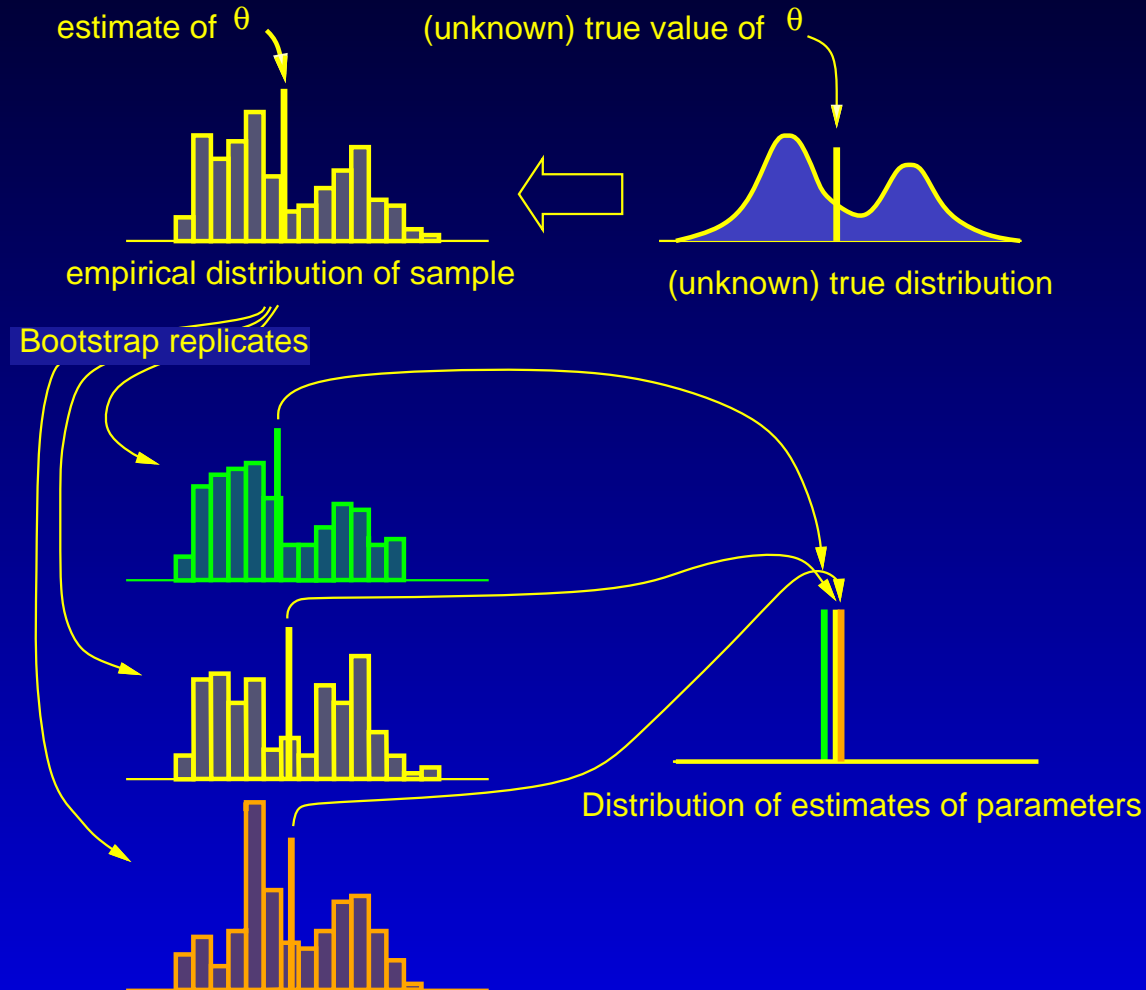


Lecture 27. Phylogeny methods, part 7 (Bootstraps, etc.)

Joe Felsenstein

Department of Genome Sciences and Department of Biology

A non-phylogeny example of the bootstrap



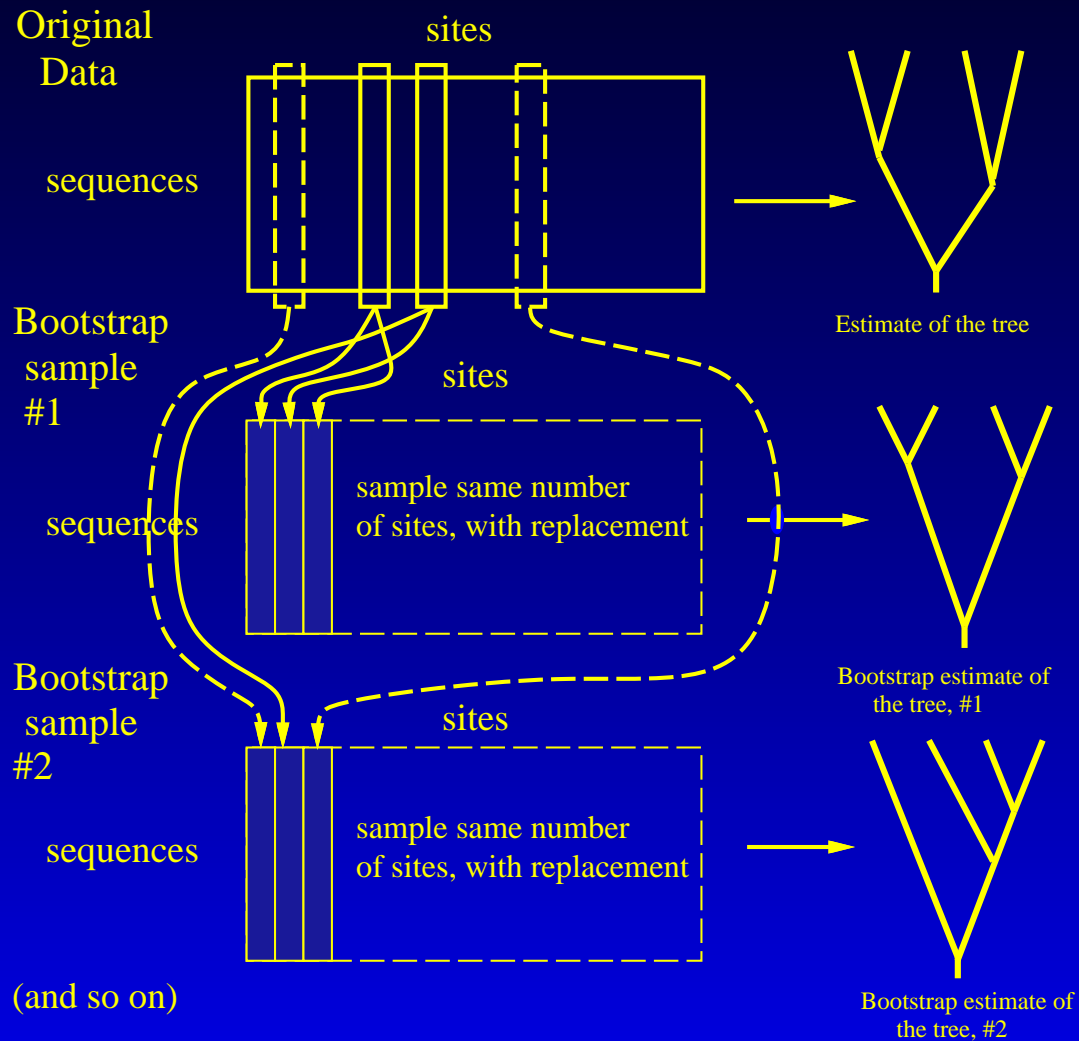
Bootstrap sampling from a distribution (a mixture of two normals) to estimate the variance of the mean

