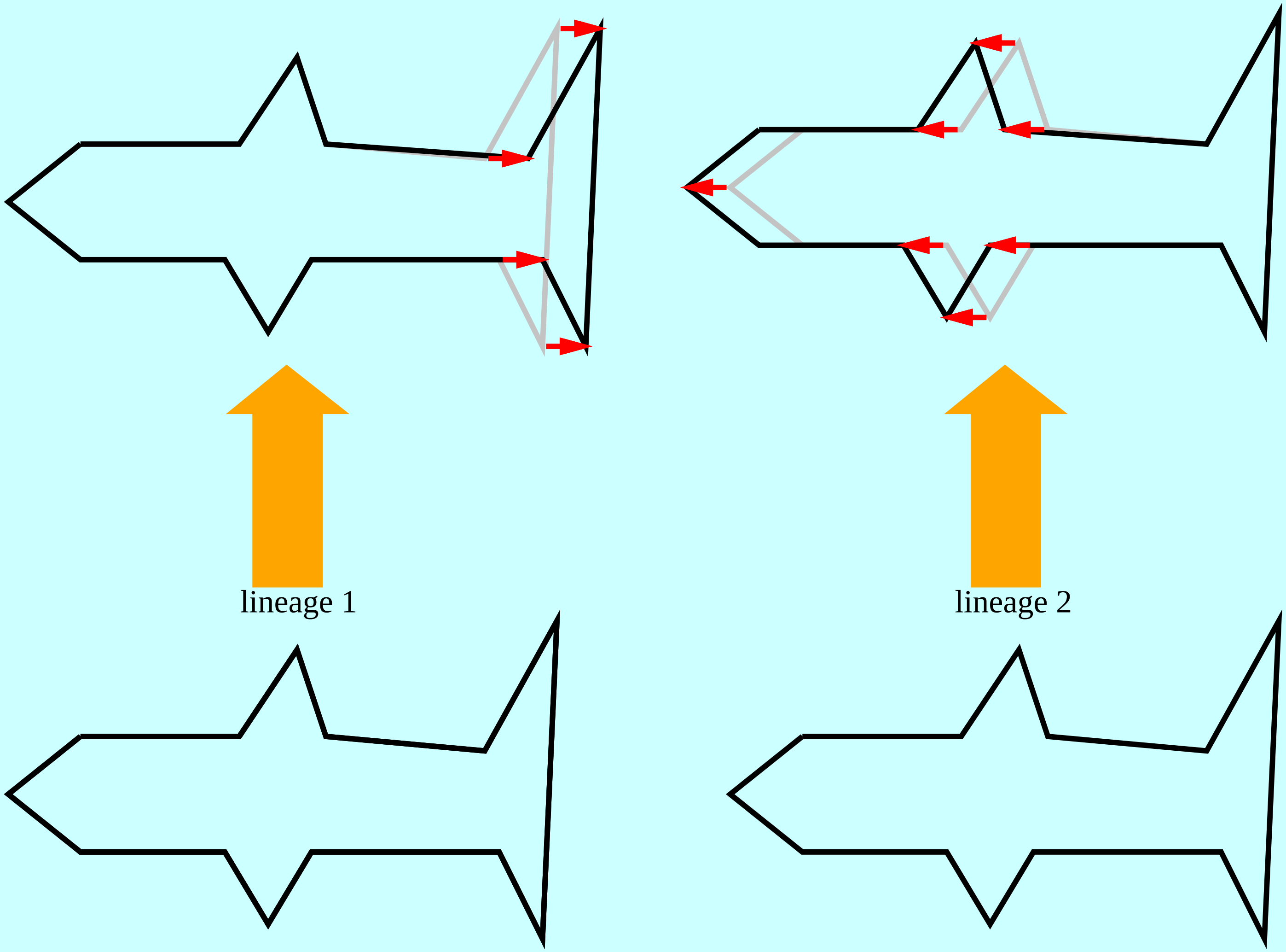
Is it possible to simply use the coordinates of landmarks $(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)$ as continuous phenotypes $x_1, y_1, \ldots, x_p, y_p$ using Brownian motion along a phylogeny?

Yes, but ...

We must do proper morphometrics (correct for translation? rotation?)

Can we superpose specimens?

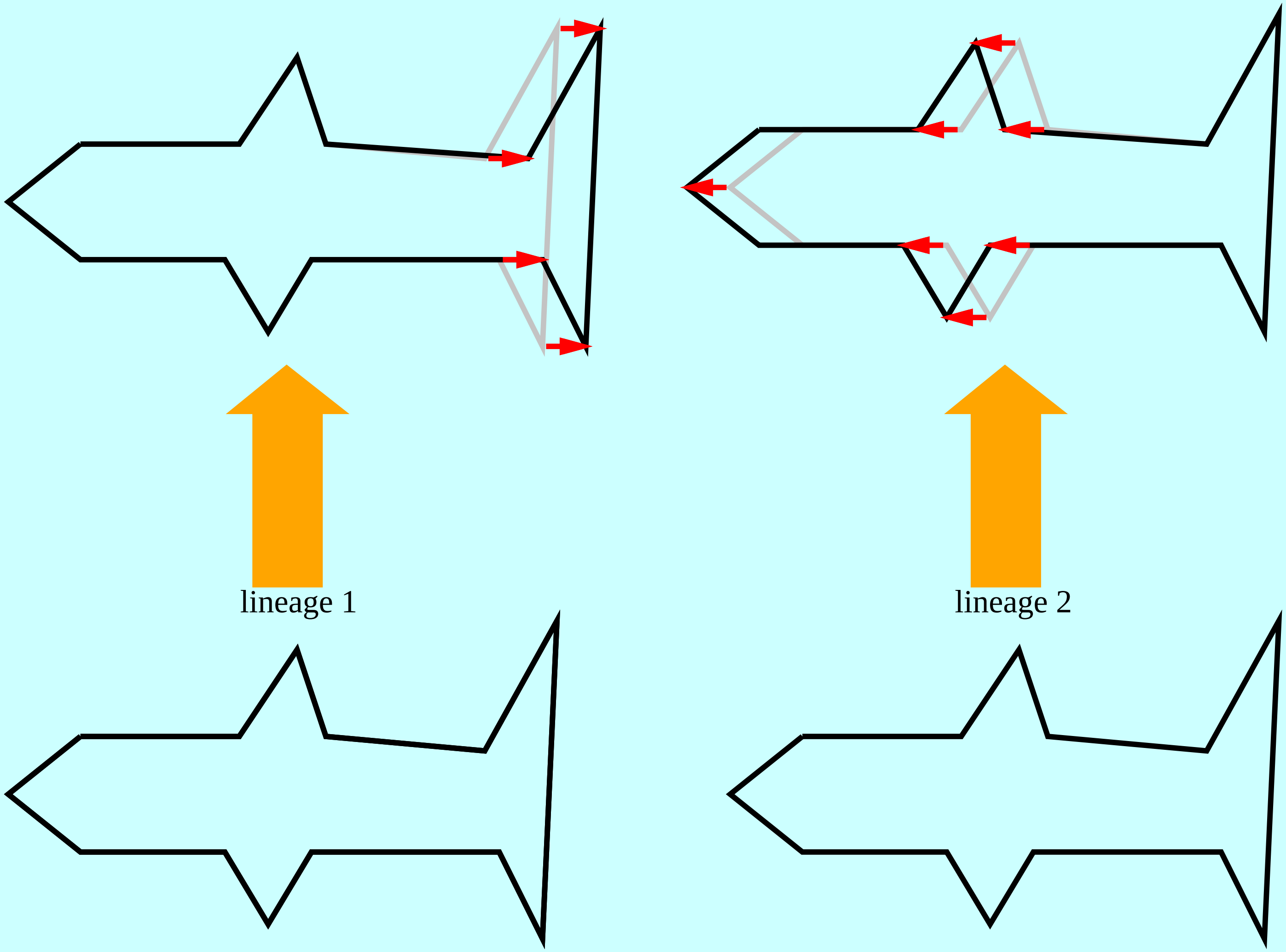
Consider two cases:



Are these different?

Why superposition is in principle impossible

Consider two cases:



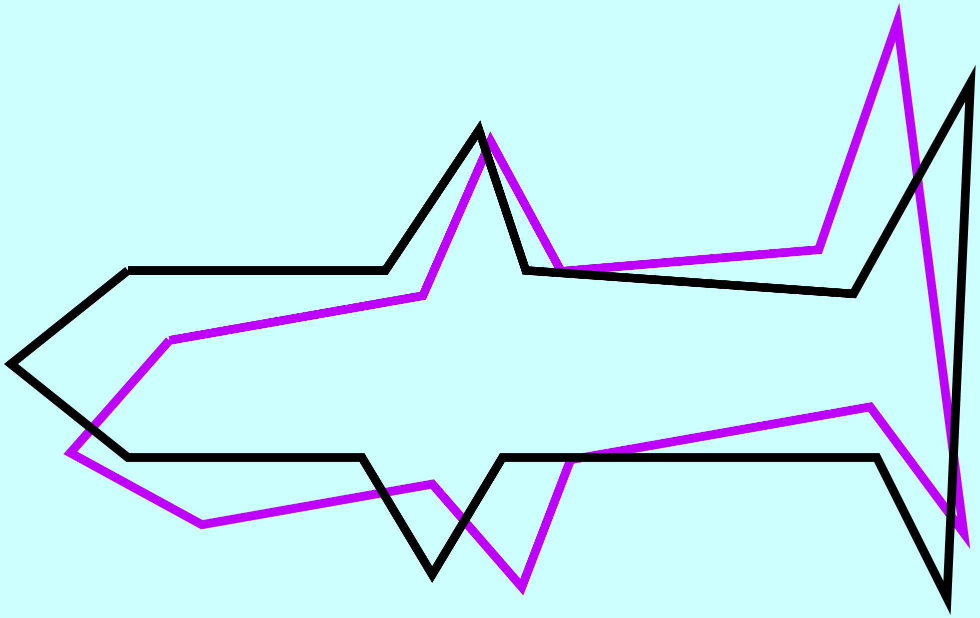
Are these different? No!

Dealing with translation

In effect one is centering each specimen so that the mean of its points is at $(0,0)$. (The assumption is that the horizontal and vertical placement of the specimen on the digitizer is not useful information).

This has the effect of dropping two degrees of freedom so that each specimen now has $2p-2$ coordinates. It now “lives” in a $(2p-2)$-dimensional space.

The annoying issue of rotation

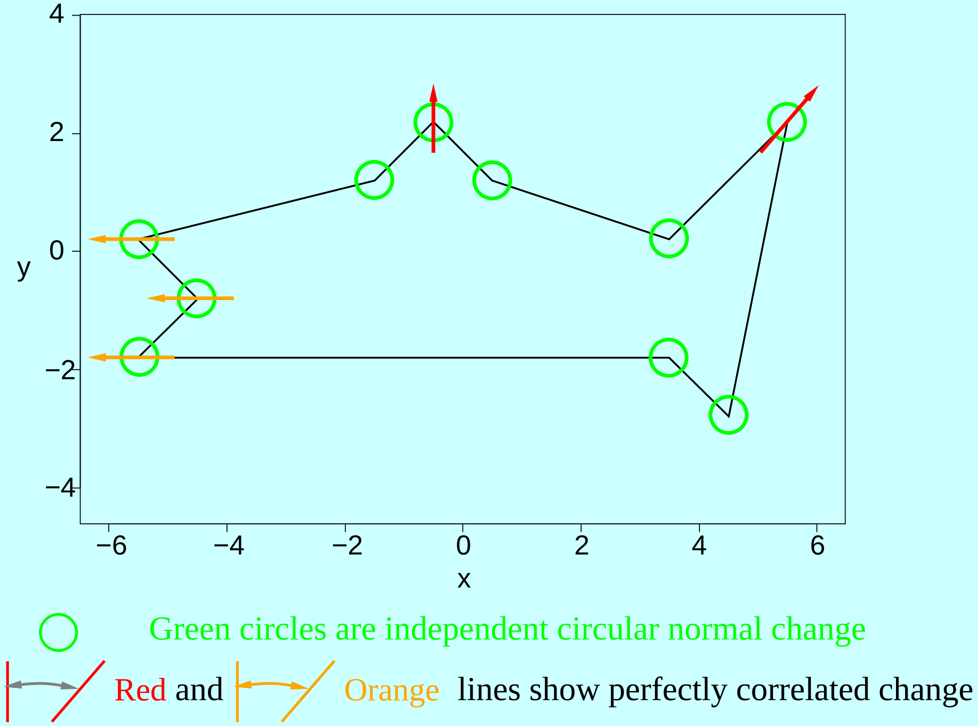


Sadly, there is no corresponding transform that tosses out rotation, as there is for translation.

We use maximum likelihood inference of the angles, changing them all until the likelihood of the dataset is maximized. This is not a perfect method because it introduces too many parameters.

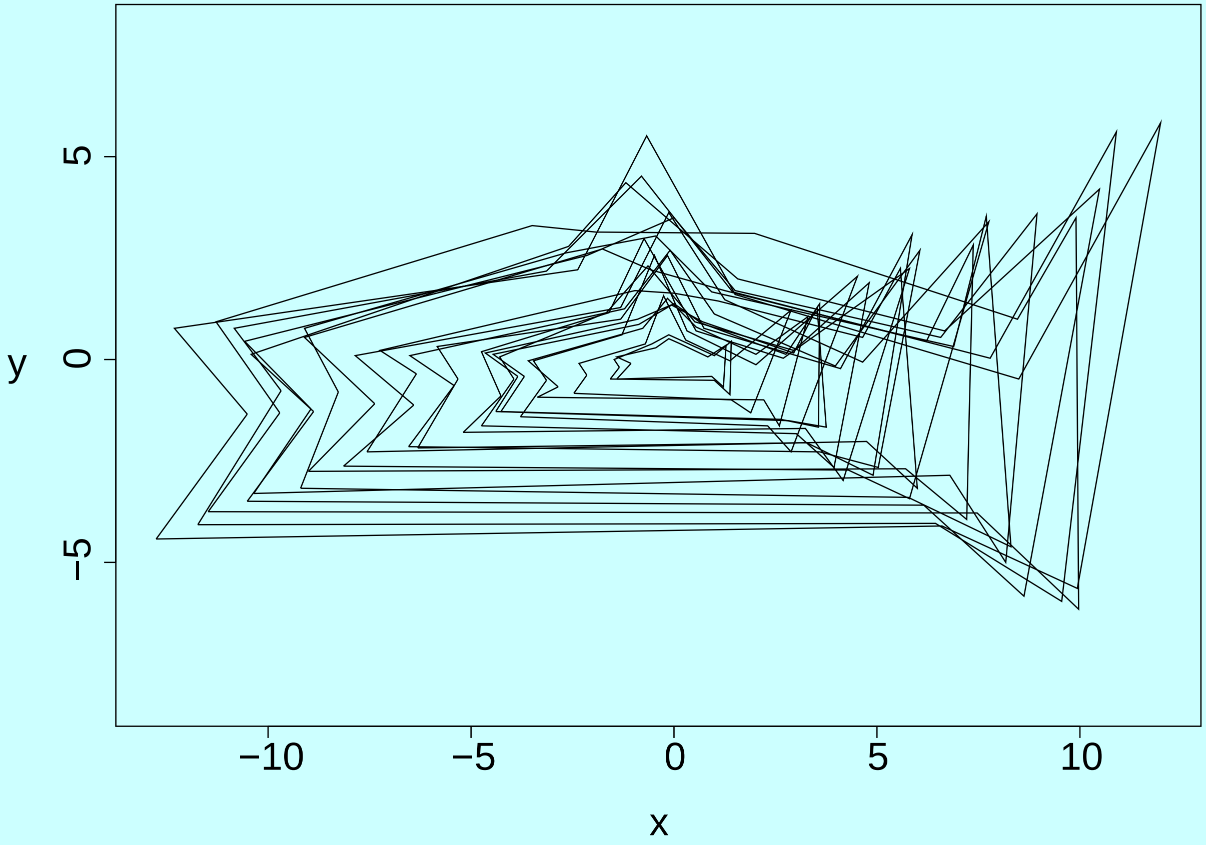
A simulation test

1. Generate 50 100-species trees by a pure birth process
2. For each evolve 100 forms by (covarying) Brownian Motion up the tree
3. These are the true covariances:

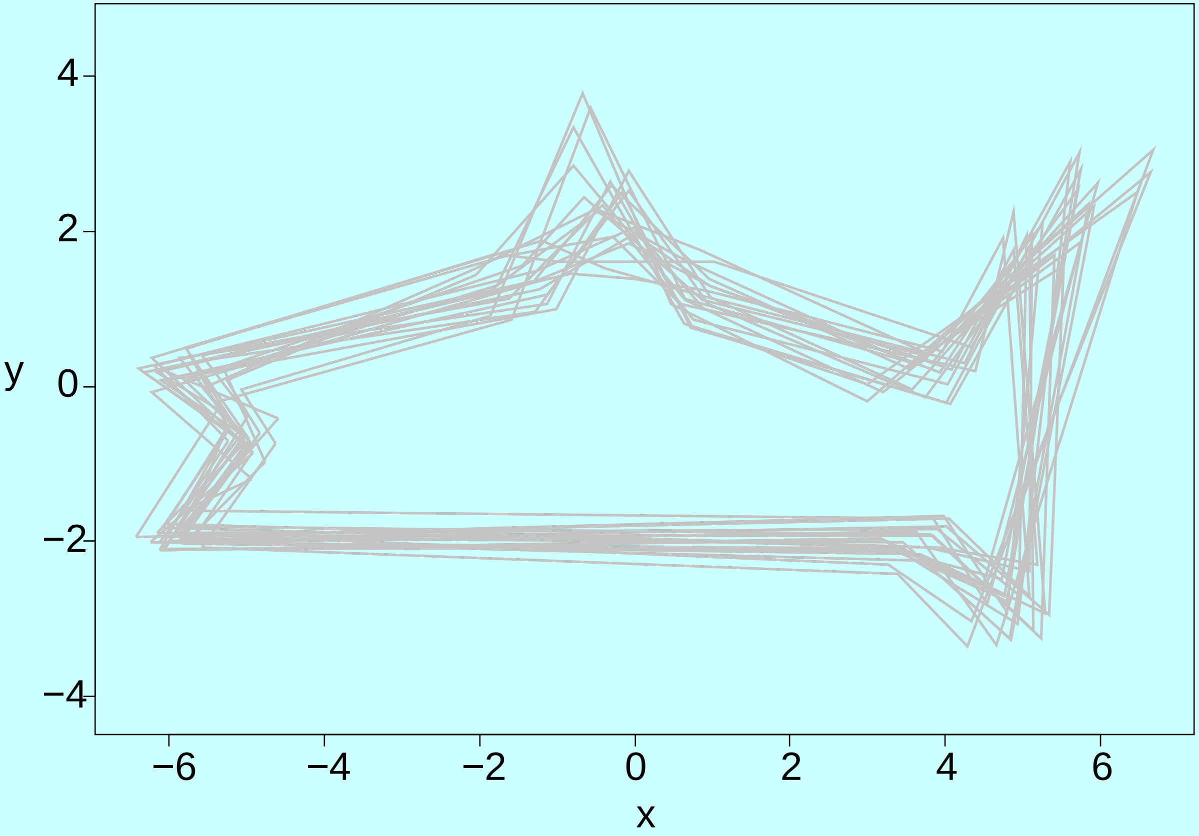


- All 10 landmarks move by independent and equal Brownian Motion of the coordinates with variance (per unit branch length) of 0.001, *plus*
- the vertical coordinate of the pectoral fin and the two coordinates of the top of the tail move in a perfectly correlated change with variance 0.016, *plus*
- the x coordinates of the nose change with a variance of 0.008 and an allometric regression on log-size of 0.5

20 of 100 fishes from data set #1, centered and rotated

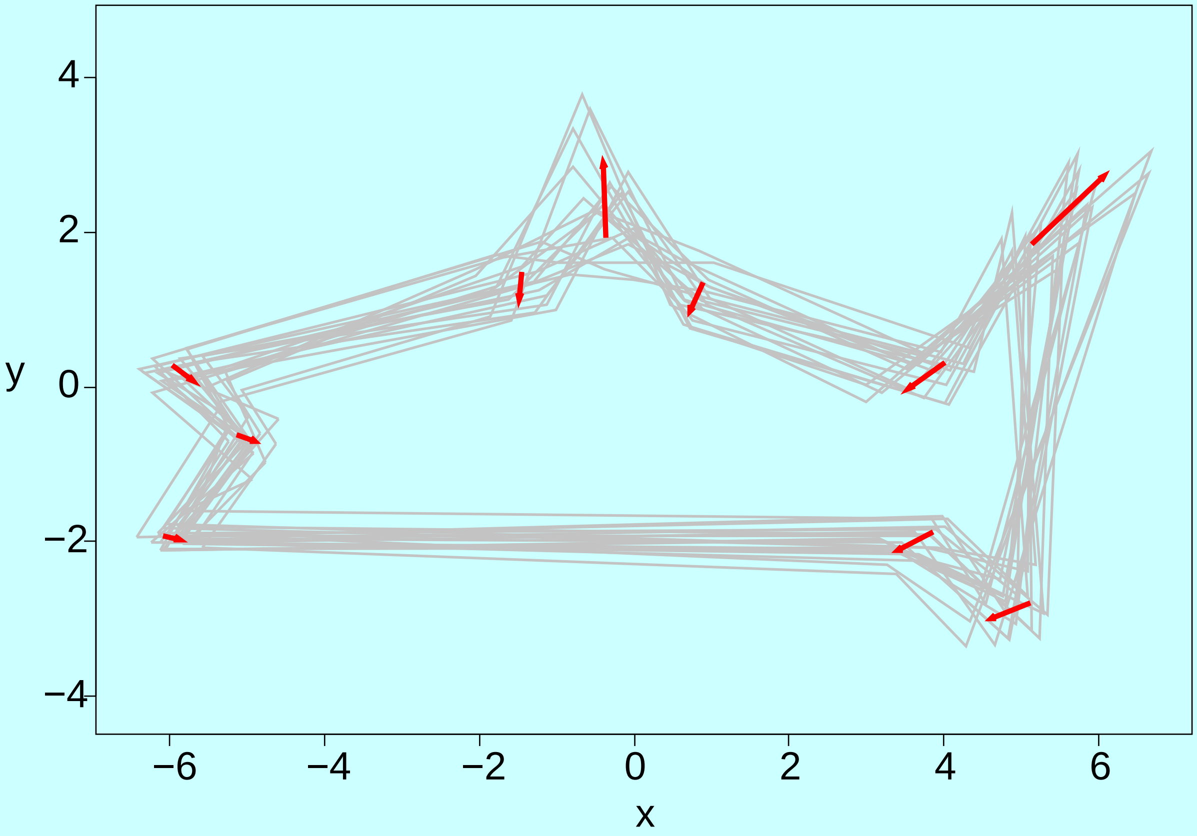


20 of the 100 fishes from data set #1, also rescaled



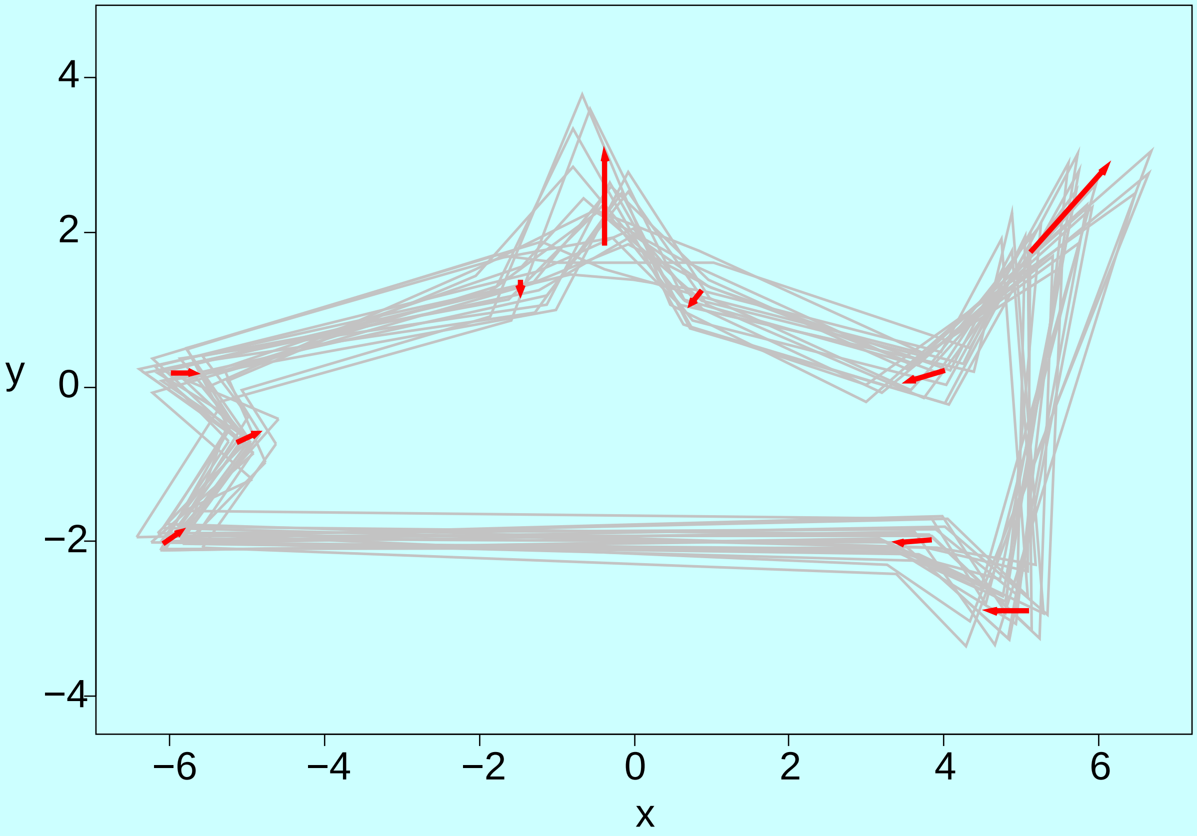
The rotating and scaling is done by maximum likelihood estimation of specimen angles and sizes.

First shape PC 1 for data set #1



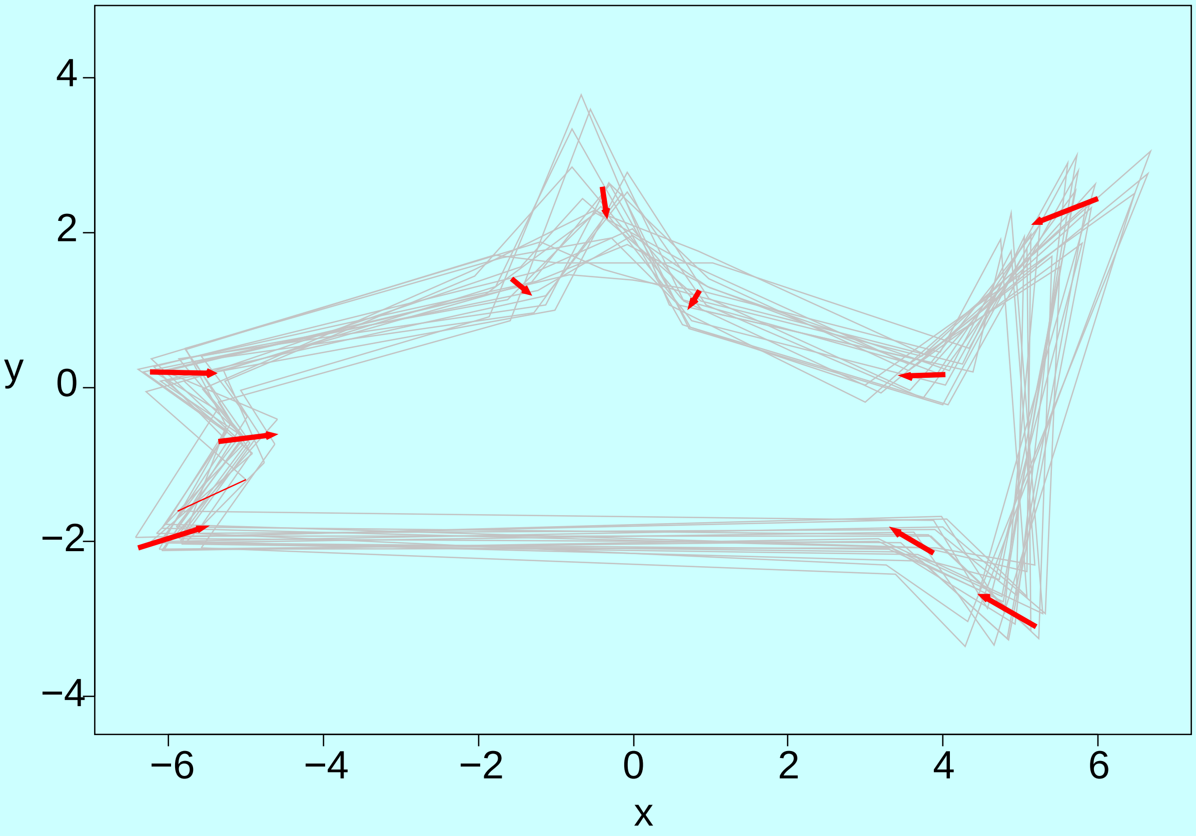
Now we’ve inferred a scale (size) component and removed it from the covariances, and then taken the first PC of the residual on size. We can see the fin component more clearly.

Making the first shape PC sparser by “medianizing”



To make PC1 be sparses we can add in a little location (not forcing the changes to maintain the centroid supeposition). This is done by subtracting from the x components, their median, and similarly for the y components. So it minimizes the L^1 norm of the PC coefficients. The result is very clear.

What do we get from the Morphometric Consensus?



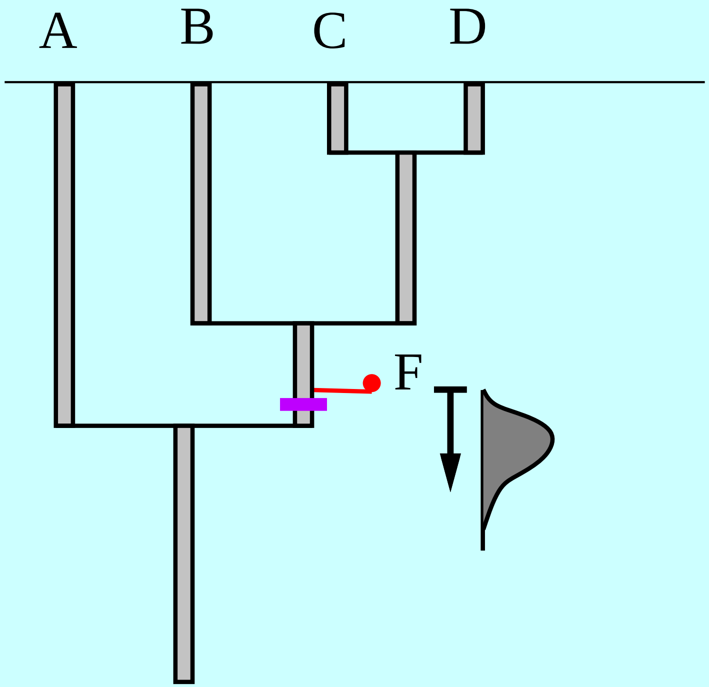
... we get a not-as-clear result with some size still there – we have ignored the tree and taken out size by standardizing centroid size, which is affected more by the fin component in the MMC methods.

Another example: Fossils and phylogenies

Similar to an approach published recently by Revell et al. (2015)

Liam Revell of University of Massachusetts, Boston (shown in Puerto Rico)

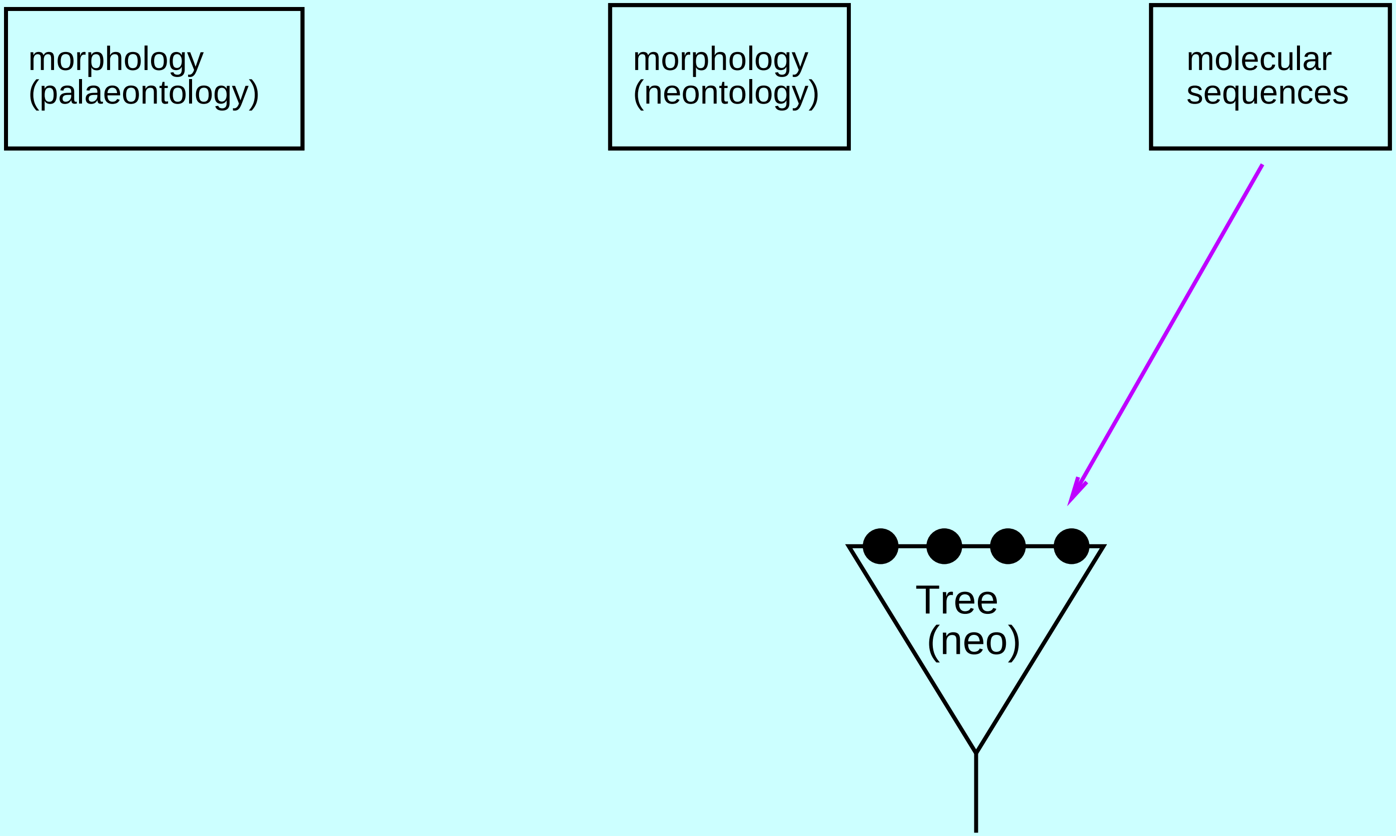
Present methods for calibration



Can take a fossil to indicate a bound on how recently a common ancestor was present. Use various priors on how much earlier or how much more recently.

But there is another way, which is being explored by me and (independently) by Alexander Pyron (2011) and by Fredrik Ronquist et al. (2012)

A better way of using fossils



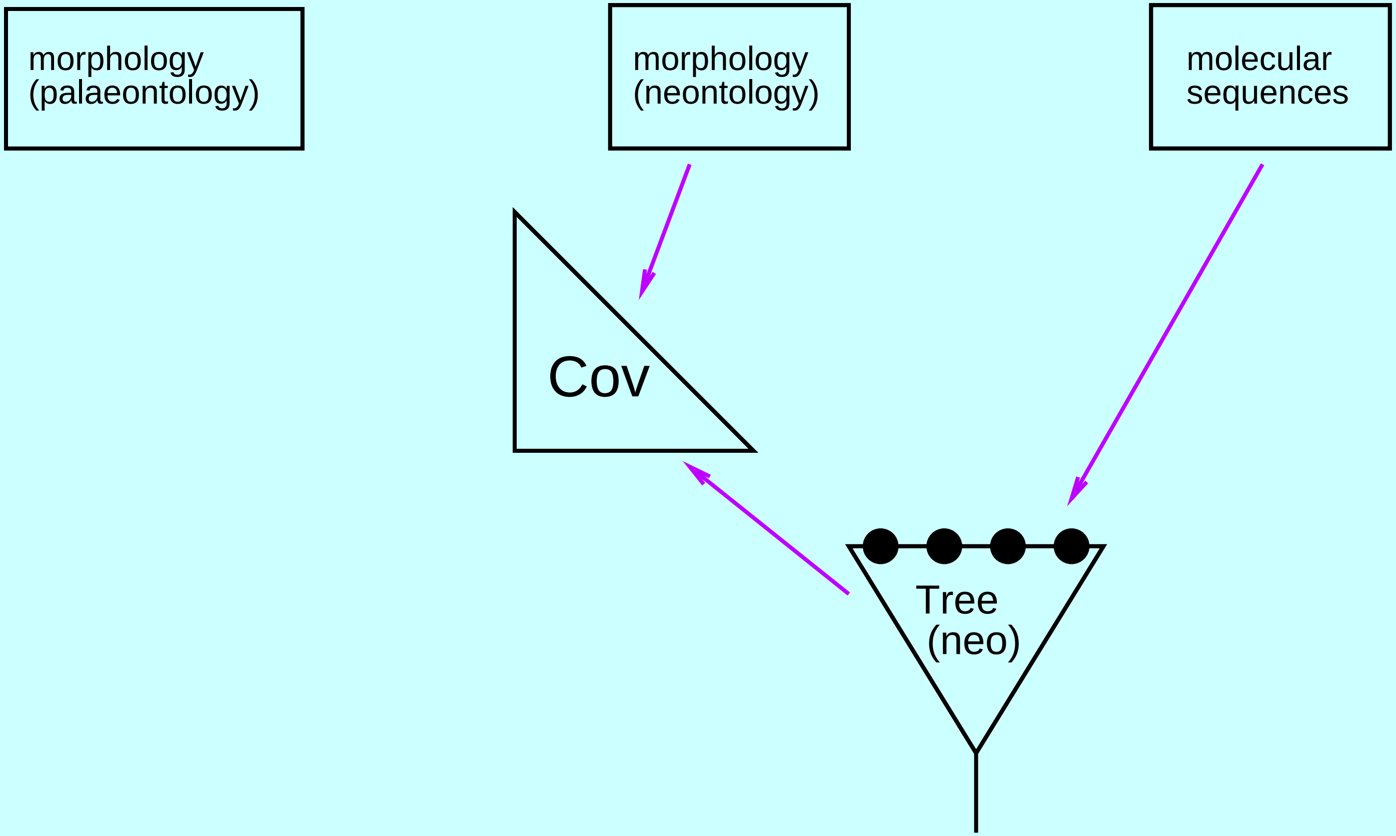
morphology
(palaeontology)

morphology
(neontology)

molecular
sequences

Infer tree of present-day species from molecular sequences

Using fossils



morphology
(palaeontology)

morphology
(neontology)

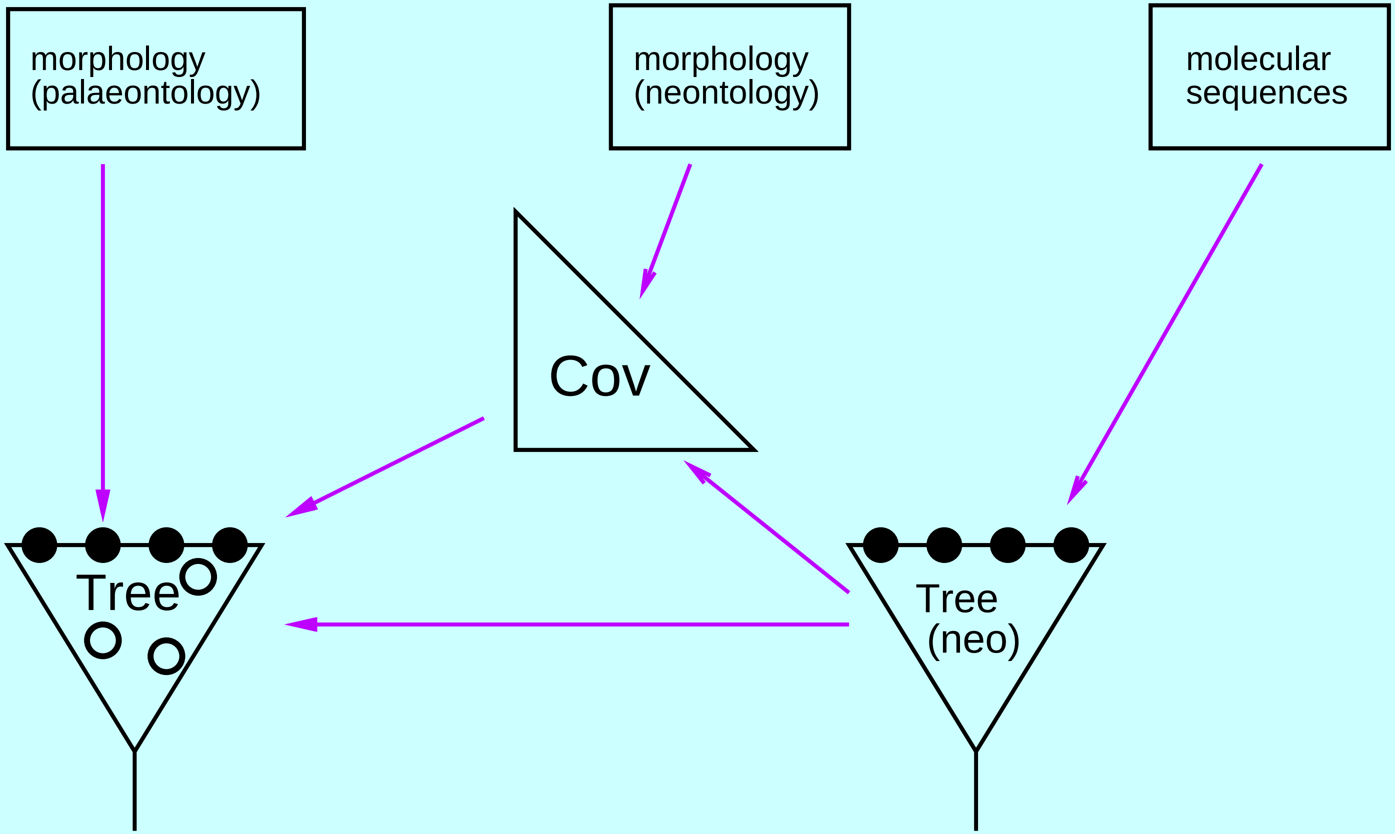
molecular
sequences

Cov

Tree
(neo)

Infer covariances of morphology using it, present-day species

Using fossils



morphology (palaeontology)

morphology (neontology)

molecular sequences

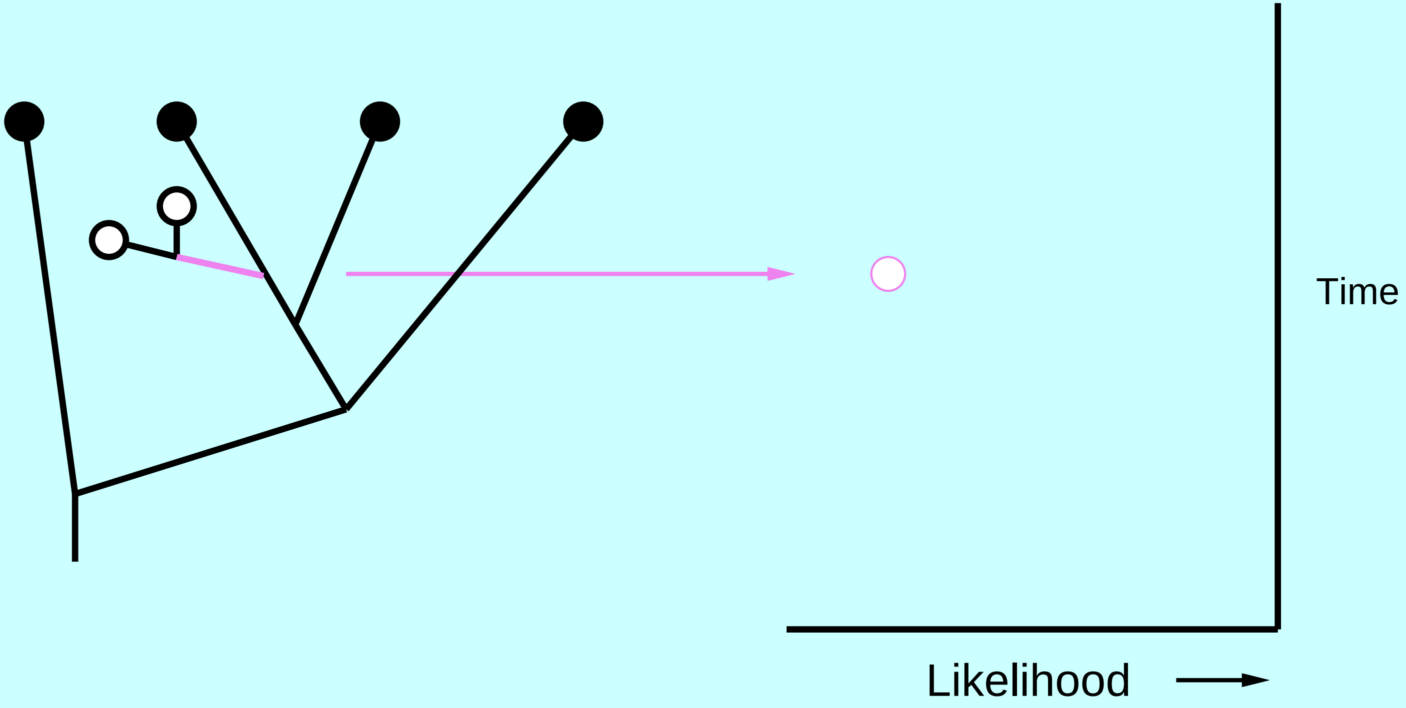
Cov

Tree

Tree (neo)