Bootstrap sampling

To infer the error in a quantity, θ , estimated from a sample of points x_1, x_2, \dots, x_n we can

- Do the following R times ($R = 1000$ or so)
- Draw a “bootstrap sample” by sampling n times with replacement from the sample. Call these $x_1^*, x_2^*, \dots, x_n^*$. Note that some of the original points are represented more than once in the bootstrap sample, some once, some not at all.
- Estimate θ from the bootstrap sample, call this $\hat{\theta}_k^*$ ($k = 1, 2, \dots, R$)
- When all R bootstrap samples have been done, the distribution of $\hat{\theta}_i^*$ estimates the distribution one would get if one were able to draw repeated samples of n points from the unknown true distribution.

Bootstrap sampling of phylogenies

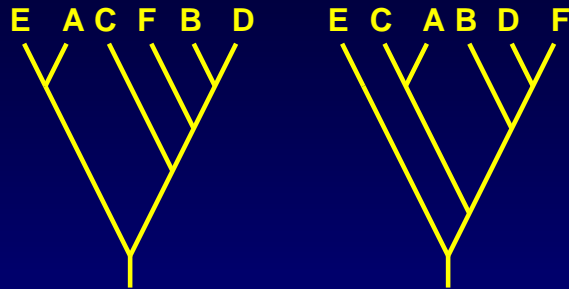


More on the bootstrap for phylogenies

- The sites are assumed to have evolved independently given the tree. They are the entities that are sampled (the x_j).
- The trees play the role of the parameter. One ends up with a cloud of R sampled trees.
- To summarize this cloud, we ask, for each branch in the tree, how frequently it appears among the cloud of trees.
- We make a tree that summarizes this for all the most frequently occurring branches.
- This is the *majority rule consensus tree* of the bootstrap estimates of the tree.

Majority rule consensus trees

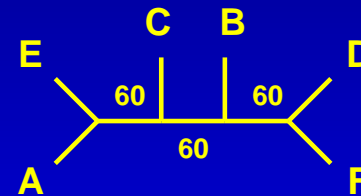
Trees:



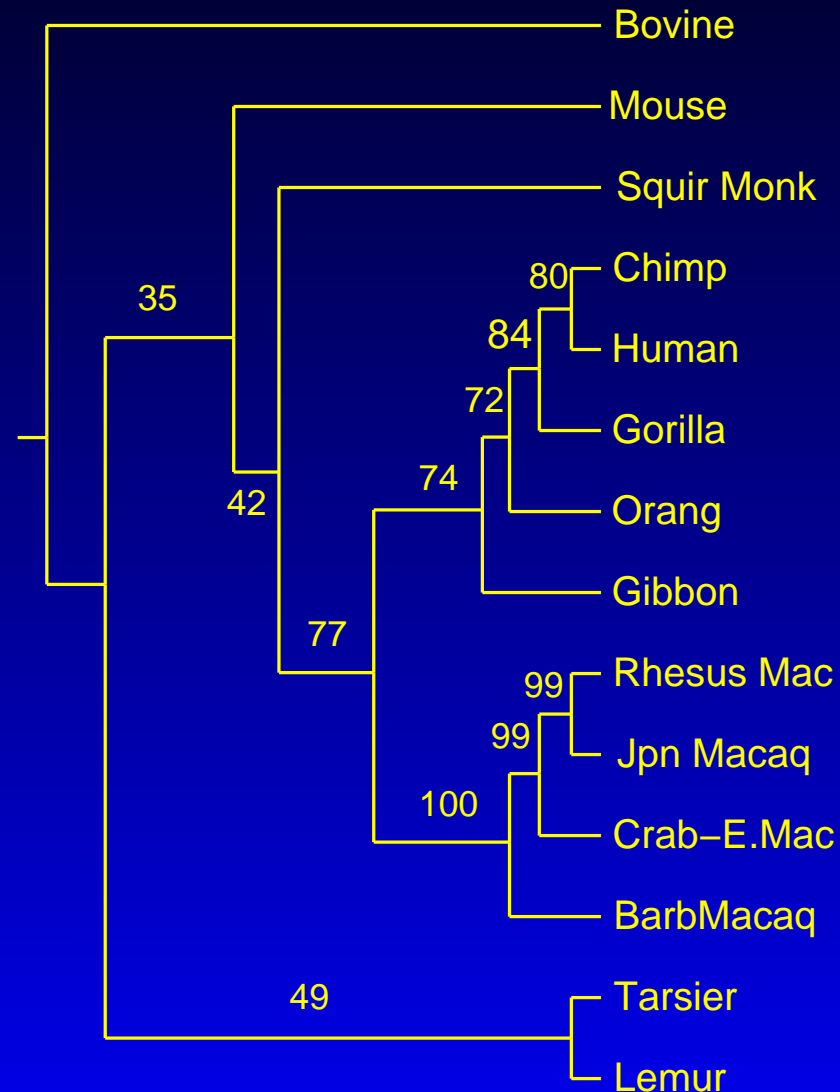
How many times each partition of species is found:

AE BCDF	3
ACE BDF	3
ACEF BD	1
AC BDEF	1
AEF BCD	1
ADEF BC	2
ABDF EC	1
ABCE DF	3

Majority-rule consensus tree of the unrooted trees:



An example of bootstrap sampling of trees



232 nucleotide, 14-species mitochondrial D-loop analyzed by parsimony, 100 bootstrap replicates

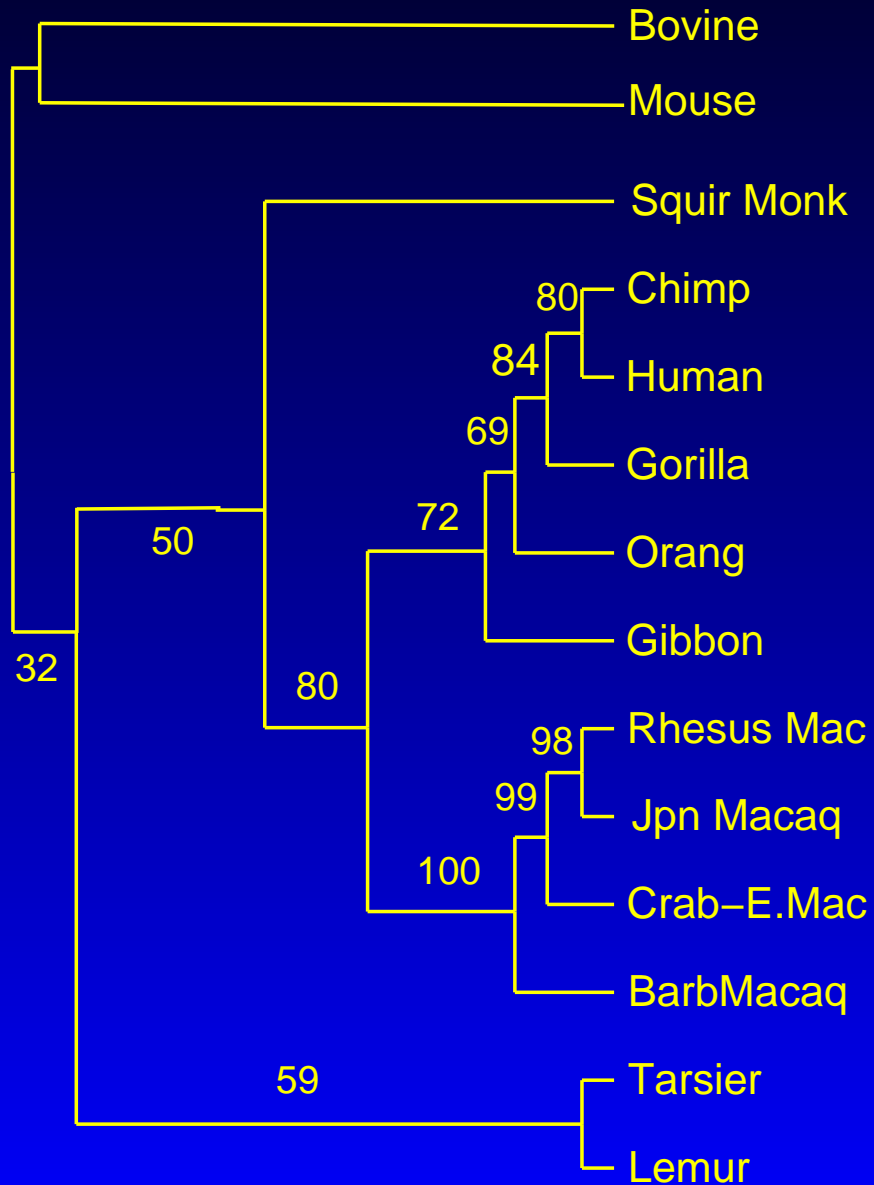
Potential problems with the bootstrap

1. Sites may not evolve independently
2. Sites may not come from a common distribution (but can consider them sampled from a mixture of possible distributions)
3. If do not know which branch is of interest at the outset, a “multiple-tests” problem means P values are overstated
4. P values are biased (too conservative)
5. Bootstrapping does not correct biases in phylogeny methods

Other resampling methods

- Delete-half jackknife. Sample a random 50% of the sites, *without* replacement.
- Delete-1/e jackknife (Farris et. al. 1996) (too little deletion from a statistical viewpoint).
- Reweighting characters by choosing weights from an exponential distribution.
- In fact, reweighting them by any exchangeable weights having coefficient of variation of 1
- Parametric bootstrap – simulate data sets of this size assuming the estimate of the tree is the truth
- (to correct for correlation among adjacent sites) (Künsch, 1989)
Block-bootstrapping – sample n/b blocks of b adjacent sites.

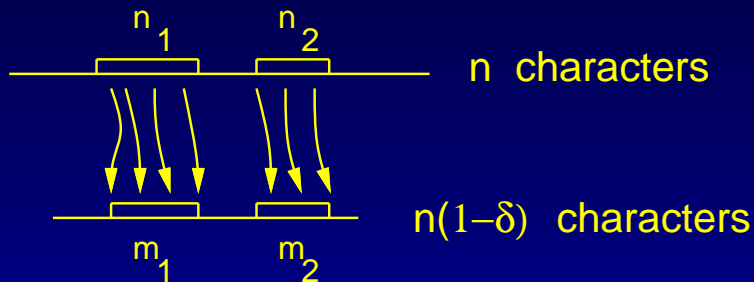
Delete half jackknife on the example



Calibrating the jackknife

Exact computation of the effects of deletion fraction for the jackknife

(suppose 1 and 2 are conflicting groups)



Bootstrap	Jackknife	
	$\delta = 1/2$	$\delta = 1/e$
$\text{Prob}(m_1 > m_2)$	0.6384	0.6441
$\text{Prob}(m_1 \geq m_2)$	0.7230	0.8040
$\text{Prob}(m_1 > m_2) + \frac{1}{2} \text{Prob}(m_1 = m_2)$	0.6807	0.7240