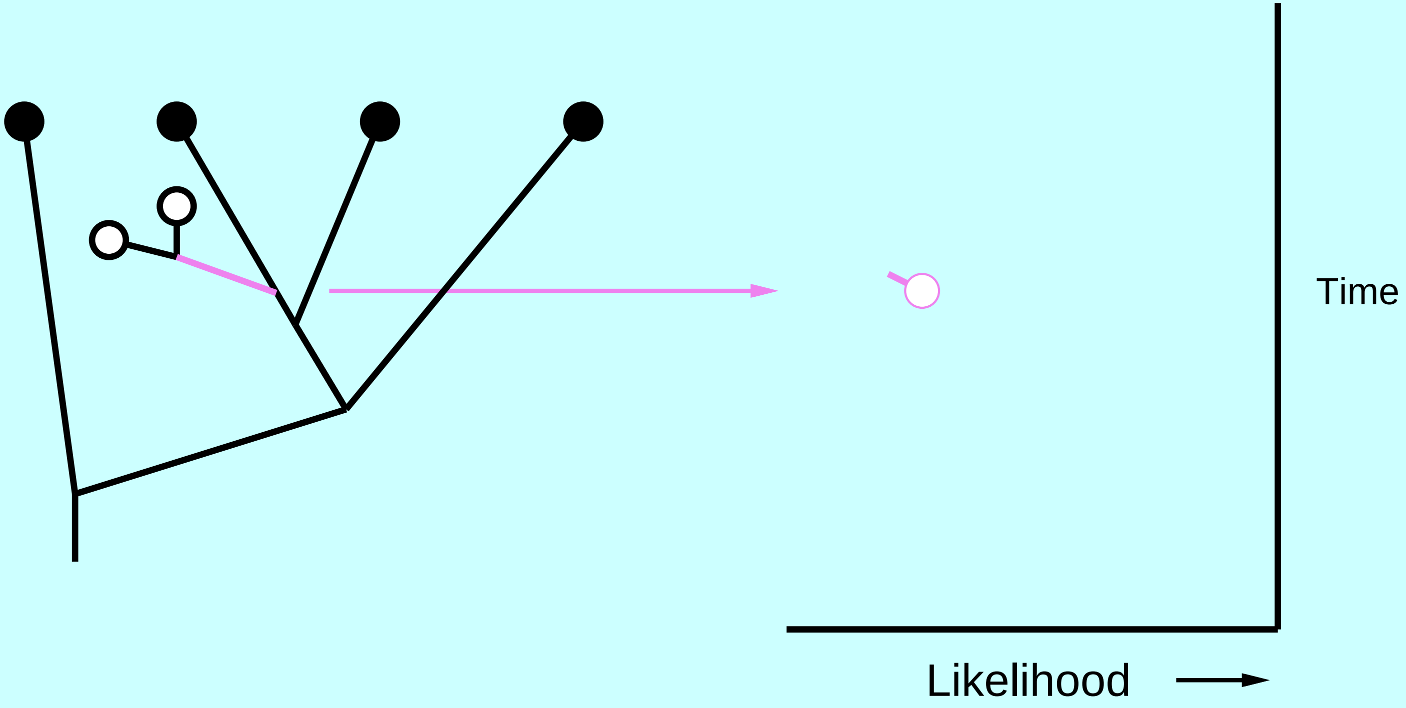
Infer placement of fossil species using their data

Using fossils



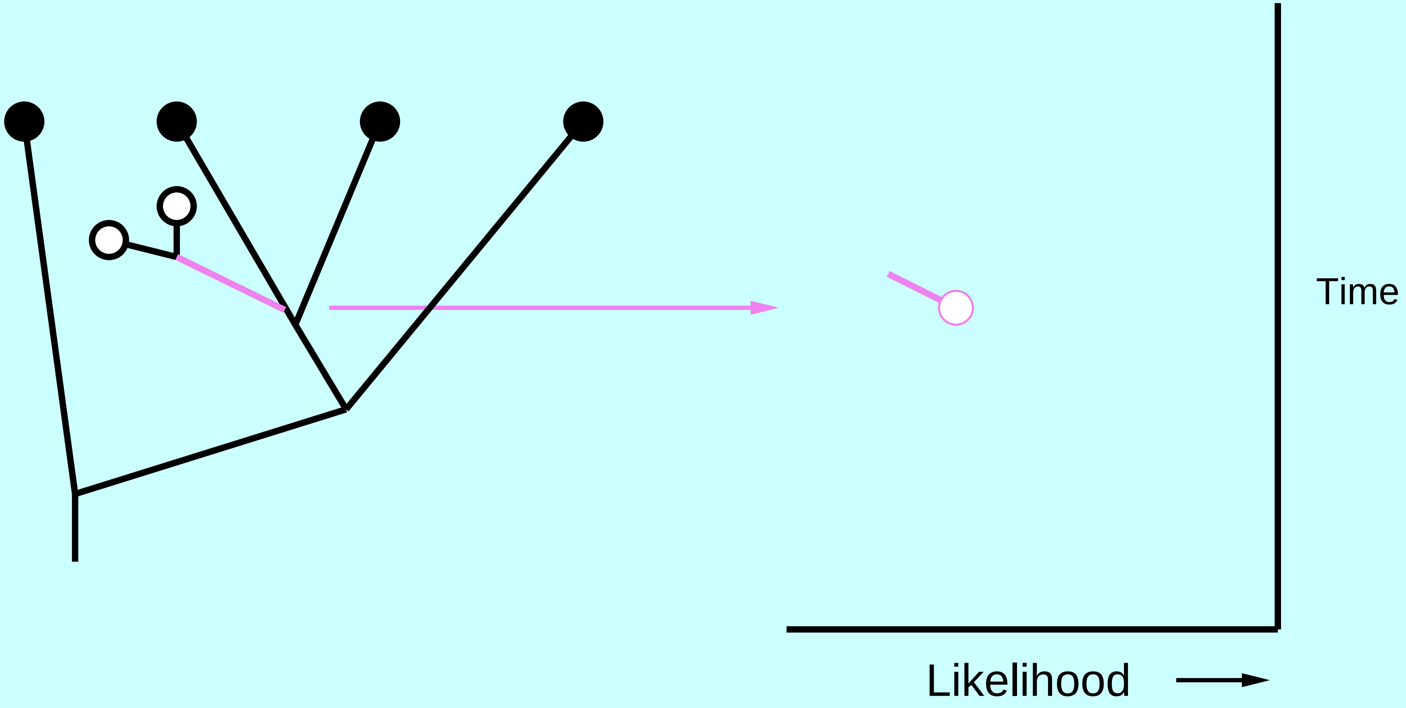
Use fossil and present-day morphology, covariances, tree,
also stratigraphic models

Using fossils



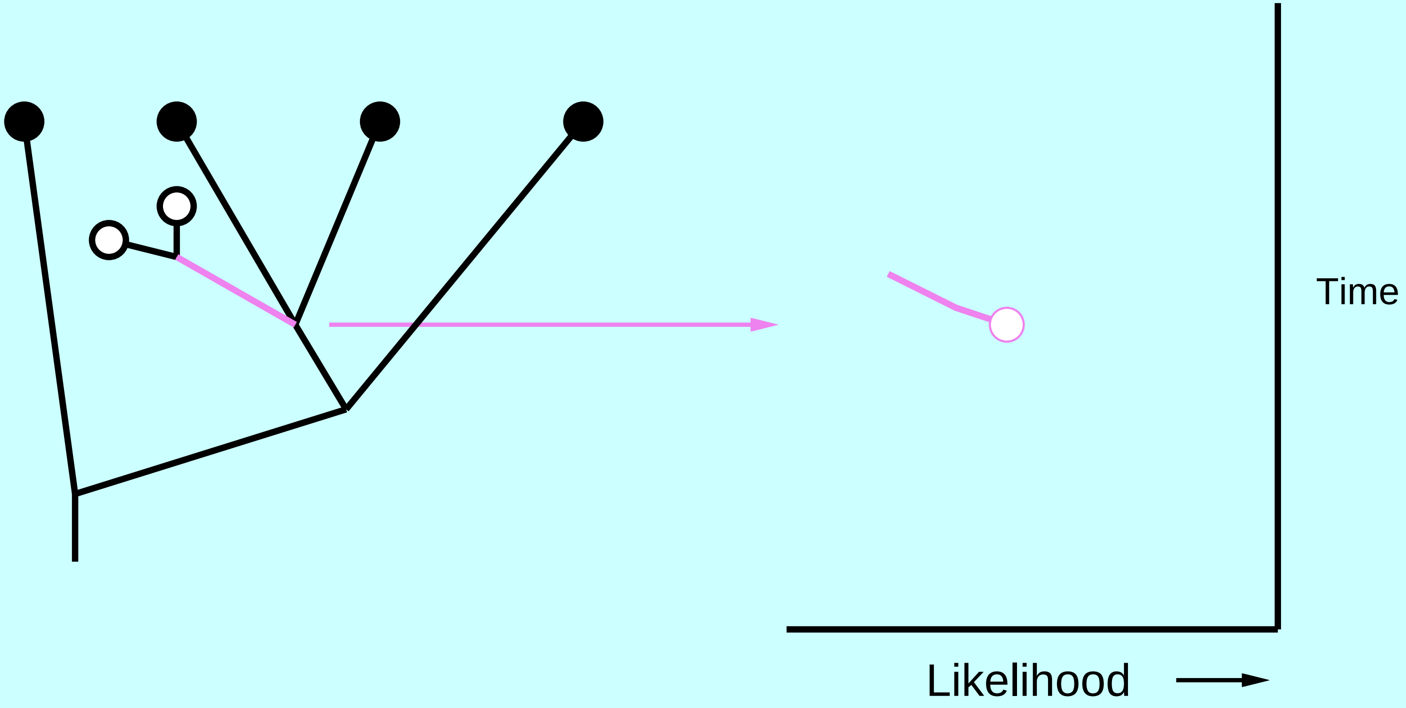
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



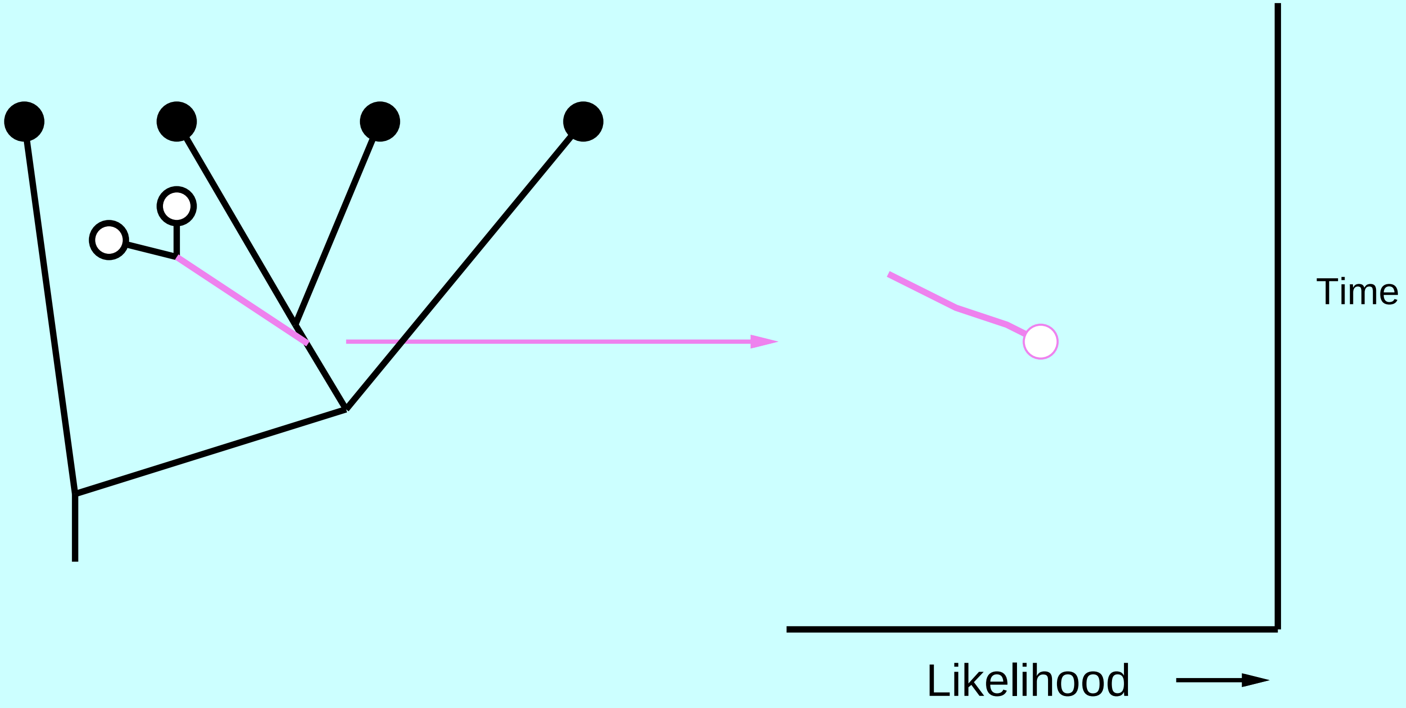
Use fossil and present-day morphology, covariances, tree,
also stratigraphic models

Using fossils



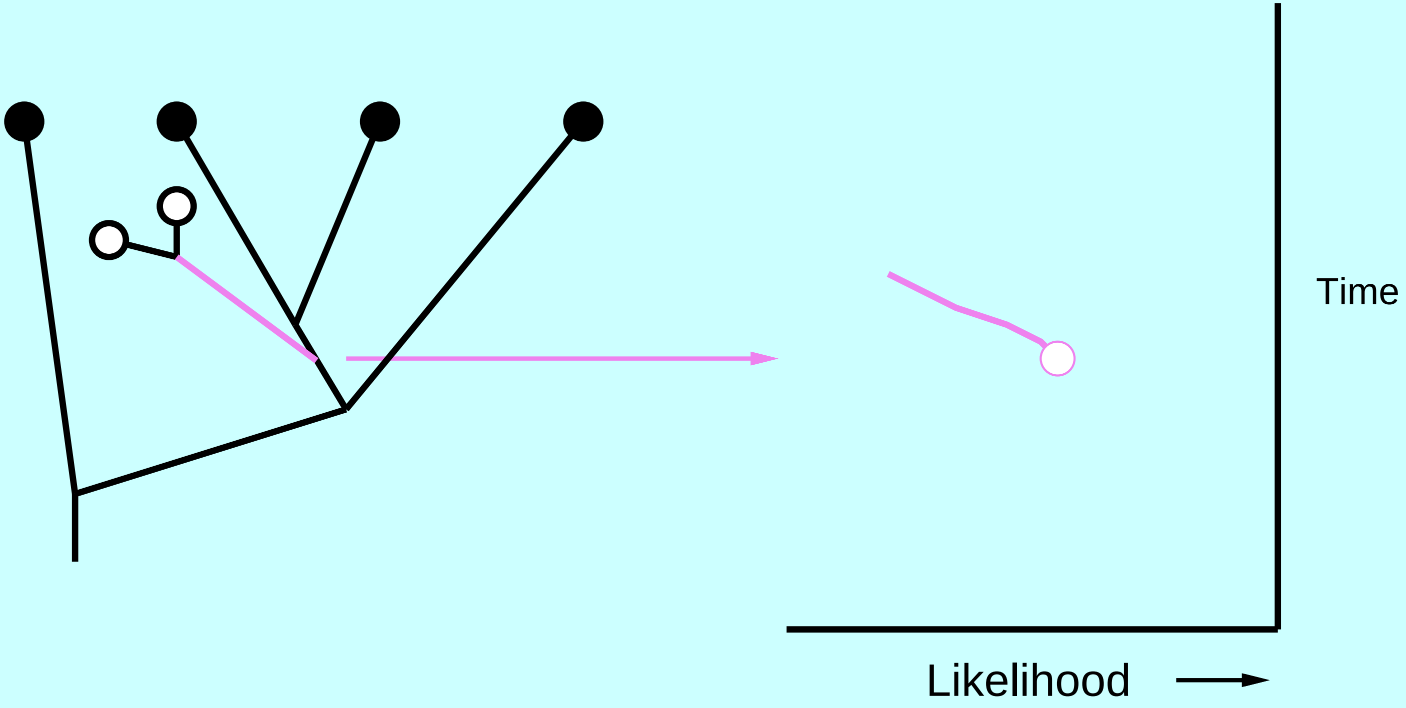
Use fossil and present-day morphology, covariances, tree,
also stratigraphic models

Using fossils



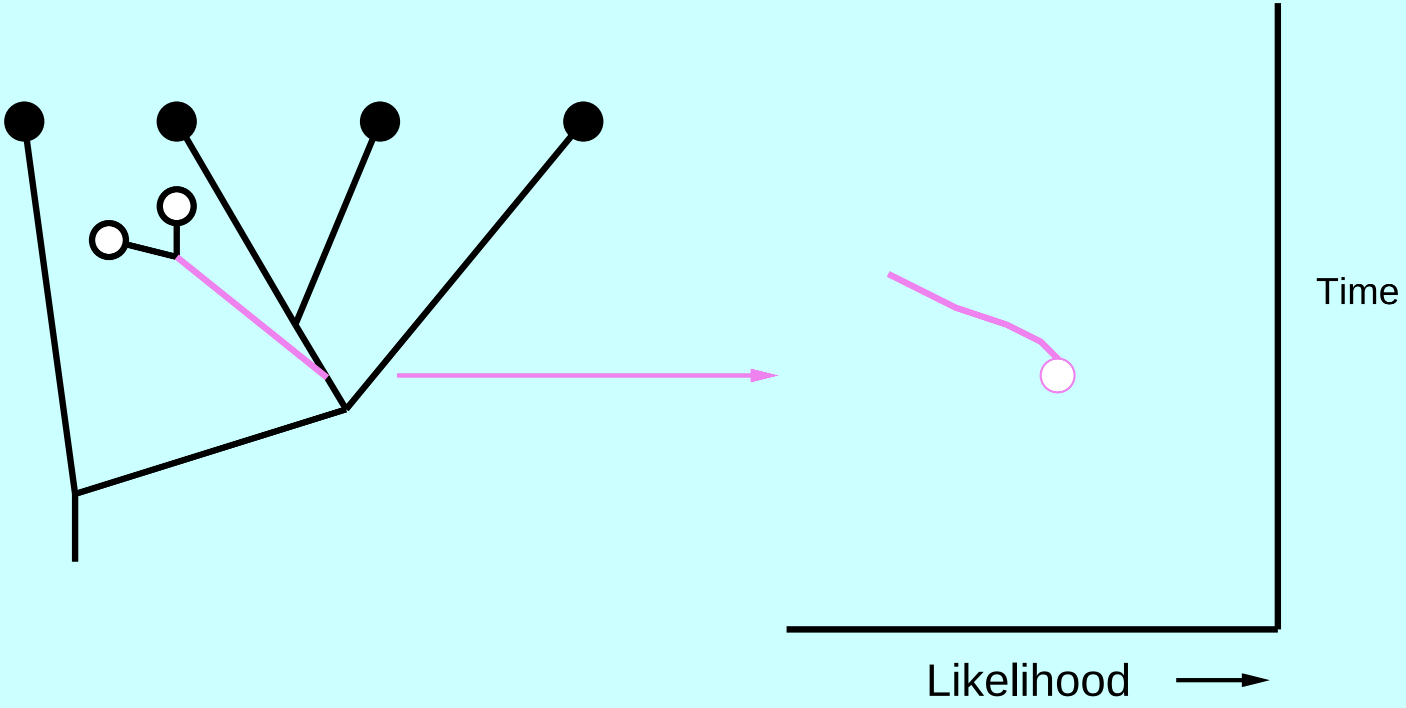
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



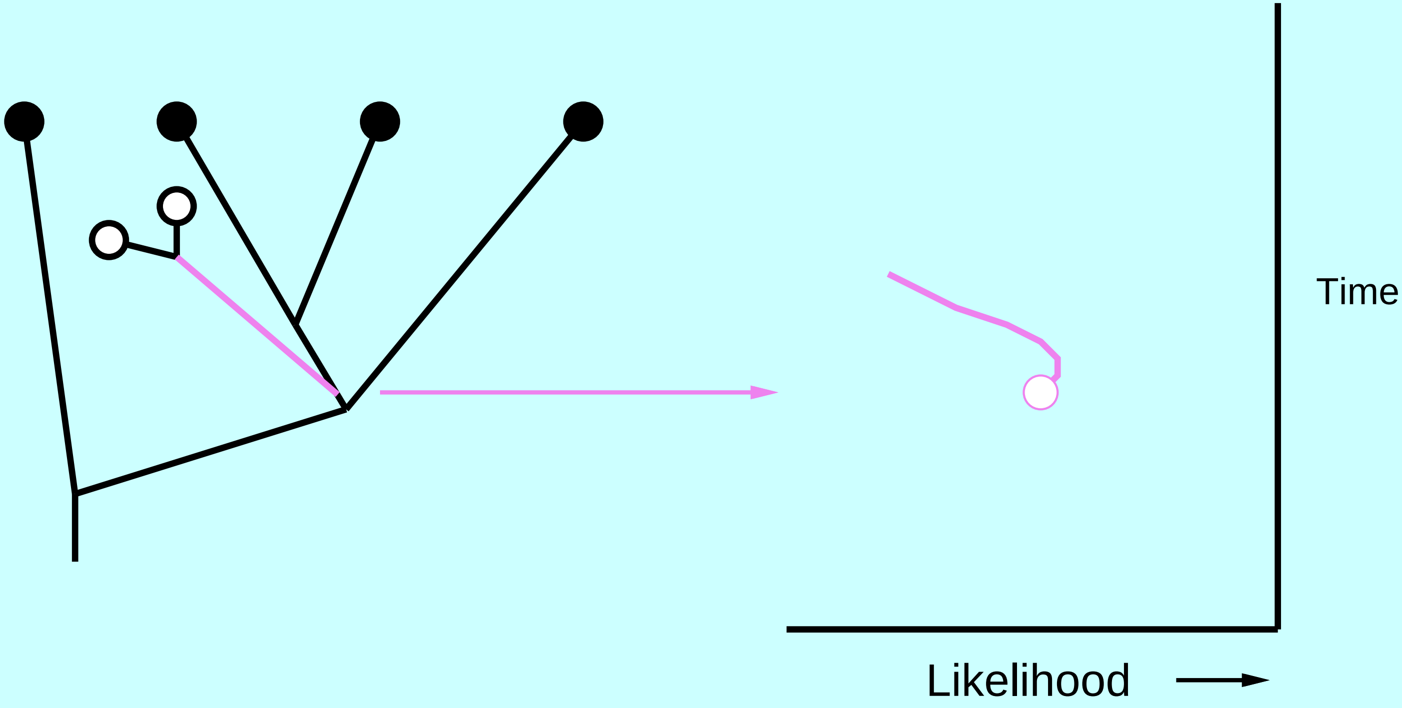
Use fossil and present-day morphology, covariances, tree,
also stratigraphic models

Using fossils



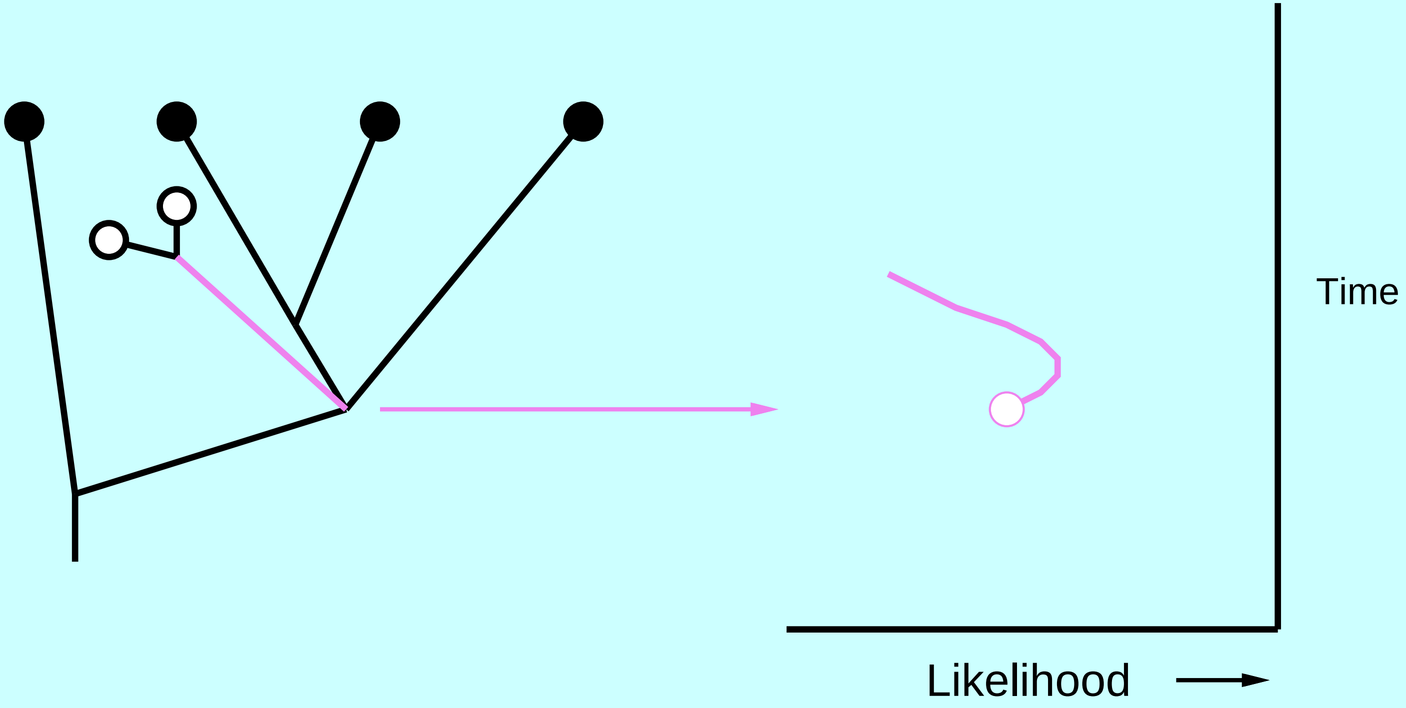
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



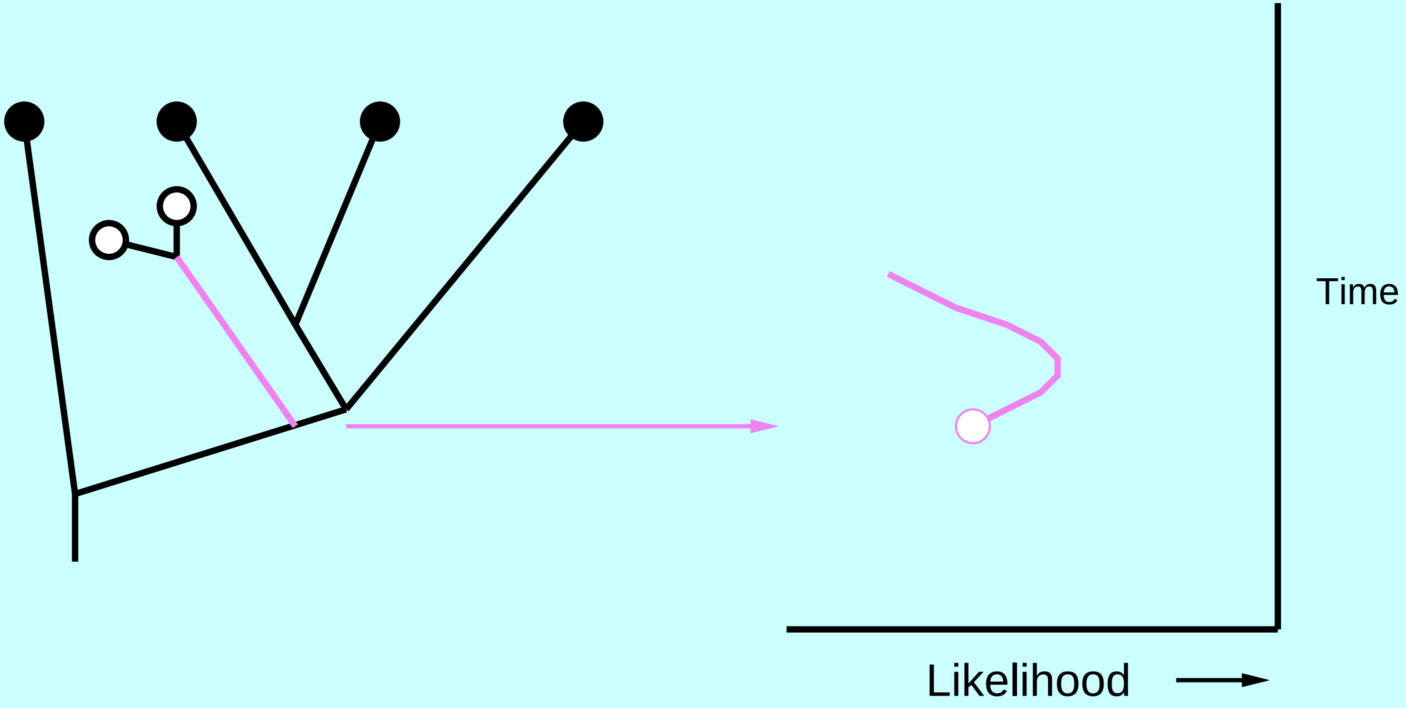
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



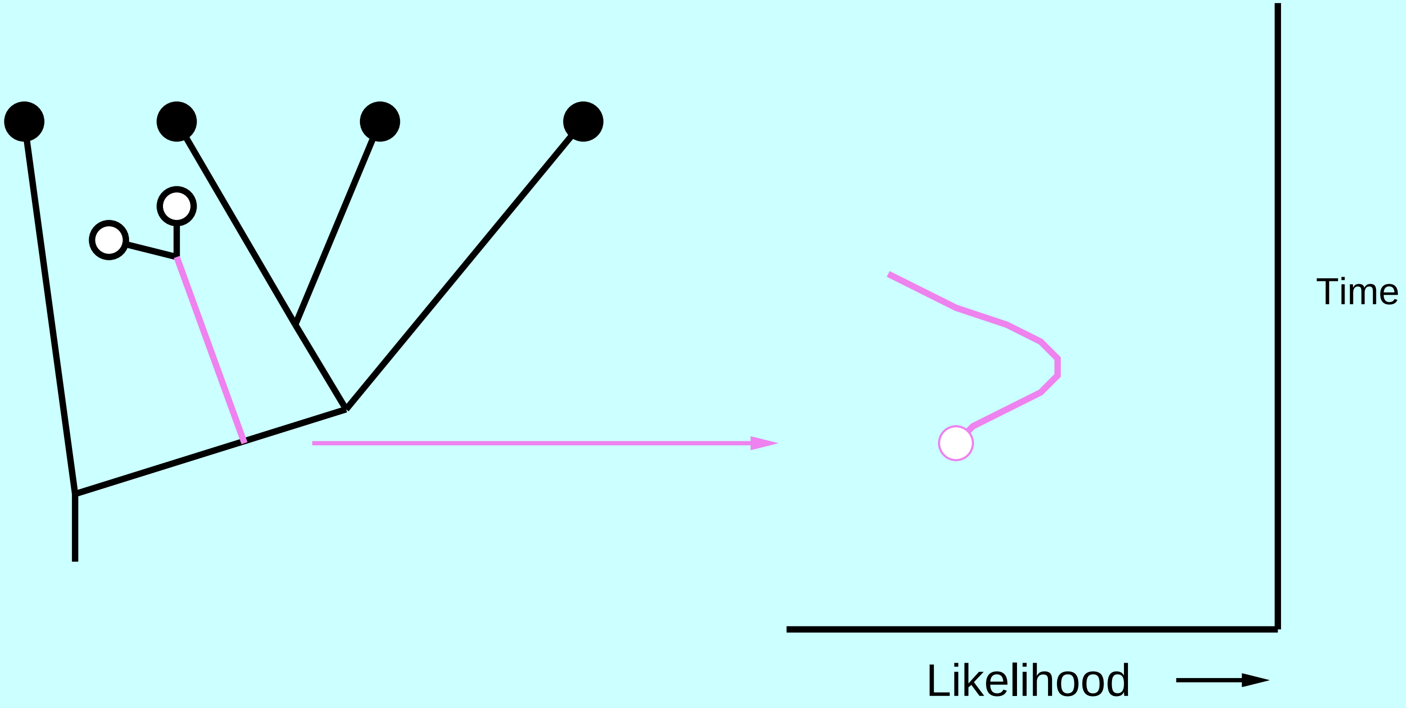
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



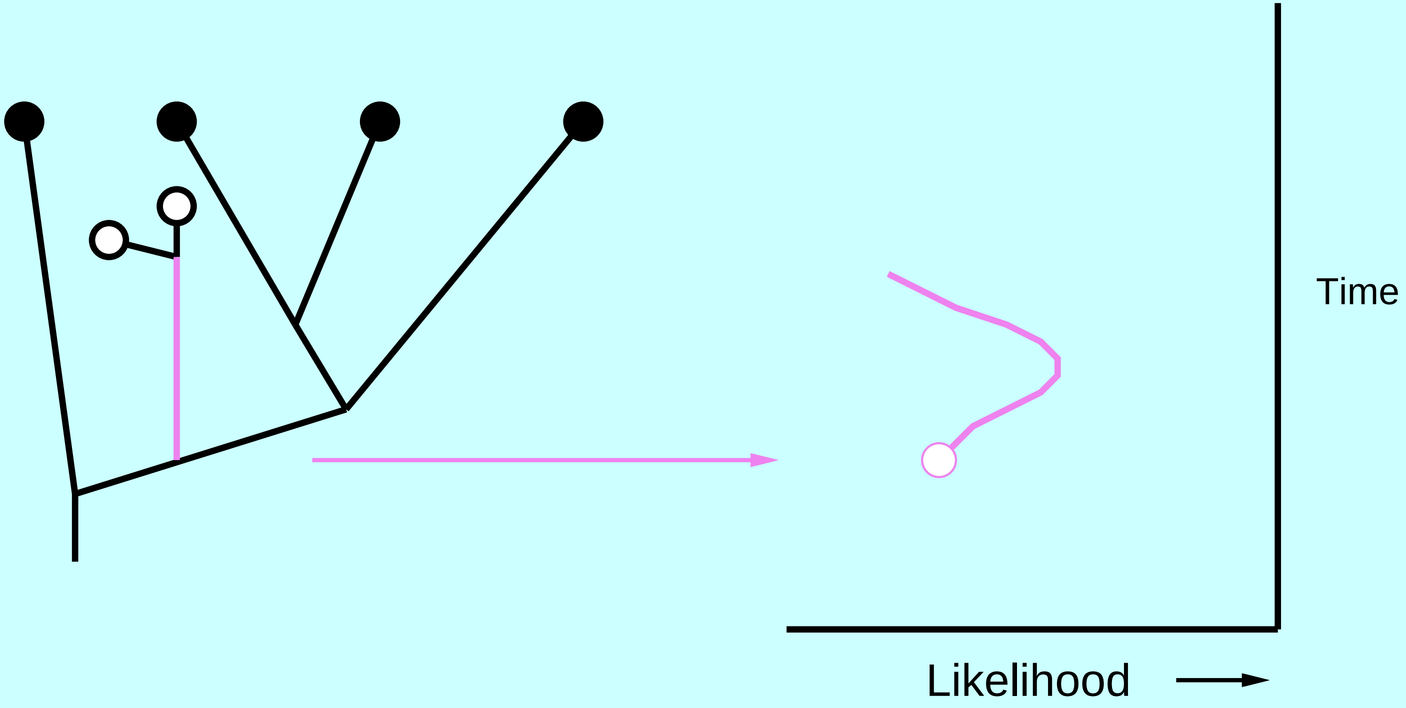
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



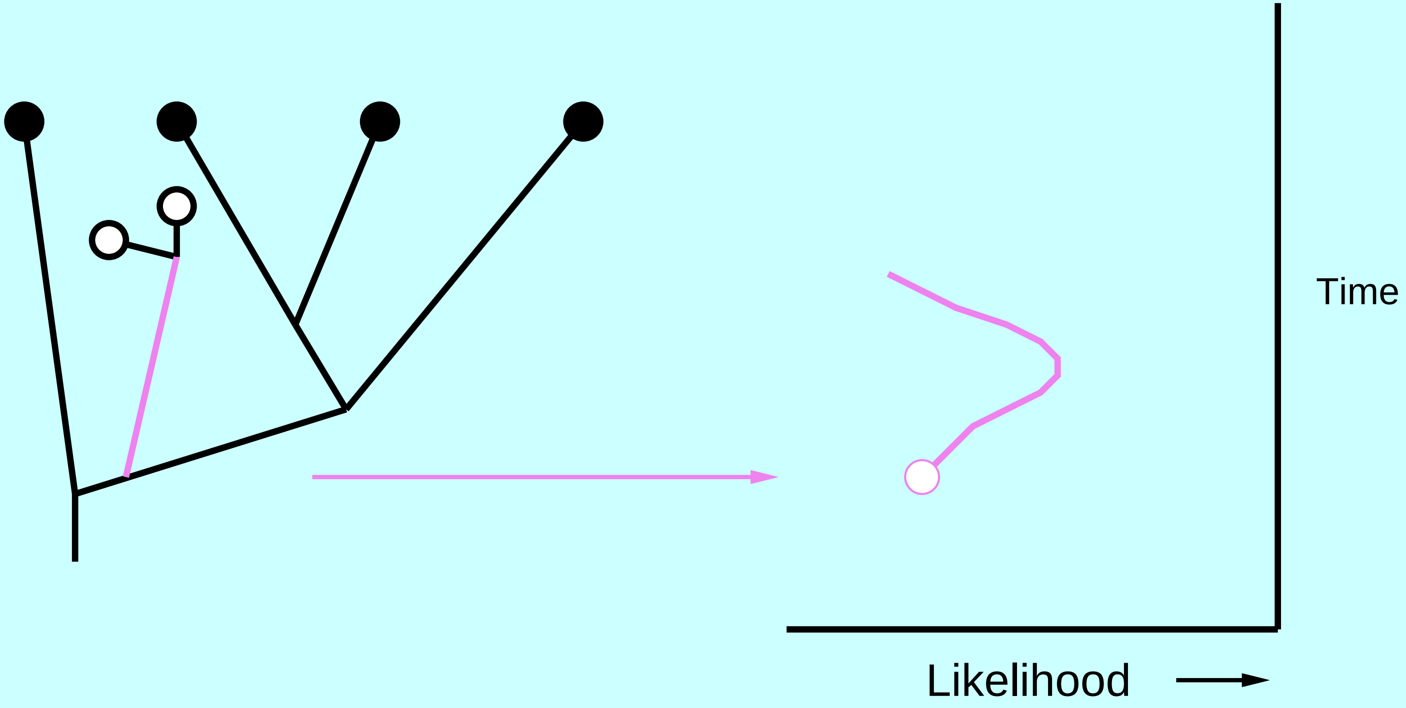
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



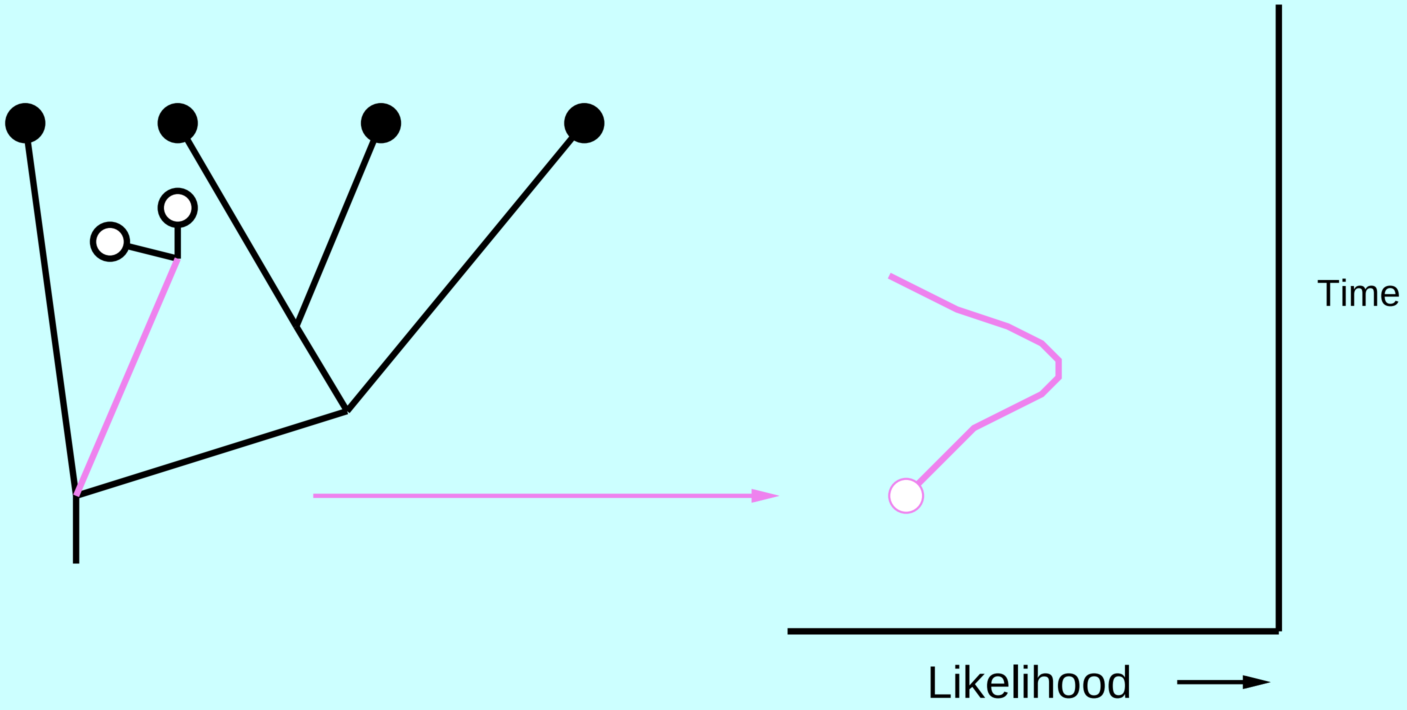
Use fossil and present-day morphology, covariances, tree,
also stratigraphic models

Using fossils



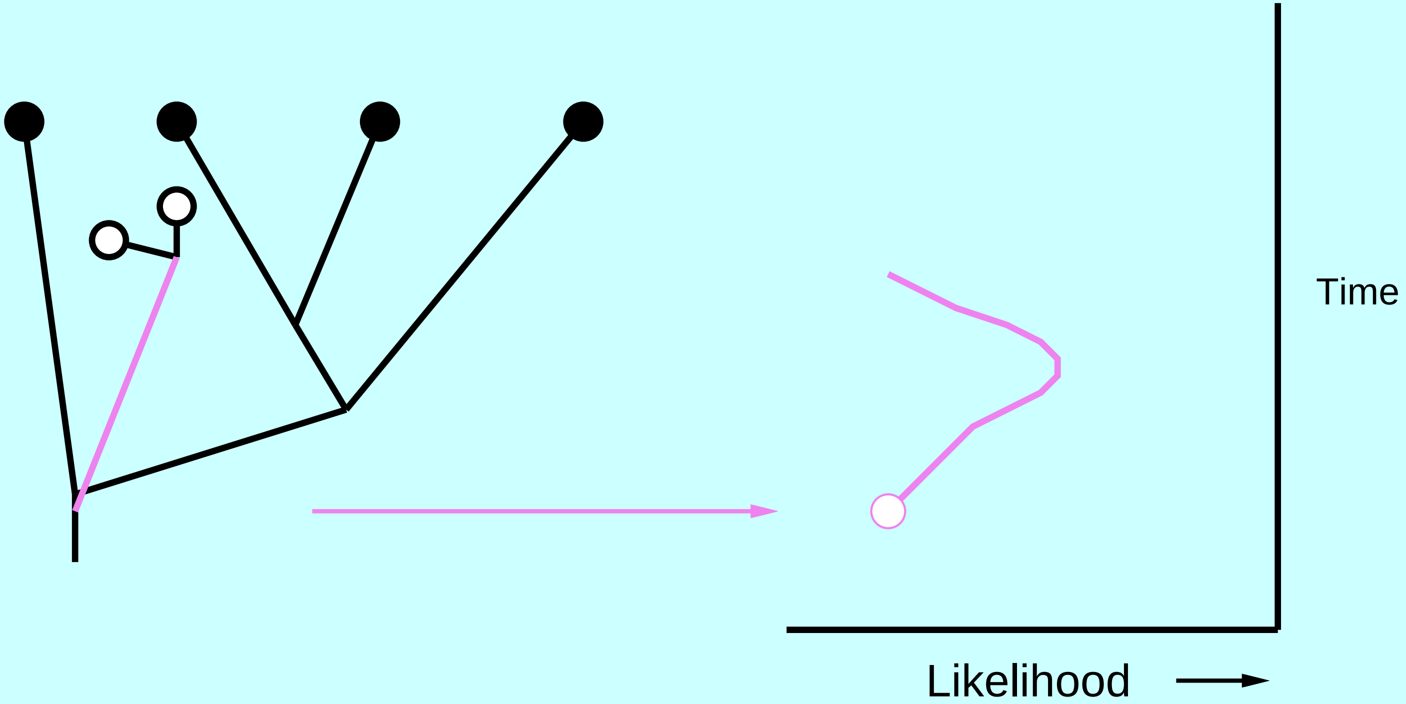
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



Use fossil and present-day morphology, covariances, tree,
also stratigraphic models

Using fossils



Use fossil and present-day morphology, covariances, tree, also stratigraphic models