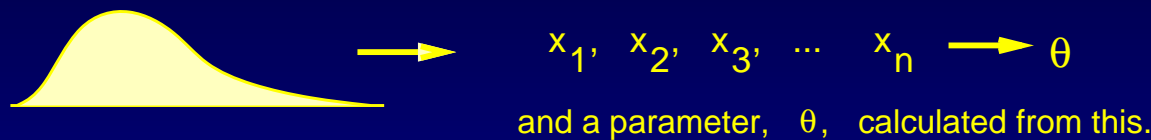
We can compute for various n 's the probabilities of getting more evidence for group 1 than for group 2

A typical result is for $n_1 = 10$, $n_2 = 8$, $n = 100$:

Parametric bootstrapping

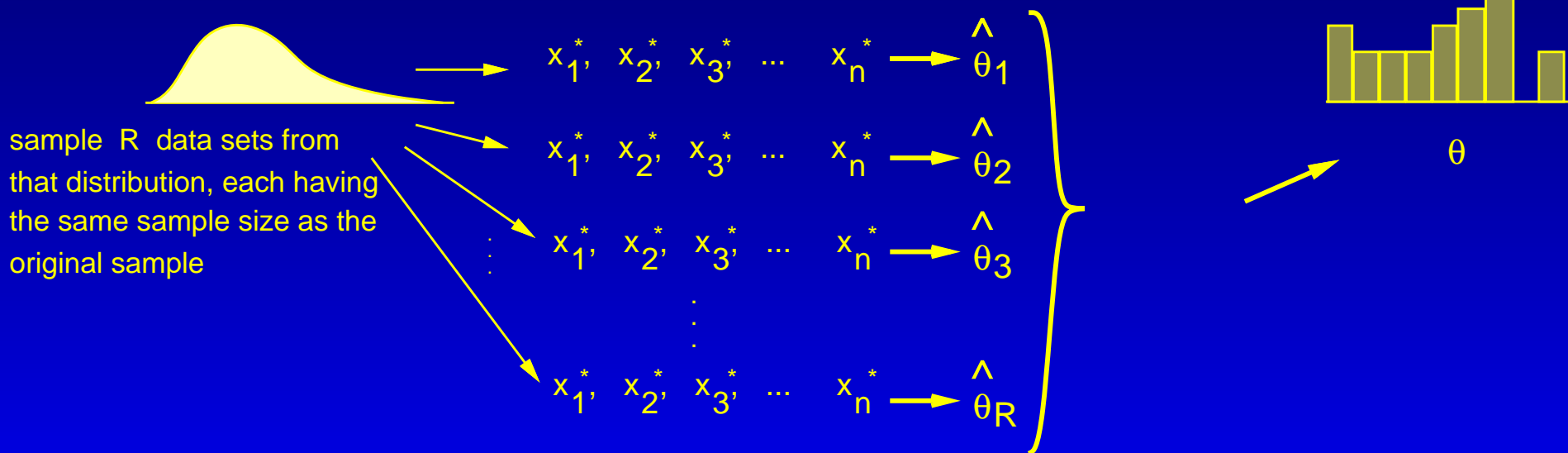
The Parametric Bootstrap (Efron, 1985)

Suppose we have independent observations drawn from a known distribution:

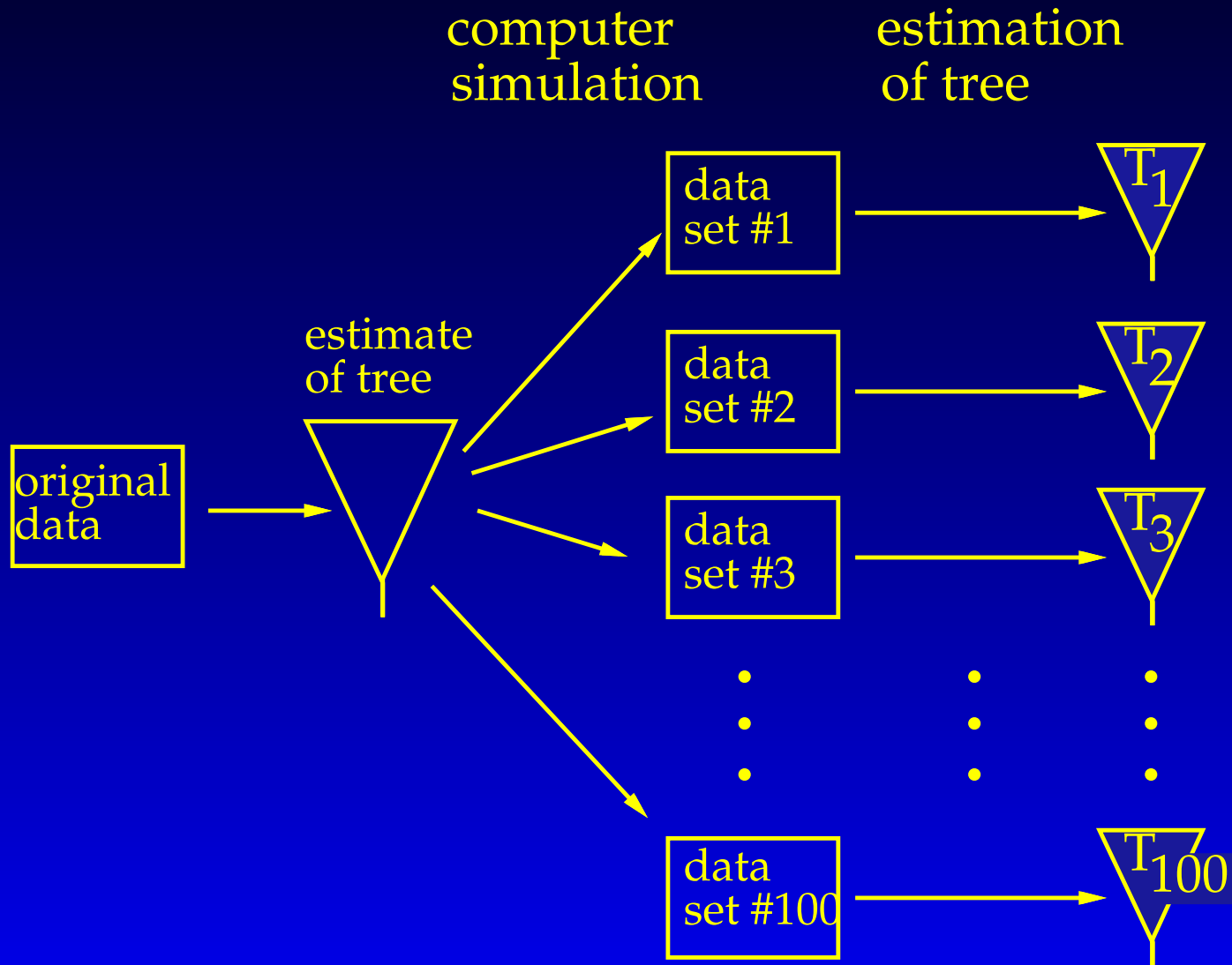


and take the distribution of the $\hat{\theta}_i$ as the estimate of the distribution from which it is drawn