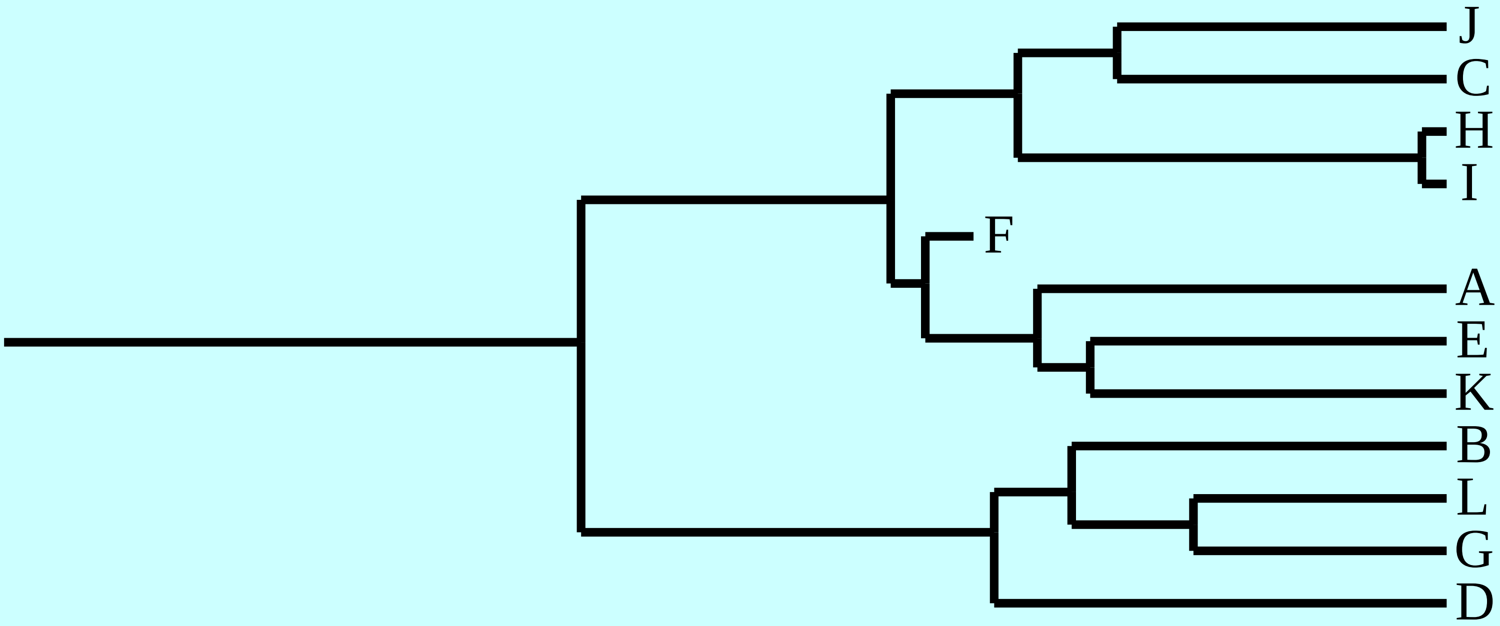
A simple result

The upshot is that to find the maximum likelihood placement of a fossil lineage, we

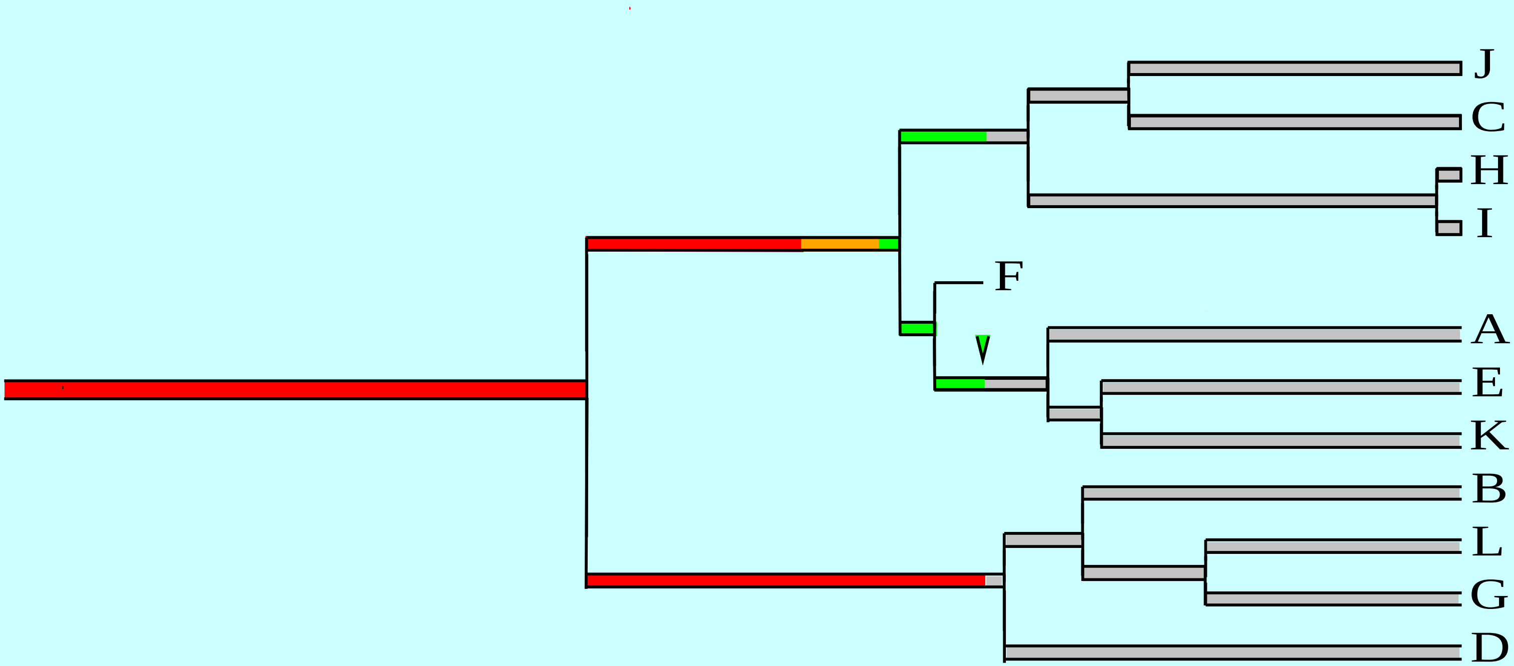
- Hook it up somewhere
- Obtain the contrasts for that tree
- Infer the phylogenetic covariances of the characters from the contrasts
- The log-likelihood for this placement is (a constant plus) $-(n-1)/2$ times the log of the determinant of the covariance matrix, minus a penalty which depends on the sum of the logs of the standard deviations of the contrasts.

So we minimize the determinant plus penalty to find the best placement. We can consider whether we can do likelihood ratio tests, too, at least for placement within a single branch.

An example: the true tree with F a fossil species

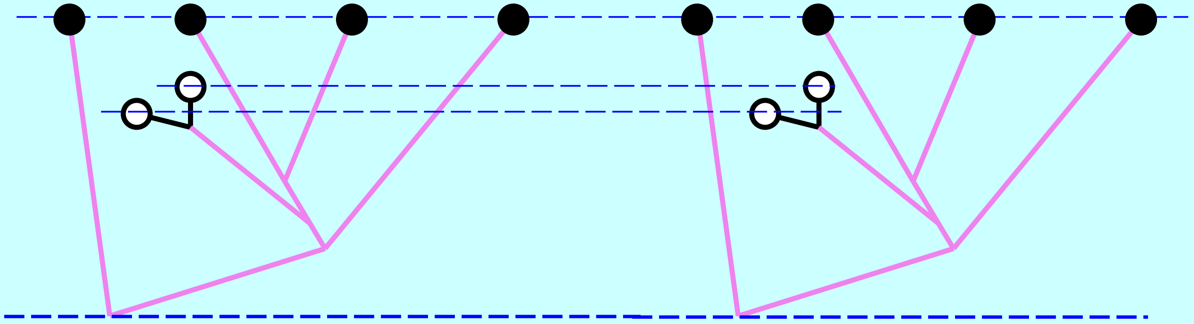


Traffic-light colors shows where fossil can be placed



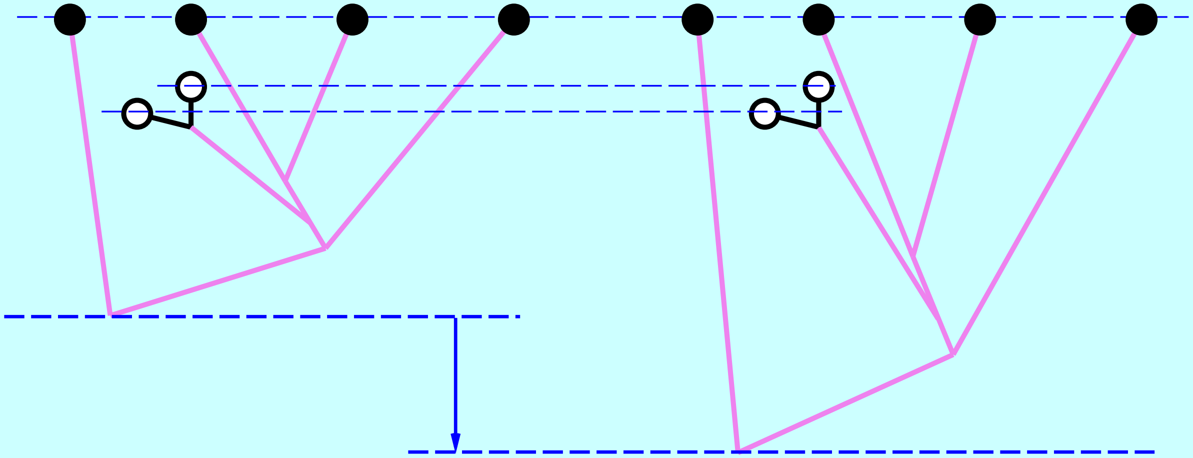
Green = within 1 log-likelihood unit, Orange = within 2 units, Red = lower than that. Green arrow is the ML placement. Gray placements are ruled out by date of the fossil.

Calibrating the molecular clock



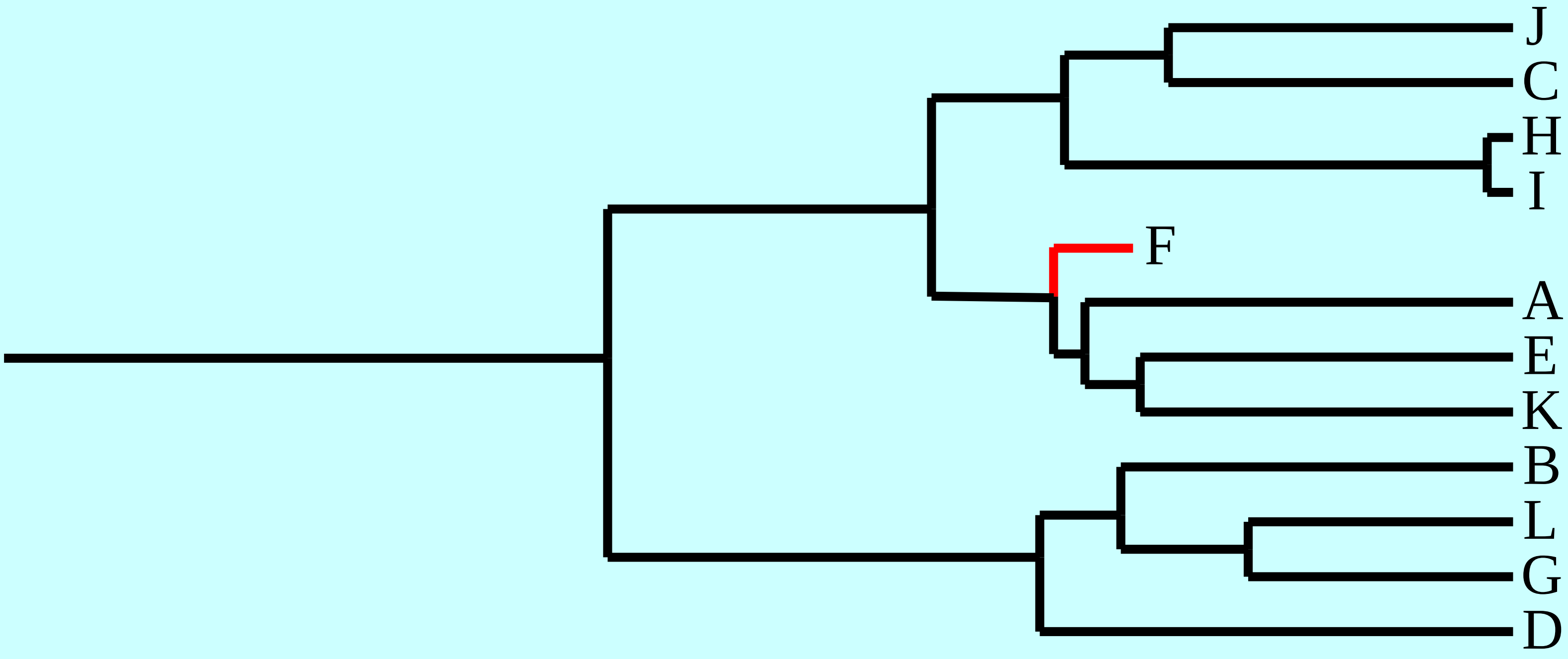
Molecular trees don’t usually have branch lengths on a time scale, and we need that. How to infer the calibration of the clock?

Calibrating the molecular clock



There will be two quantities to infer, the scaling of the molecular tree on the time scale, and the placement of the connection to the fossil. We make an ML estimate and accept other values that are not rejected by a Likelihood Ratio Test with 2 degrees of freedom.

Calibrating the molecular clock



For example if (not a real example) the placement of F turned out to be as shown, with the branch length shown in red, that in turn scales the whole molecular tree, as we know the time of F.

A qualification

- The present method takes the molecular tree as known.

A qualification

- The present method takes the molecular tree as known.
- Uncertainty in it could be modelled by doing the analysis multiple times on bootstrap samples (or Bayesian posterior samples) of the tree estimates.

A qualification

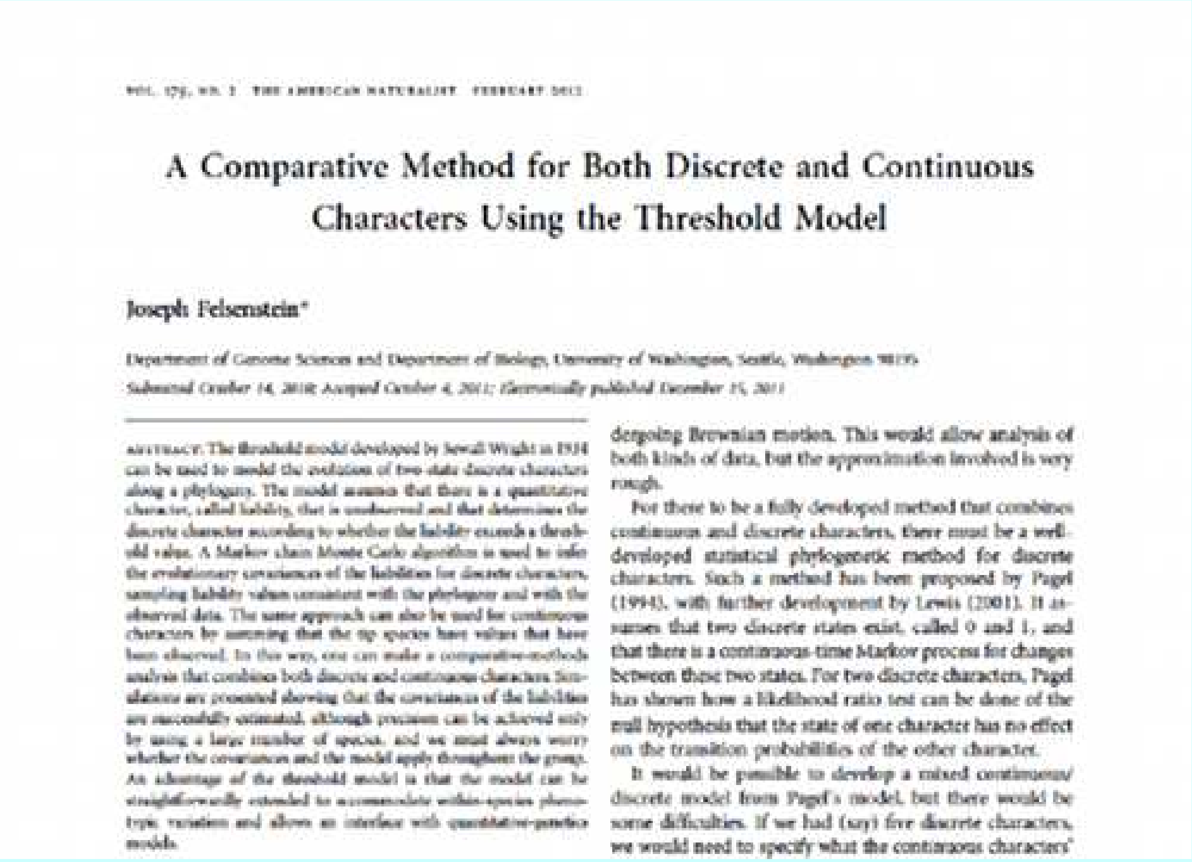
- The present method takes the molecular tree as known.
- Uncertainty in it could be modelled by doing the analysis multiple times on bootstrap samples (or Bayesian posterior samples) of the tree estimates.
- Pyron and Ronquist both use a more comprehensive “total evidence” approach of allowing the morphological data to influence Bayesian inference of the tree.

A qualification

- The present method takes the molecular tree as known.
- Uncertainty in it could be modelled by doing the analysis multiple times on bootstrap samples (or Bayesian posterior samples) of the tree estimates.
- Pyron and Ronquist both use a more comprehensive “total evidence” approach of allowing the morphological data to influence Bayesian inference of the tree.
- I suspect this will have little effect if there is a lot of molecular data, so I am sticking with this approach.

A third example: A threshold model for 0/1 characters

This was published in *American Naturalist* in 2012:



VOL. 179, NO. 2 THE AMERICAN NATURALIST FEBRUARY 2012

A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model

Joseph Felsenstein*

Department of Genome Sciences and Department of Biology, University of Washington, Seattle, Washington 98195

Submitted October 14, 2010; Accepted October 4, 2011; Electronically published December 15, 2011

ABSTRACT: The threshold model developed by Sewall Wright in 1934 can be used to model the evolution of two-state discrete characters along a phylogeny. The model assumes that there is a quantitative character, called liability, that is unobserved and that determines the discrete character according to whether the liability exceeds a threshold value. A Markov chain Monte Carlo algorithm is used to infer the evolutionary covariances of the liabilities for discrete characters, sampling liability values consistent with the phylogeny and with the observed data. The same approach can also be used for continuous characters by assuming that the tip species have values that have been observed. In this way, one can make a comparative-methods analysis that combines both discrete and continuous characters. Simulations are presented showing that the covariances of the liabilities are successfully estimated, although precision can be achieved only by using a large number of species, and we must always worry whether the covariances and the model apply throughout the group. An advantage of the threshold model is that the model can be straightforwardly extended to accommodate within-species phenotypic variation and allows an interface with quantitative-genetics models.

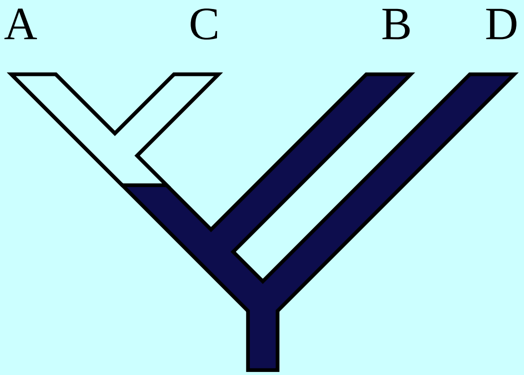
dergoing Brownian motion. This would allow analysis of both kinds of data, but the approximation involved is very rough.

For there to be a fully developed method that combines continuous and discrete characters, there must be a well-developed statistical phylogenetic method for discrete characters. Such a method has been proposed by Pagel (1994), with further development by Lewis (2001). It assumes that two discrete states exist, called 0 and 1, and that there is a continuous-time Markov process for changes between these two states. For two discrete characters, Pagel has shown how a likelihood ratio test can be done of the null hypothesis that the state of one character has no effect on the transition probabilities of the other character.

It would be possible to develop a mixed continuous/discrete model from Pagel's model, but there would be some difficulties. If we had (say) five discrete characters, we would need to specify what the continuous characters'

Current methods for statistical treatment of 0/1 characters

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

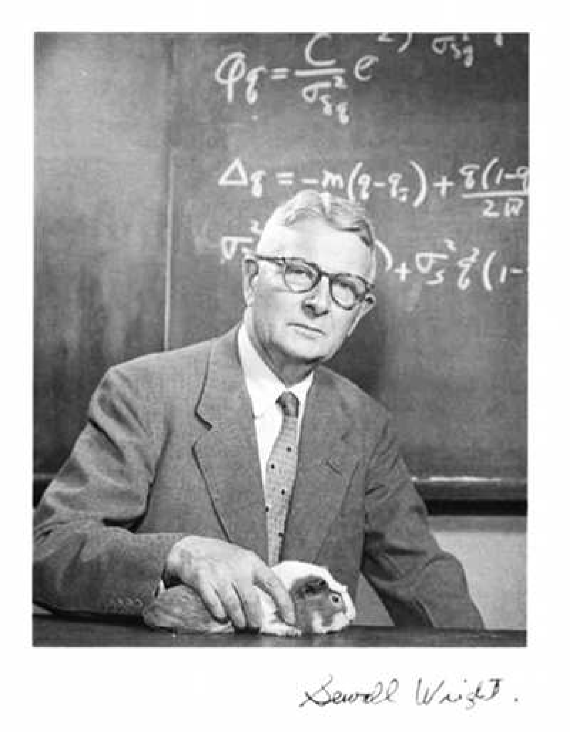


$$0 \underset{\beta\, dt}{\overset{\alpha\, dt}{\rightleftarrows}} 1$$

Pagel allows inference of whether change is correlated, on a known tree.
Lewis infers the tree, but does not allow for correlations among characters.
Neither takes into account contributions to a 0/1 character from multiple underlying loci.

A better model: threshold model

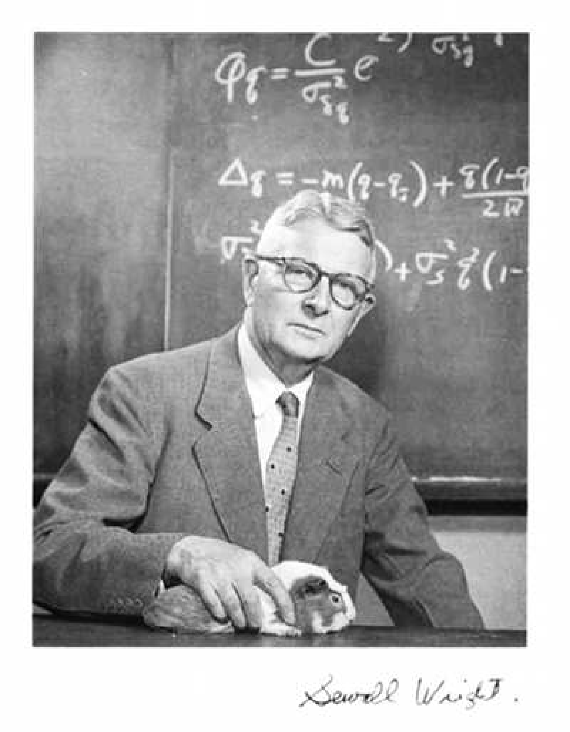
A relevant model was invented in 1934 by



Sewall Wright (1889-1988)
shown here in 1954

The threshold model

A relevant model was invented in 1934 by



Sewall Wright (1889-1988)

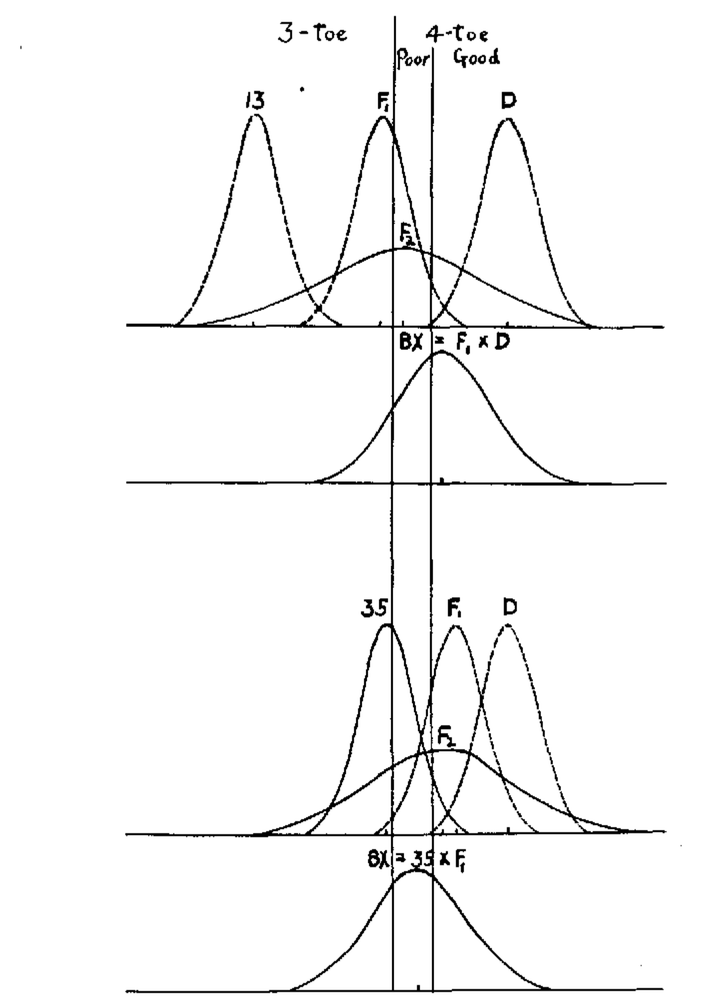
rumor has it he then turned and absent-mindedly
erased the board with the guinea pig

Sewall Wright, at the University of Chicago, 1928

In 1928

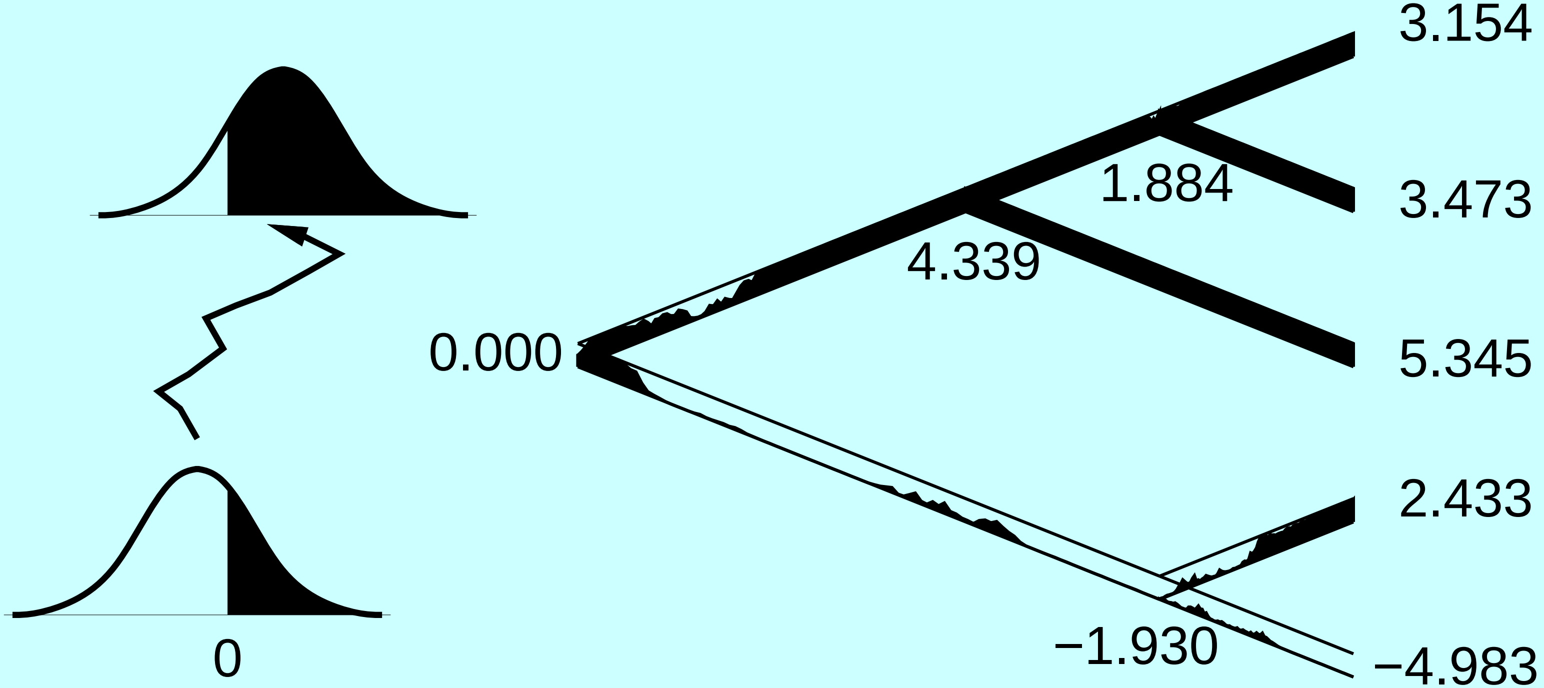
Same place
in May, 2013

The threshold model



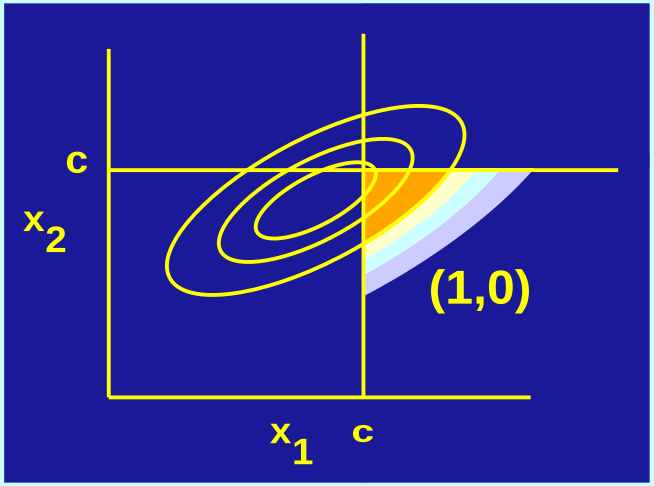
Sewall Wright (1934), guinea pig digit number
(from Wright's follow-up 1934 second paper)

The threshold model on a tree



Computing the likelihood

With two species, one character:



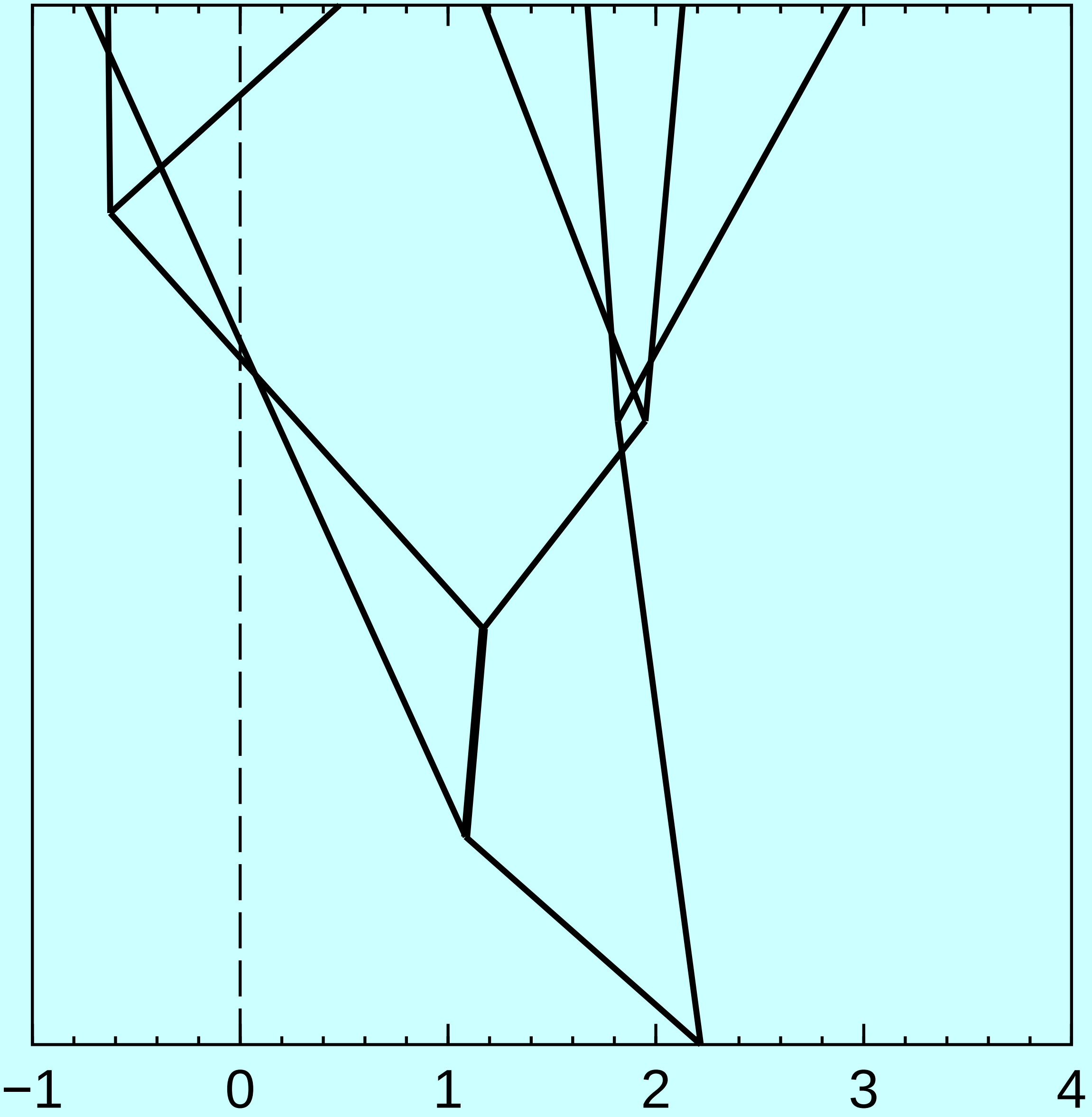
Disadvantages:

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

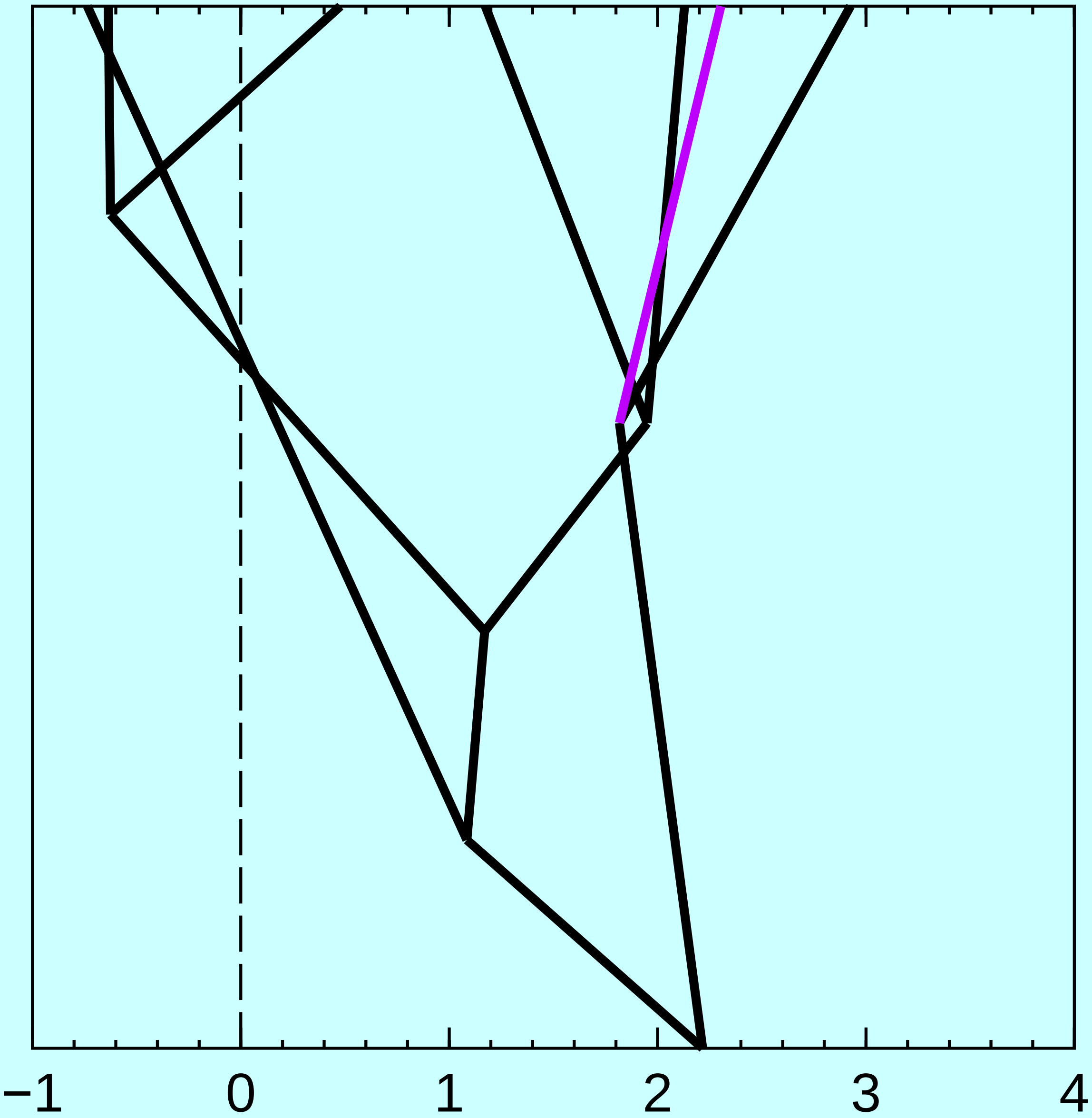
With 5 species, one character:

$$
\begin{aligned}
\mathsf{L} &= \mathrm{Prob}\,(1,1,0,1,1) \\
&= \int_0^\infty \int_0^\infty \int_{-\infty}^0 \int_0^\infty \int_0^\infty \varphi(\mathsf{x}_1,\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4,\mathsf{x}_5 \mid \mathrm{Tree})\ \mathsf{dx}_1\ \mathsf{dx}_2\ \mathsf{dx}_3\ \mathsf{dx}_4\ \mathsf{dx}_5
\end{aligned}
$$