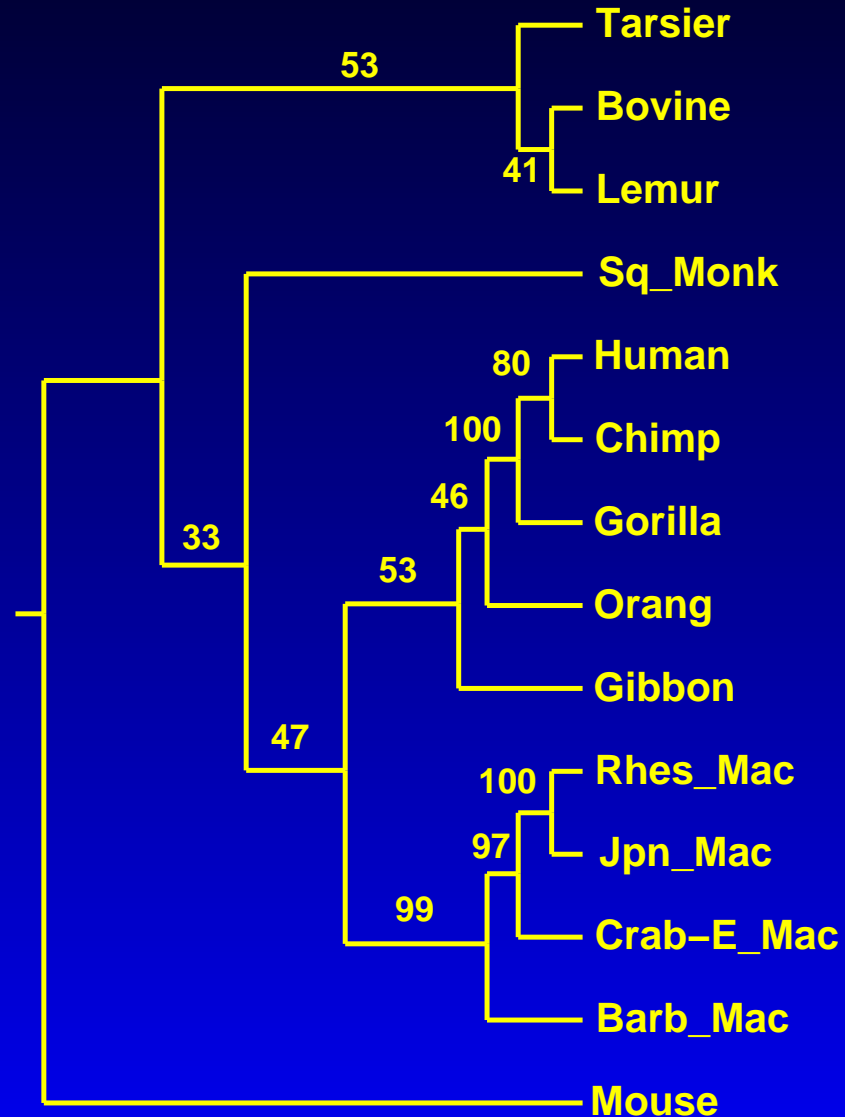
- To infer the variability of θ
- Use the current estimate, $\hat{\theta}$
- Use the distribution that has that as its true parameter



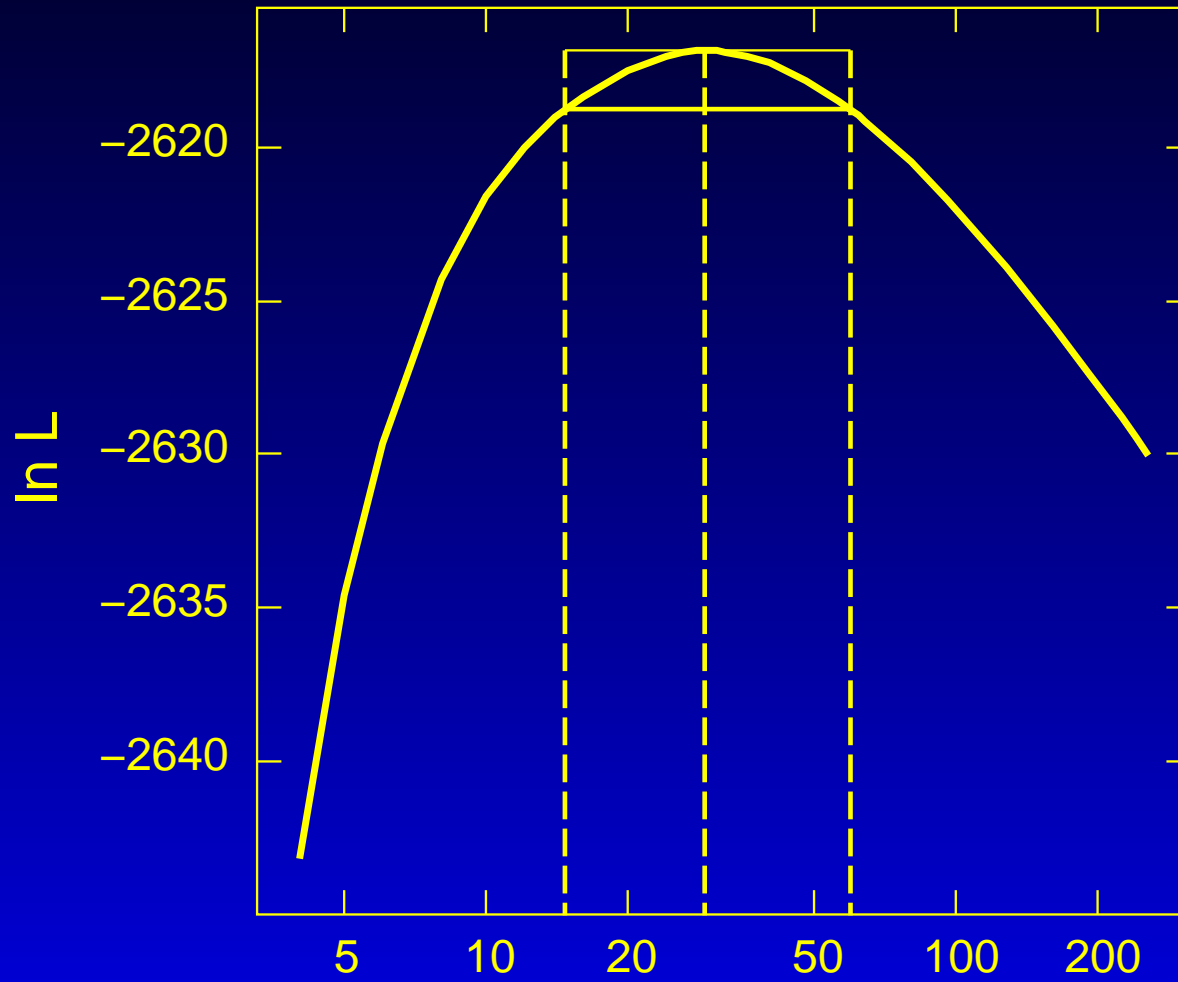
The parametric bootstrap for phylogenies