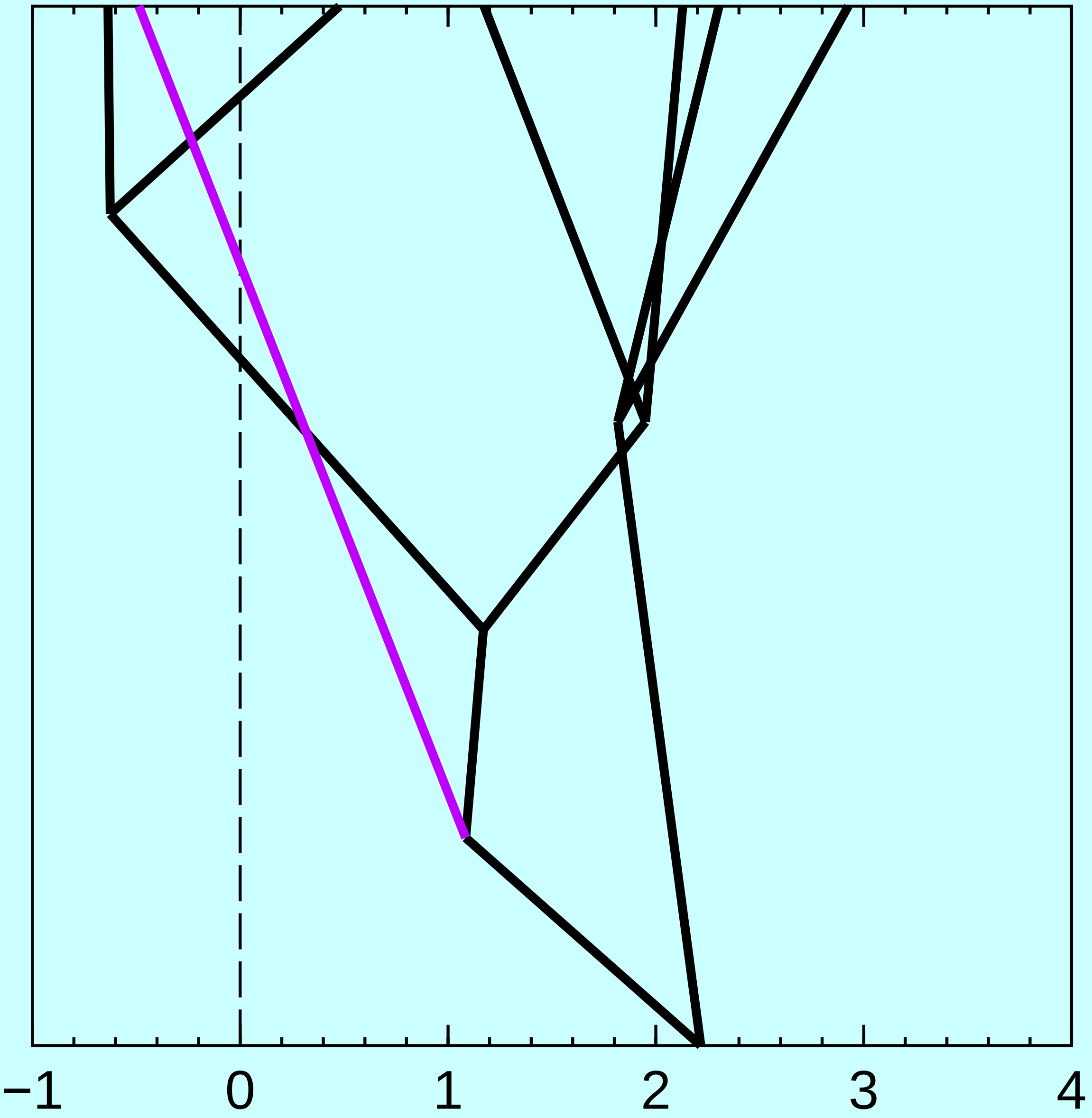
MCMC sampling of liabilities



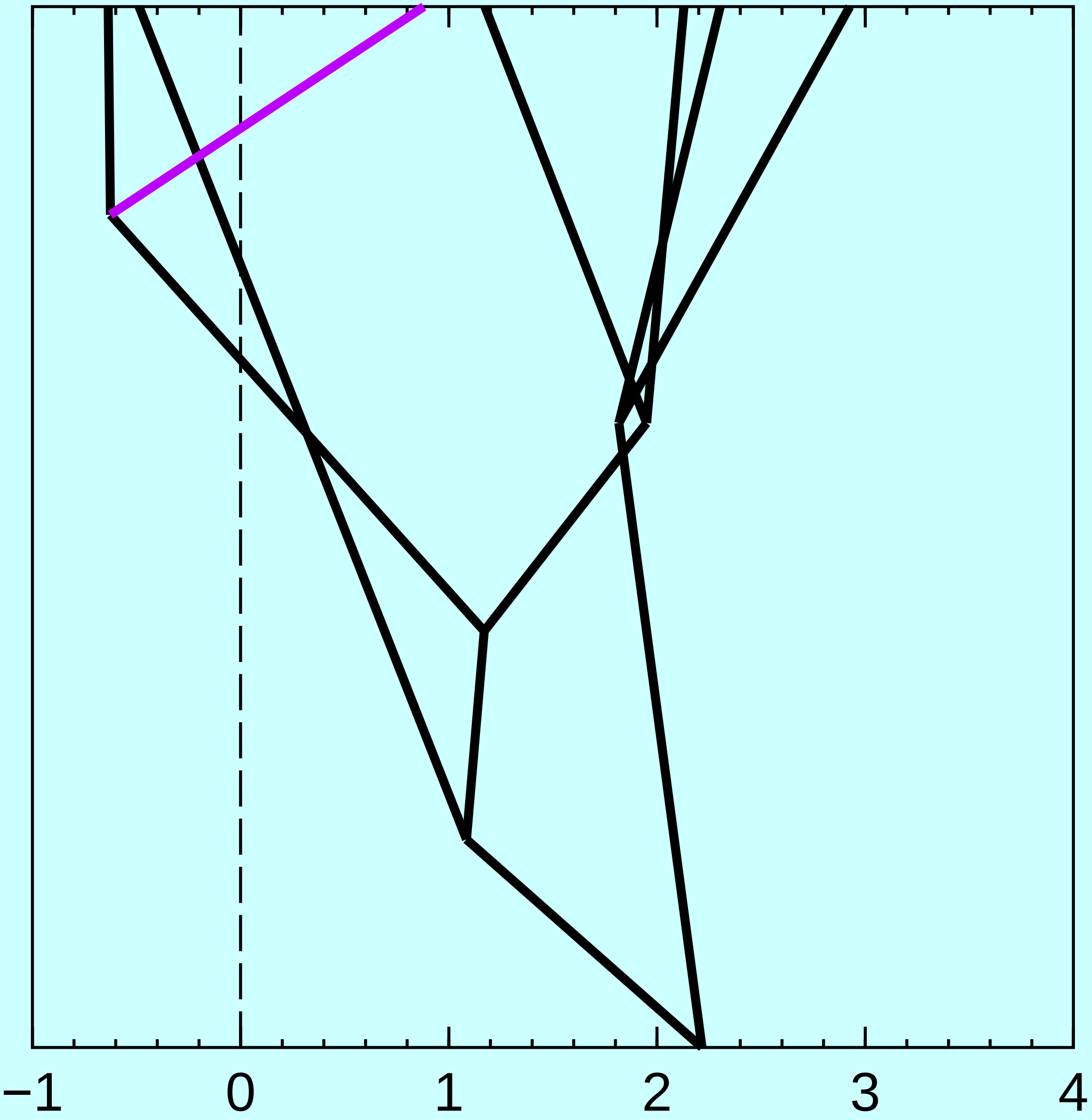
MCMC sampling of liabilities



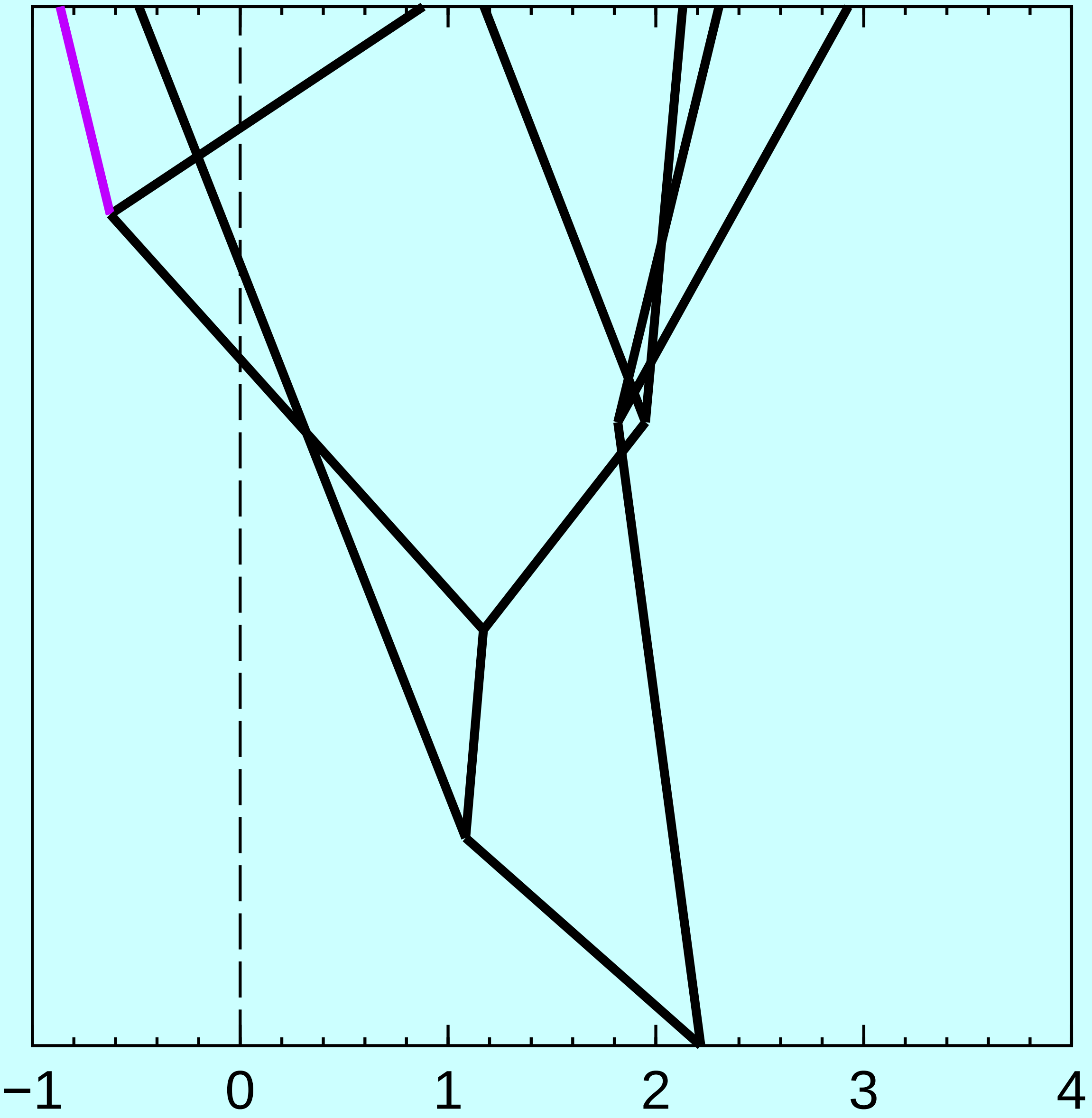
MCMC sampling of liabilities



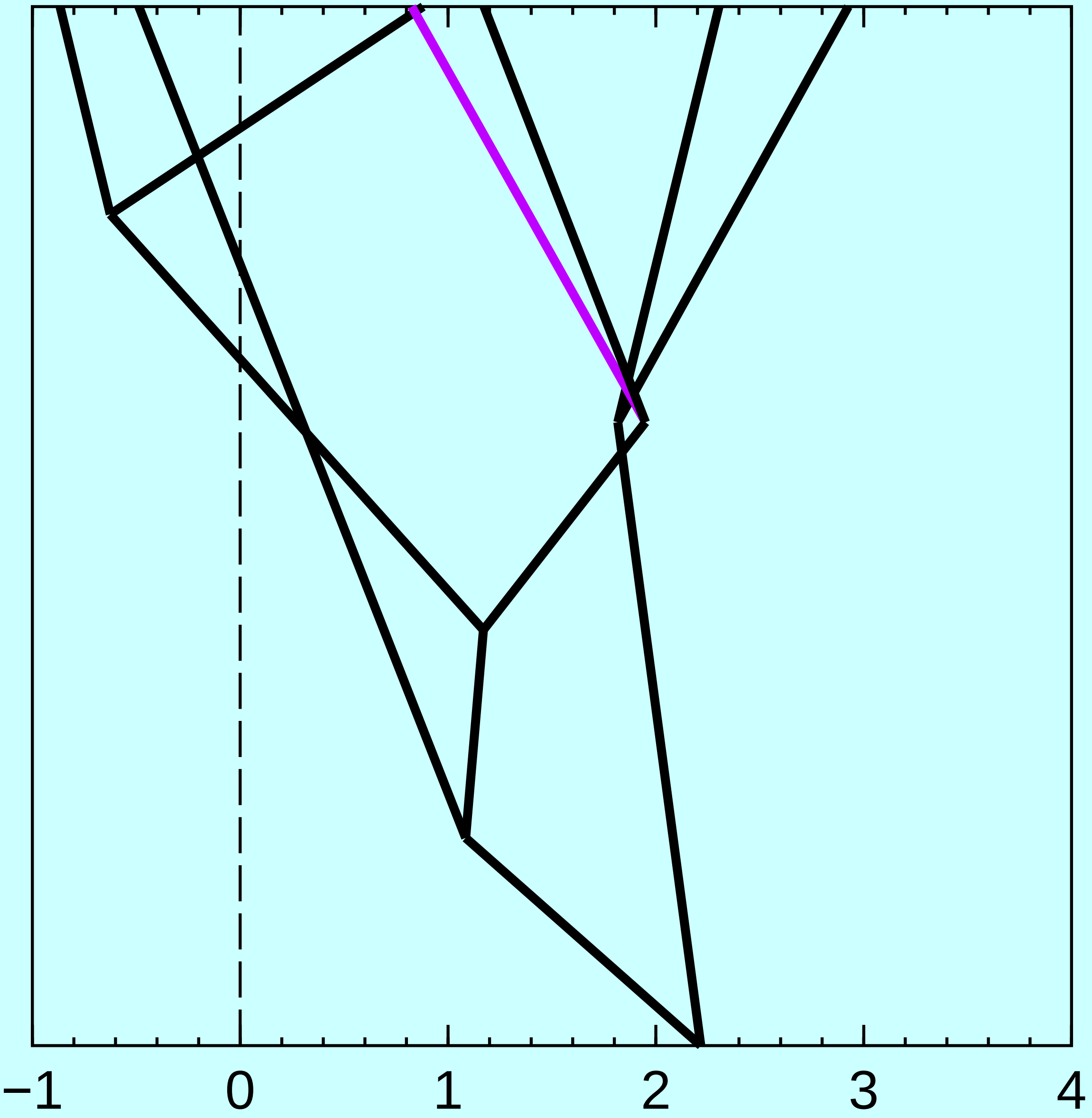
MCMC sampling of liabilities



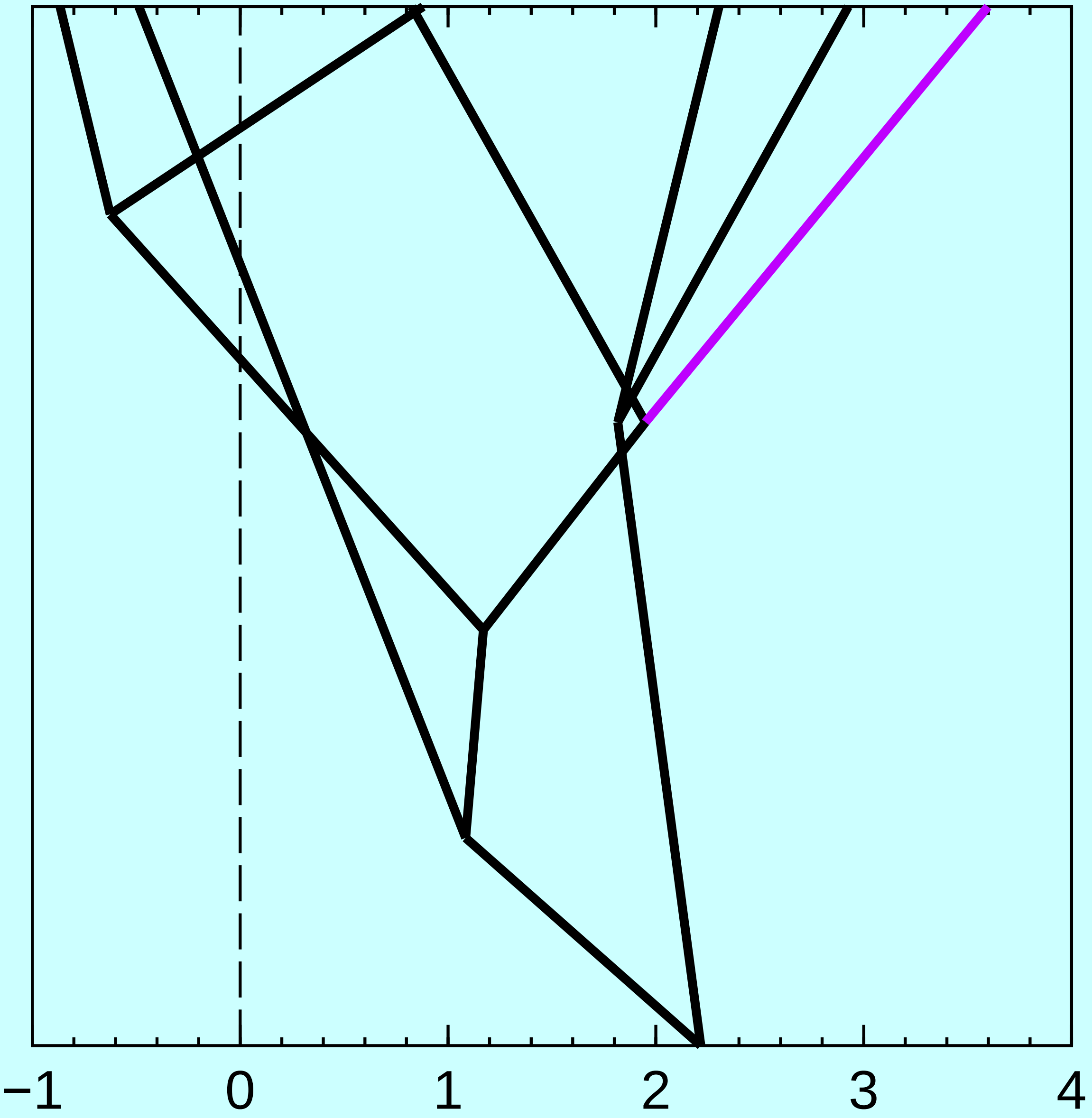
MCMC sampling of liabilities



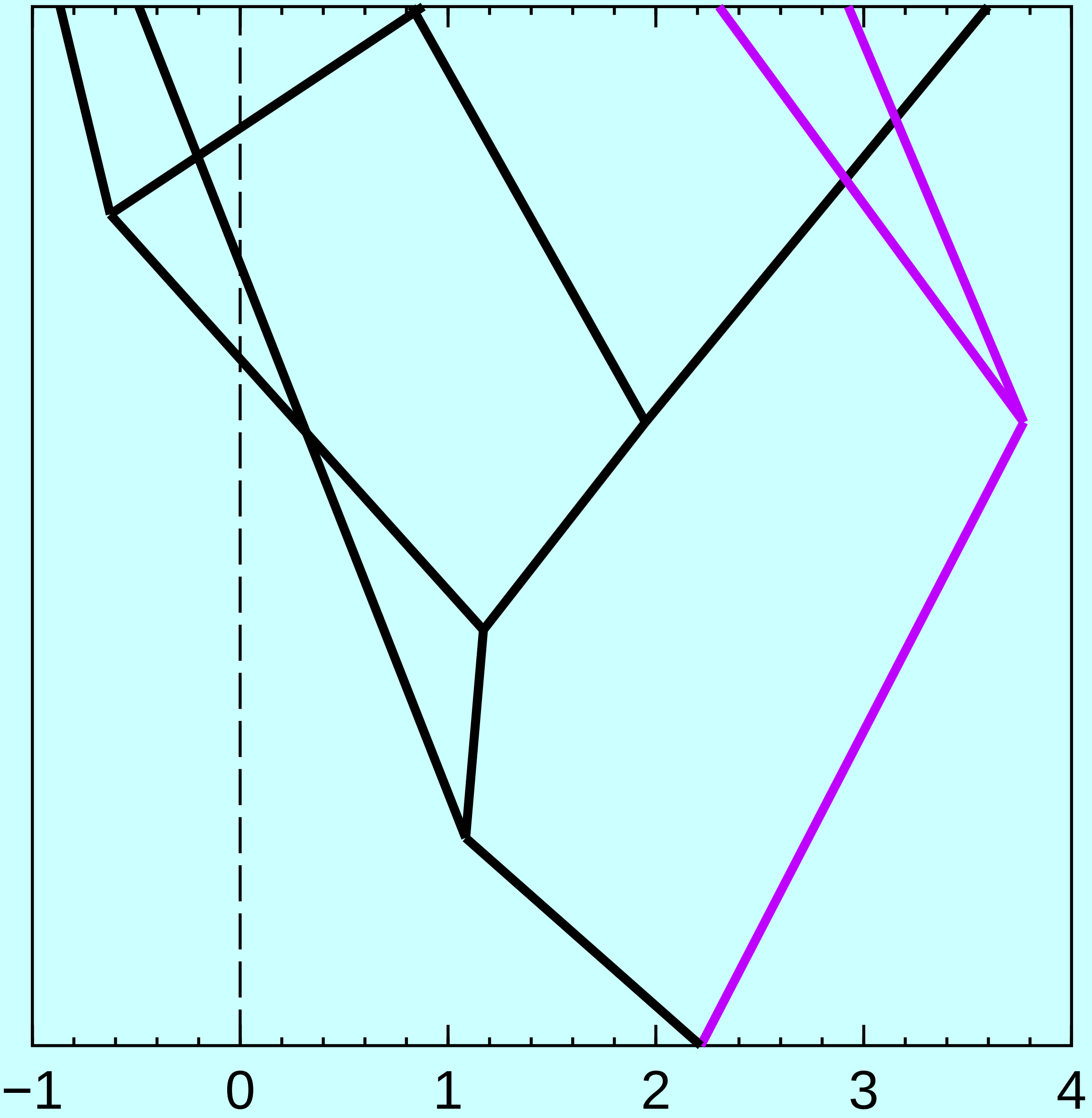
MCMC sampling of liabilities



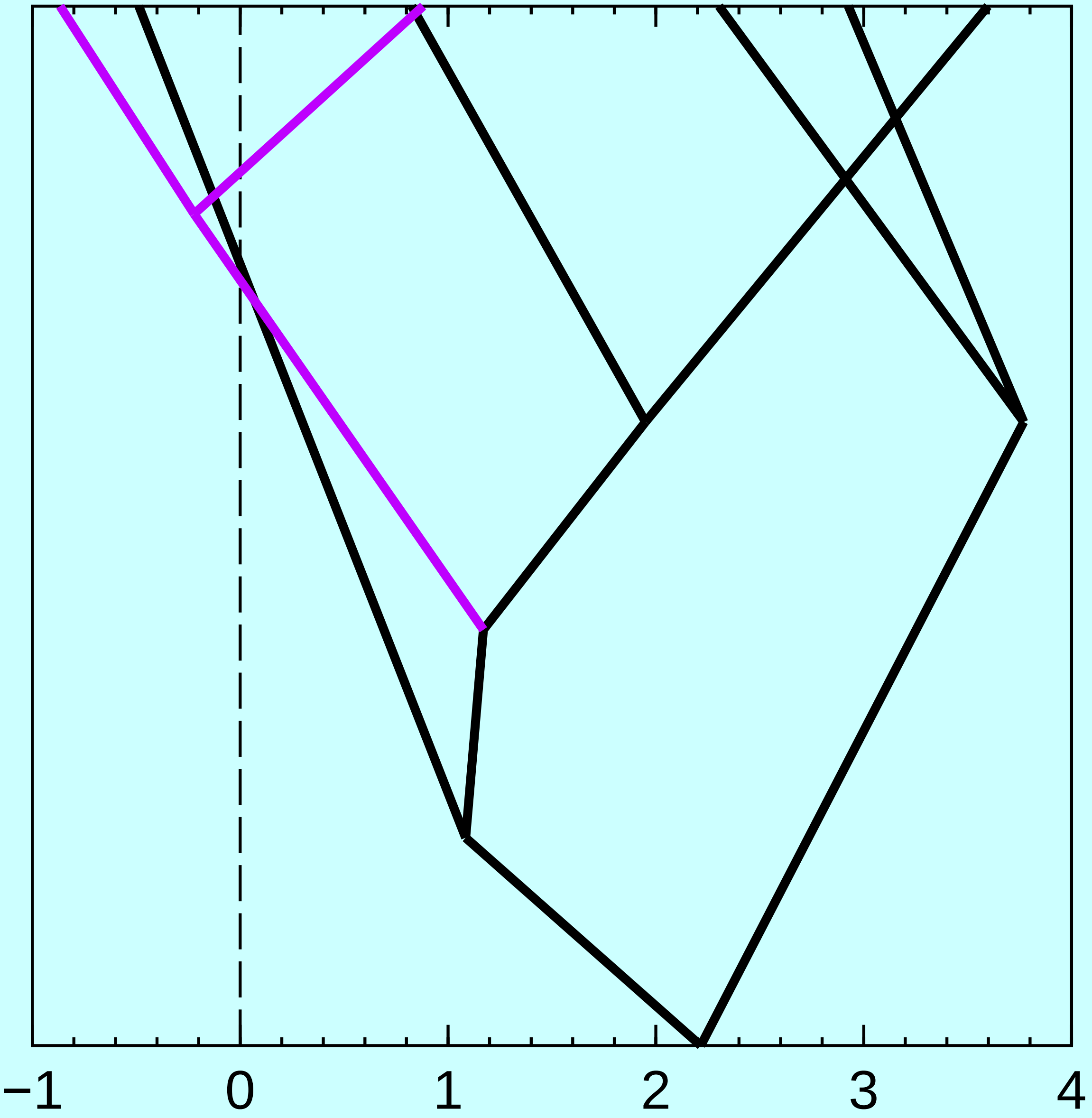
MCMC sampling of liabilities



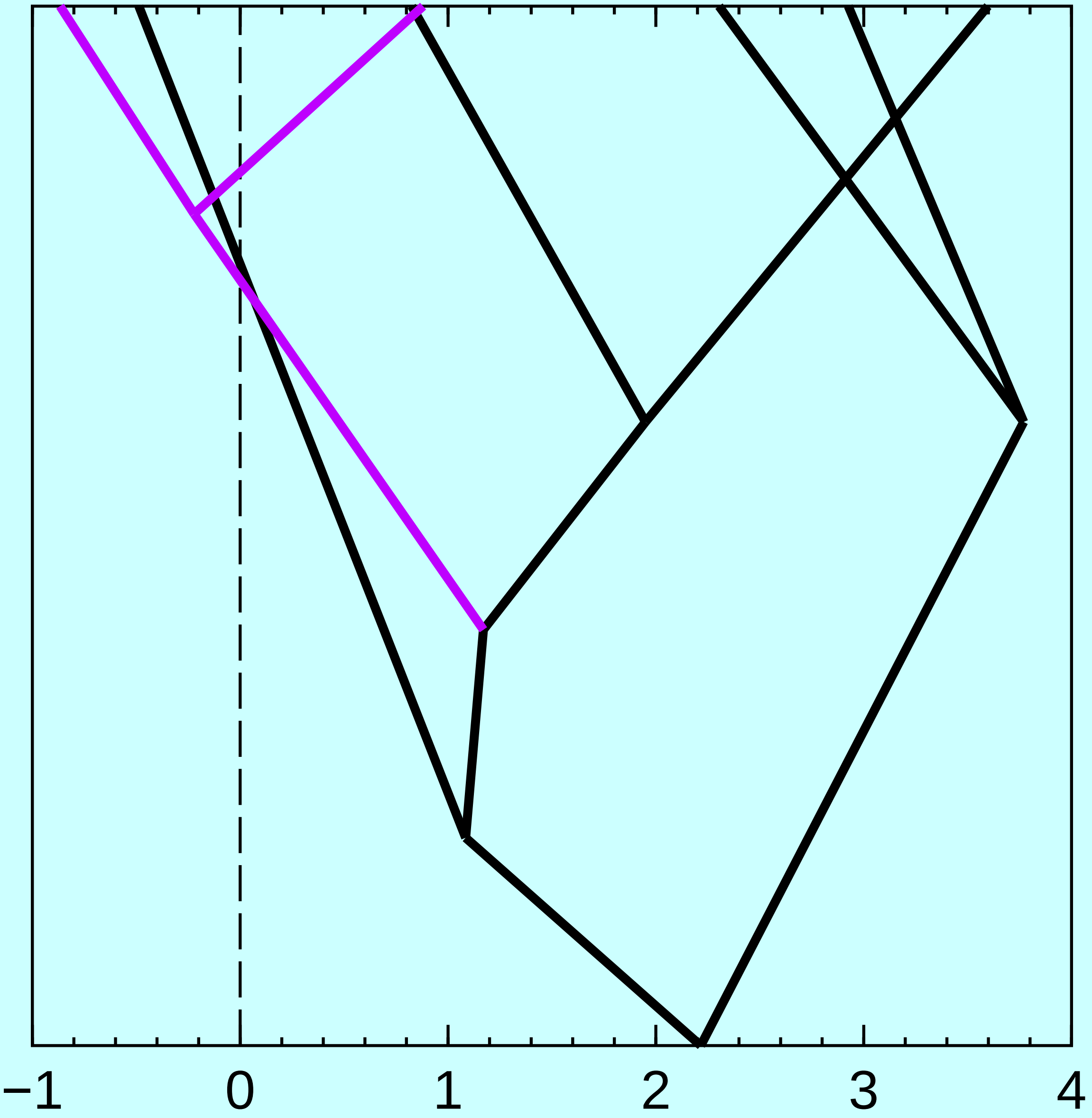
MCMC sampling of liabilities



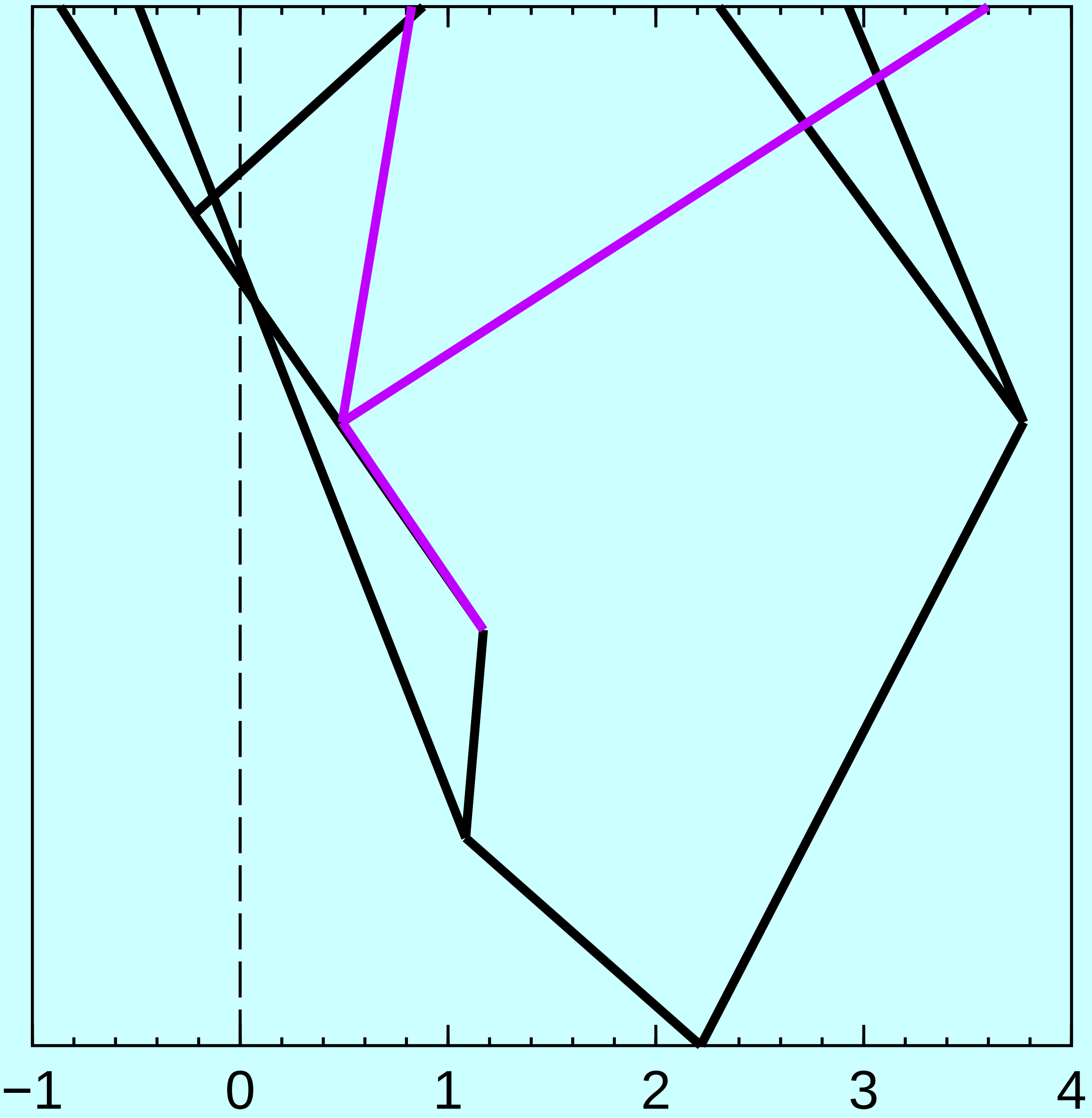
MCMC sampling of liabilities



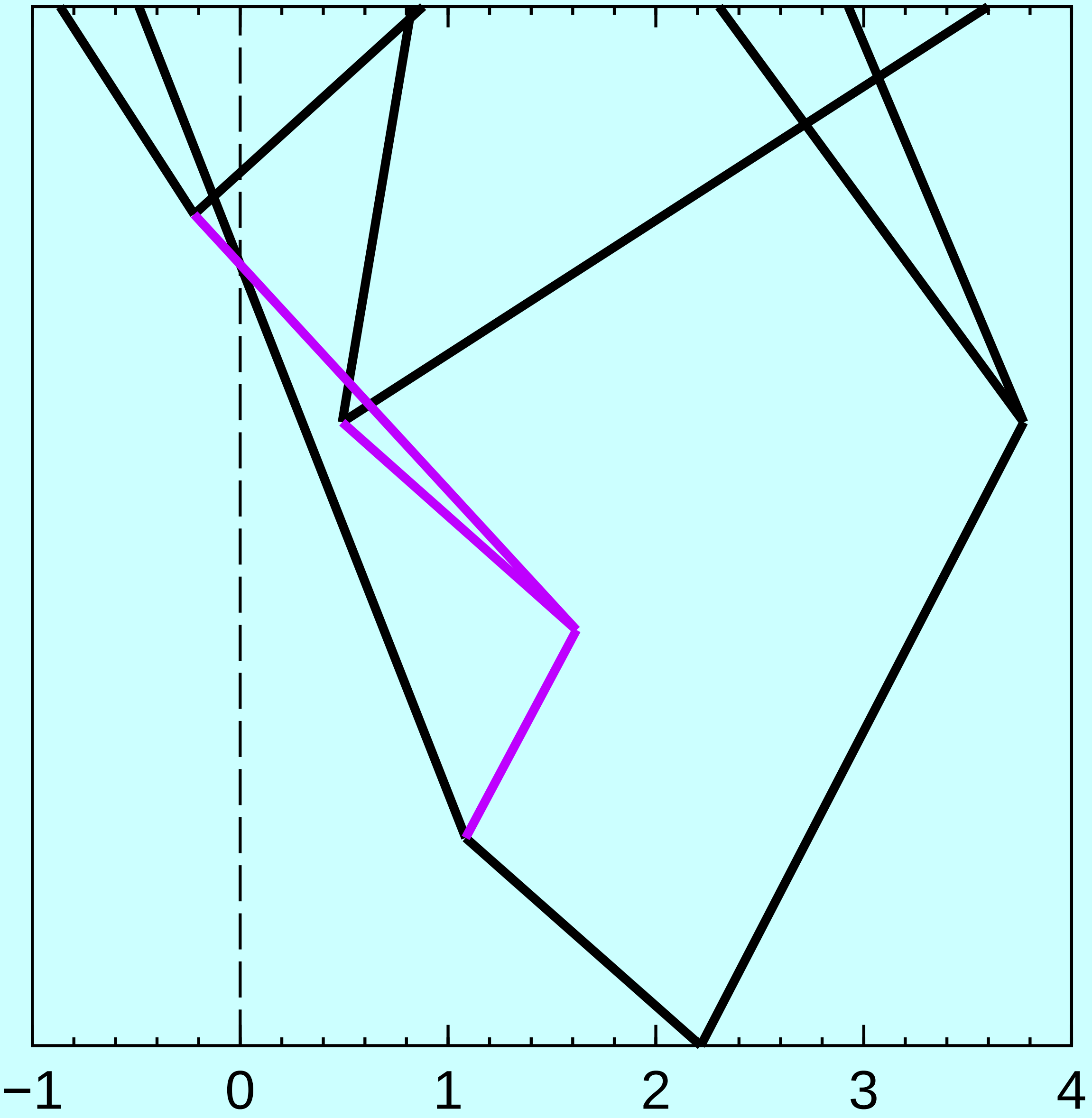
MCMC sampling of liabilities



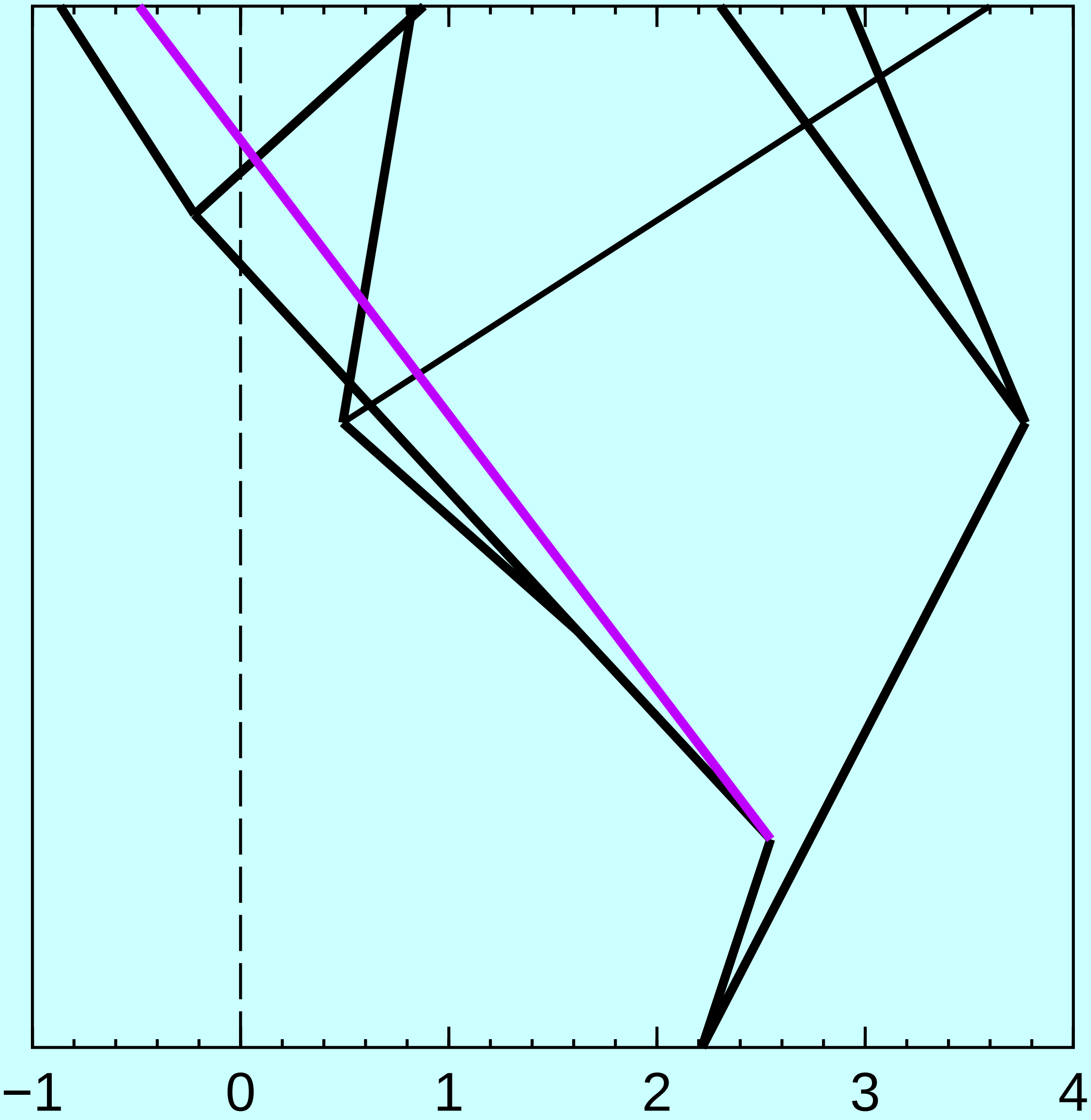
MCMC sampling of liabilities



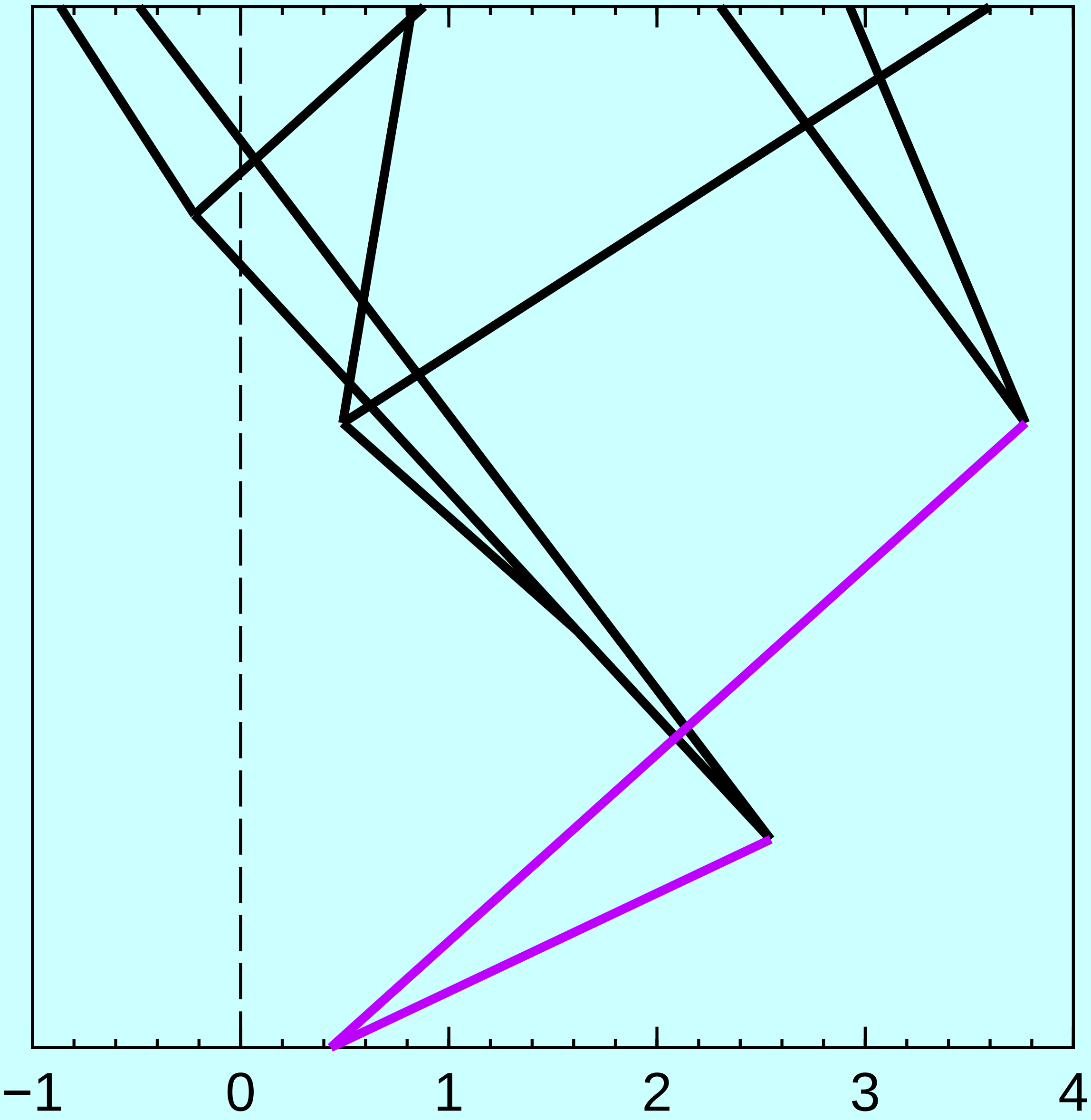
MCMC sampling of liabilities



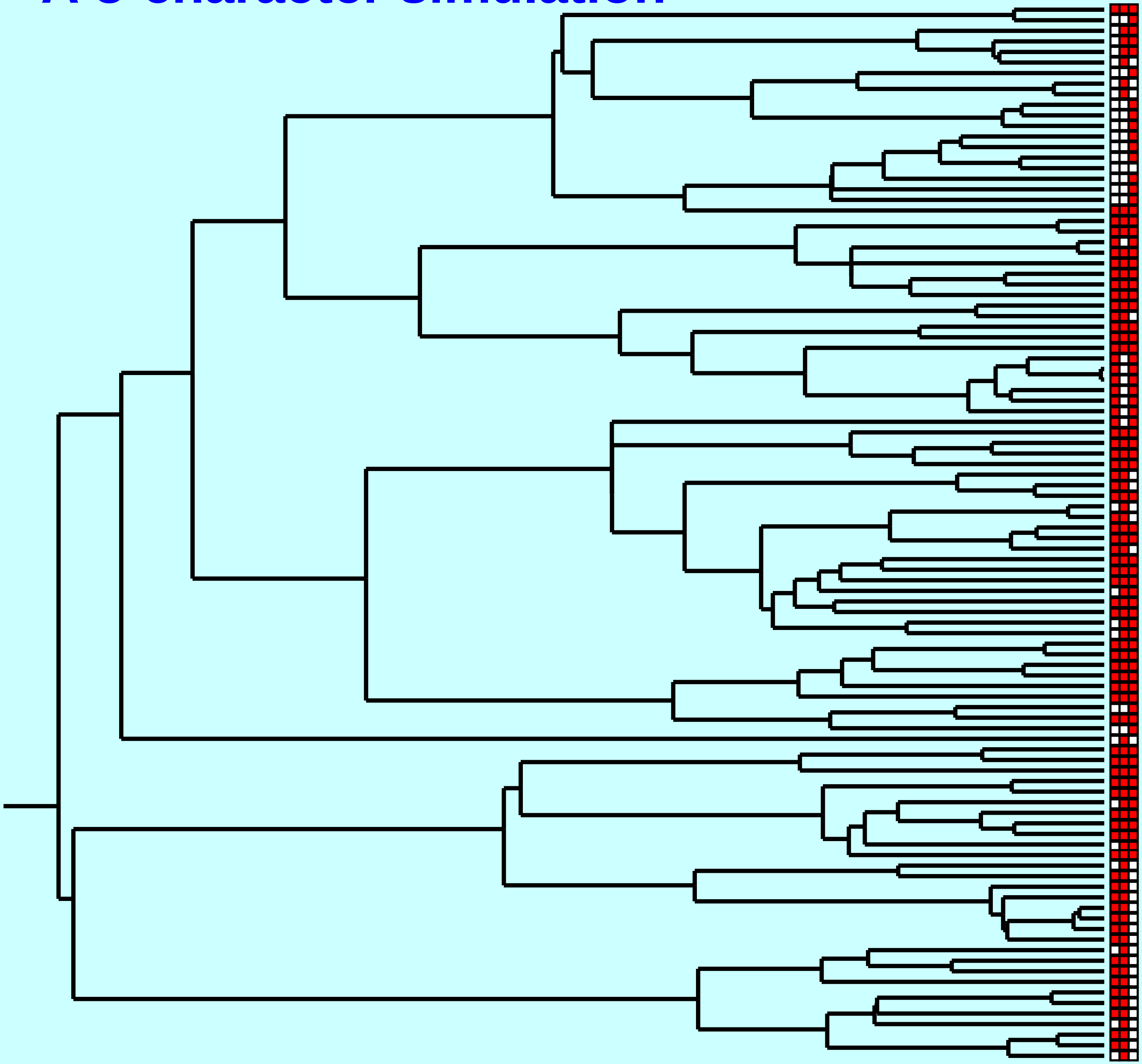
MCMC sampling of liabilities



MCMC sampling of liabilities



A 3-character simulation



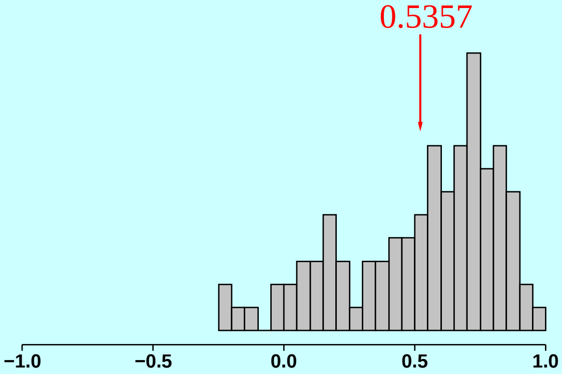
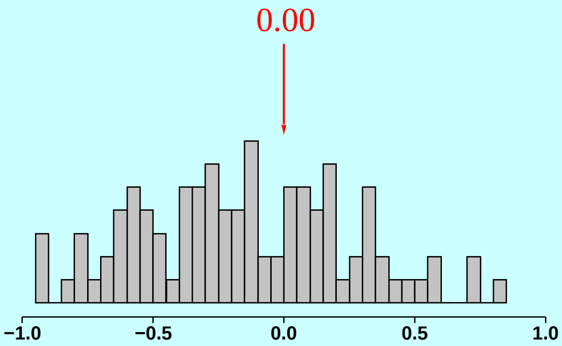
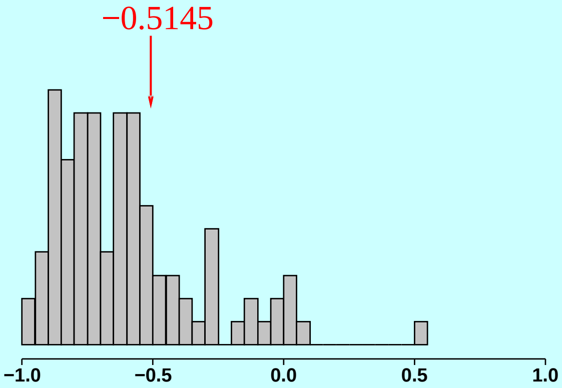
A 3-character simulation

For these true covariances:

$$\begin{bmatrix} 1.64 & 0.8 & 0 \\ 0.8 & 1.36 & -0.6 \\ 0 & -0.6 & 1 \end{bmatrix}$$

100 data sets with 100-species trees were analyzed.

Inferred correlation coefficients

	character 1	character 2
character 1		
character 2		
character 3		

What about QTLs?

(QTL = Quantitative Trait Locus)

- We can integrate these methods with QTL inference.

What about QTLs?

(QTL = Quantitative Trait Locus)

- We can integrate these methods with QTL inference.
- Not only identify QTLs, but to see them change across species, including some QTLs causing variation within some species, some within others.

What about QTLs?

(QTL = Quantitative Trait Locus)

- We can integrate these methods with QTL inference.
- Not only identify QTLs, but to see them change across species, including some QTLs causing variation within some species, some within others.
- Could even allow us to infer on which of two correlated characters the selection really acted.

The Reunion

- For the last 40-50 years population-genetic work within species has been (mostly) isolated from work on molecular evolution between species.
- Now we are in a gradual Reunion of these two lines of work (*not a New Synthesis, though*) as observations can be made that connect them (coalescents across species boundaries, Ds/Dn inferences, etc.)
- As this happens, Russ Lande's vision will become more and more of a reality – quantitative genetics will become directly relevant to multi-species evolutionary biology.

More generally we are seeing increased connections between

- Within- and between-species work
- Morphological and genomic studies
- Paleontological and neontological studies

What we can ... and cannot ... infer

- BUT ... we have limited power from any one sample of species. Biologists must learn to accept that, and find ways to propagate that uncertainty through the analysis that flow from these inferences. We cannot (ever!) have a Fly-On-The-Wall account of evolution.

What we can ... and cannot ... infer

- BUT ... we have limited power from any one sample of species. Biologists must learn to accept that, and find ways to propagate that uncertainty through the analysis that flow from these inferences. We cannot (ever!) have a Fly-On-The-Wall account of evolution.
- Furthermore we must always be sensitive to the limits of our models – as we expand the tree to less related groups, the models are called severely into question.

Thanks to ...

Felsenstein / Rudd
Family Funds

"Instead of painting the kitchen"

NSF

for several
grants to me
and Fred

NIH NIGMS

for several past
grants to me
and Mary Kuhner

Felsenstein / Rudd
family funds

(more and
more necessary)

... and of course to the Colegio Nacional and UNAM for inviting me here.

References

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30 (2):** 314-334. **[One of Russ's major papers using the constant-variances approximation]**

Stebbins, G. L. 1950. Variation and Evolution in Plants. Columbia University Press, New York. **[Describes selective covariance and cites Tedin (1926) for it]**

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60:** 466-481. **[Pioneering paper on using morphology and molecules to place fossils and date divergences]**

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, Early Access, 10.1093/sysbio/sys058 **[Another pioneering paper on using morphology and molecules to place fossils and date divergences]**

References

Revell, L. J., D. L. Mahler, R. G. Reynolds, and G. J. Slater. 2015. Placing cryptic, recently extinct, or hypothesized taxa into an ultrametric phylogeny using continuous character data: A case study with the lizard *Anolis roosevelti*. *Evolution* **69:** 1027-1035. **[Equivalent to the approach advocated here – computations different in detail but results should be the same]**

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]**

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]**

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]**

References

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* **19:** 445-471. **[Review with mention of usefulness of threshold model]**

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. 2004. *Inferring Phylogenies.* Sinauer Associates, Sunderland, Massachusetts. **Mentions threshold model**

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. **[The threshold 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 threshold method for both continuous and discrete characters, together]**