An example of the parametric bootstrap

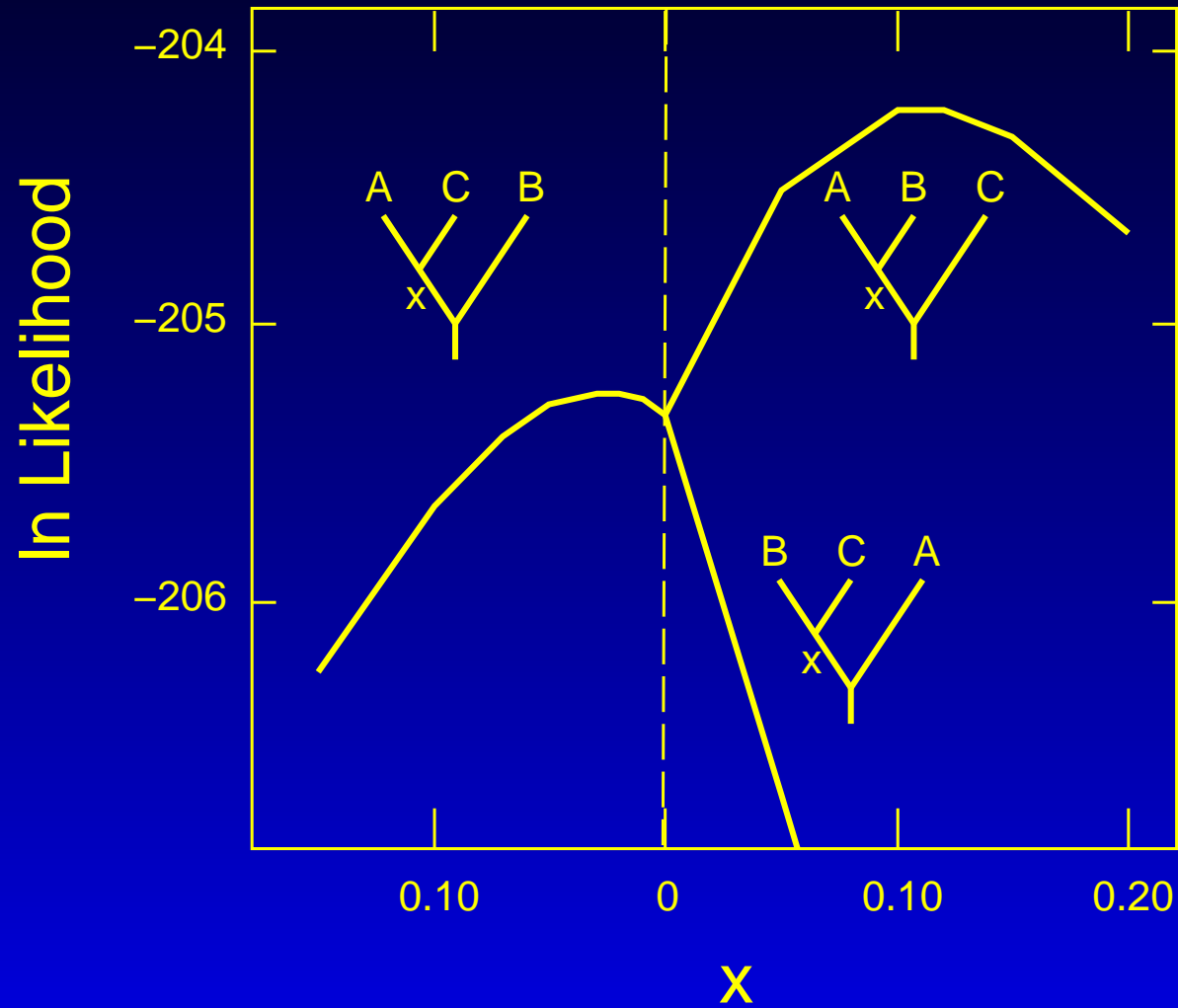


Likelihood ratio confidence limits on Ts/Tn ratio

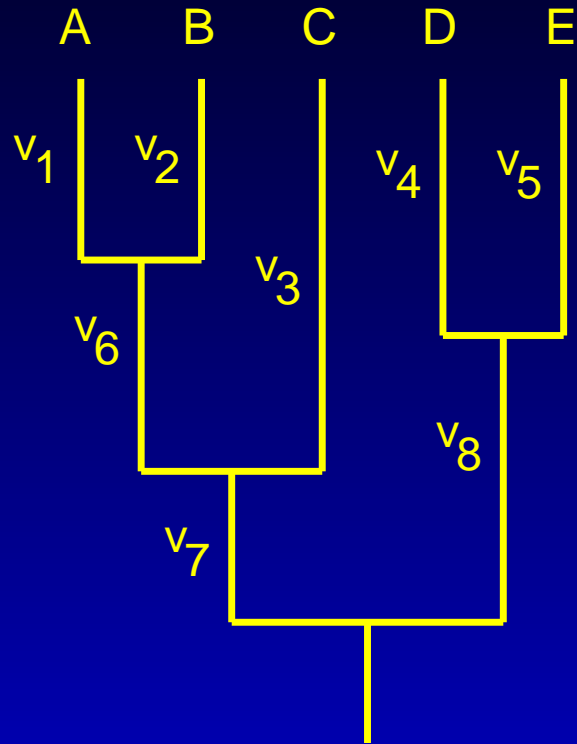


Transition / transversion ratio
for the 14-species primate data set

Likelihoods in tree space – a 3-species clock example



The constraints for a molecular clock



Constraints for a clock

$$v_1 = v_2$$

$$v_4 = v_5$$

$$v_1 + v_6 = v_3$$

$$v_3 + v_7 = v_4 + v_8$$

Testing for a molecular clock

To test for a molecular clock:

- Obtain the likelihood with no constraint of a molecular clock (For primates data with $T_s/T_n = 30$ we get $\ln L_1 = -2616.86$)
- Obtain the highest likelihood for a tree which is constrained to have a molecular clock: $\ln L_0 = -2679.0$
- Look up $2(\ln L_1 - \ln L_0) = 2 \times 62.14 = 124.28$ on a χ^2 distribution with $n - 2 = 12$ degrees of freedom (in this case the result is significant)

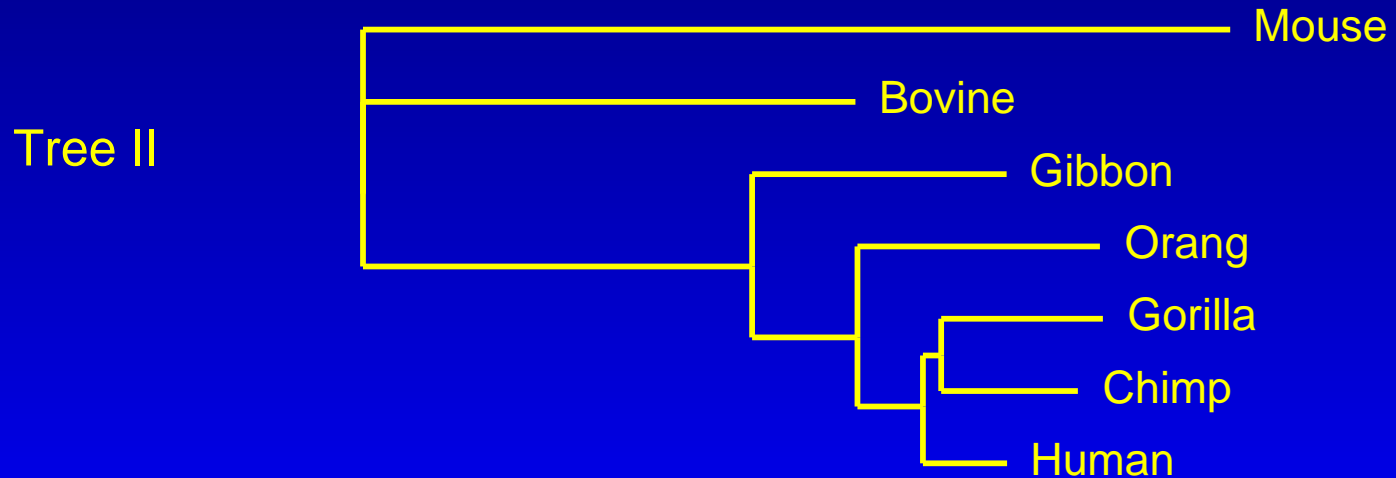
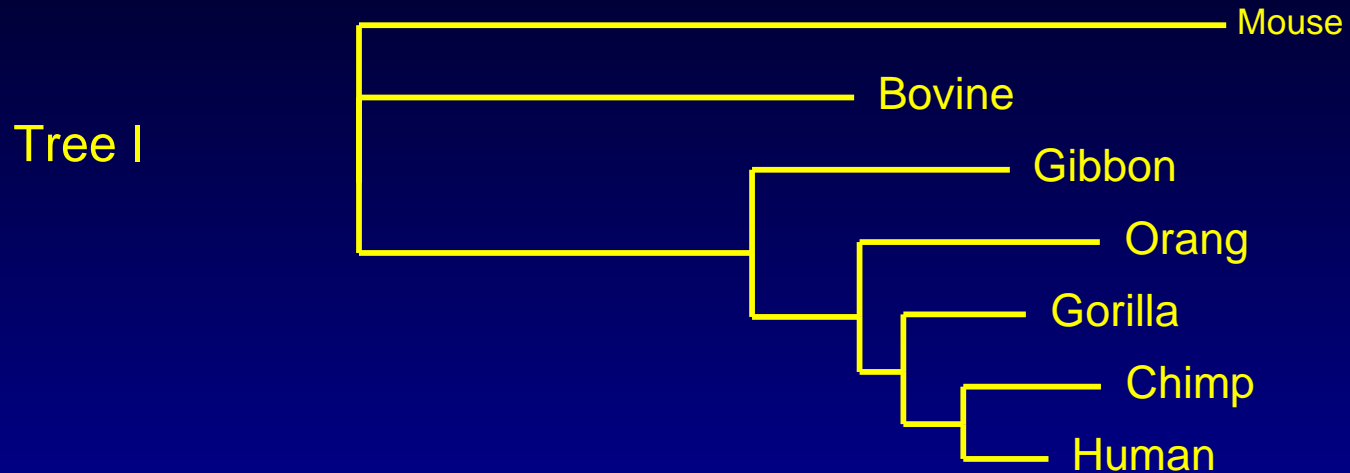
Goldman's simulation test of Likelihood Ratios

Goldman (1993) suggests that, in cases where we may wonder whether the Likelihood Ratio Test statistic really has its desired χ^2 distribution we can:

- Take our best estimate of the tree
- Simulate on it the evolution of data sets of the same size
- For each replicate, calculate the LRT statistic
- Use this as the distribution and see where the actual LRT value lies in it (e.g.: in the upper 5%?)

This, of course, is a parametric bootstrap.

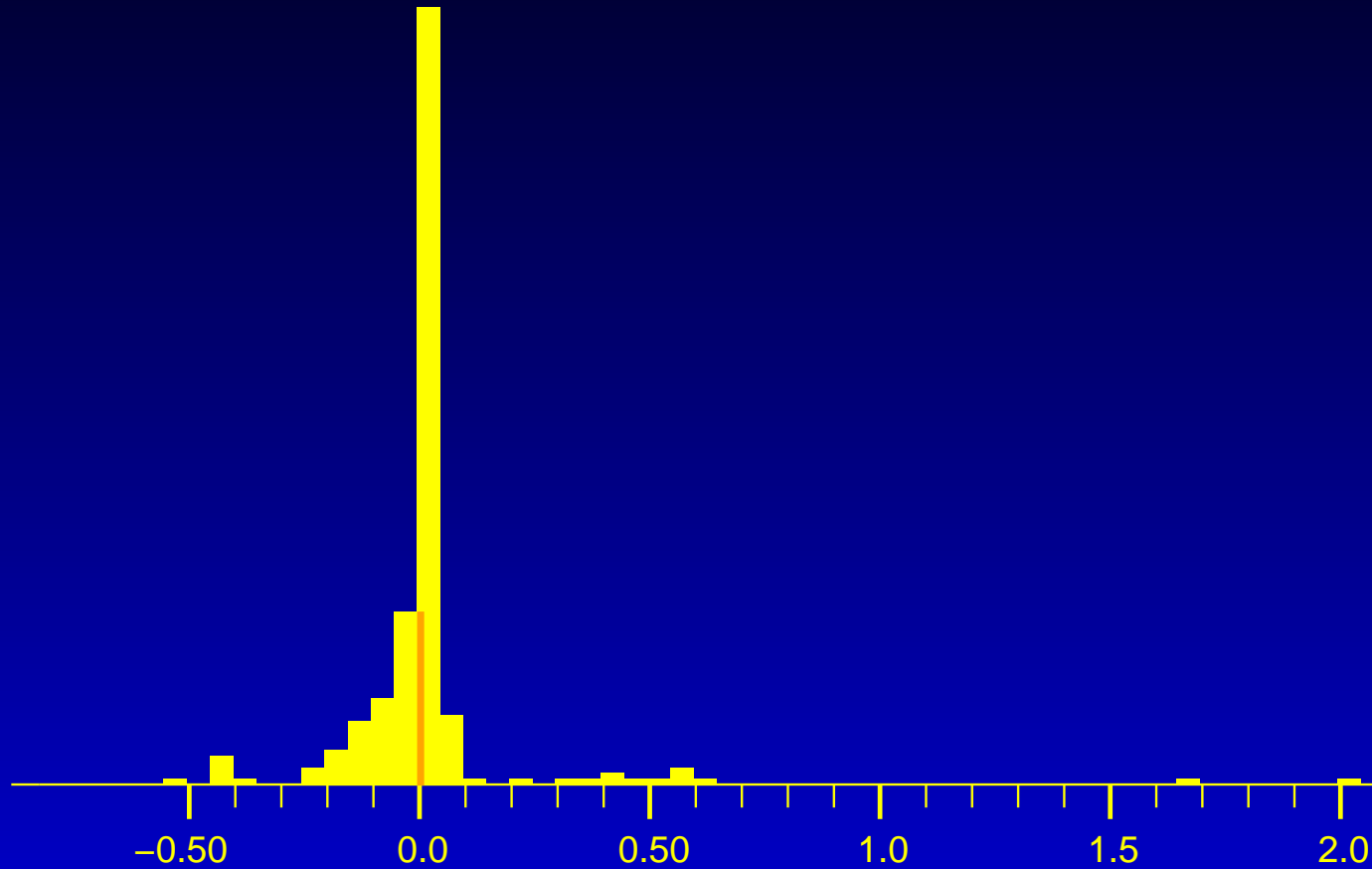
Two trees to be tested by paired sites tests



Differences in log likelihoods site by site

Tree \ site	1	2	3	4	5	6	...	231	232	In L
I	-2.971	-4.483	-5.673	-5.883	-2.691	-8.003	...	-2.971	-2.691	-1405.61
II	-2.983	-4.494	-5.685	-5.898	-2.700	-7.572	...	-2.987	-2.705	-1408.80
Diff	+0.012	+0.111	+0.013	+0.015	+0.010	-0.431	...	+0.012	+0.010	+3.19

Histogram of log likelihood differences



Difference in log likelihood at site

Paired sites tests

- Winning sites test (Prager and Wilson, 1988). Do a sign test on the signs of the differences.
- z test (me, 1993 in PHYLIP documentation). Assume differences are normal, do z test of whether mean (hence sum) difference is significant.
- t test. Swofford et. al., 1996: do a t test (paired)
- Wilcoxon ranked sums test (Templeton, 1983).
- RELL test (Kishino and Hasegawa, 1989 per my suggestion). Bootstrap resample sites, get distribution of difference of totals.

In this example

- Winning sites test. 160 of 232 sites favor tree I. $P < 3.279 \times 10^{-9}$
- z test. Difference of log-likelihood totals is 0.948104 standard deviations from 0, $P = 0.343077$. Not significant.
- t test. Same as z test for this large a number of sites.
- Wilcoxon ranked sums test. Rank sum is 4.82805 standard deviations below its expected value, $P = 0.000001378765$
- RELL test. 8,326 out of 10,000 samples have a positive sum, $P = 0.3348$ (two-sided)

Bayesian methods

In the Bayesian framework, one can avoid the separate calculation of confidence intervals. The posterior distribution of trees shows us how much credence to give different trees (for example, it assigns probabilities to different tree topologies).

The unresolved issue is how to summarize this posterior distribution in the best way. In this respect Bayesian methods leave you in a situation analogous to having the cloud of bootstrap-sampled trees without yet having summarized them.

References

- Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**: 795-803. **[Bremer support]**
- Cavender, J. A. 1977. Taxonomy with confidence. *Mathematical Biosciences* **40**: 271-280 (Erratum, vol. 44, p. 308, 1979) **[First paper on testing trees]**
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**: 1-26. **[The original bootstrap paper]**
- Efron, B. 1985. Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72**: 45-58. **[The parametric bootstrap]**
- Farris, J. S., V. A. Albert, M. Källersjö, D. Lipscomb, and A. G. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**: 99-124. **[The delete-1/e jackknife for phylogenies]**
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791. **[The bootstrap first applied to phylogenies]**

more references

- Felsenstein, J. and H. Kishino. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Systematic Biology* **42**: 193-200. [A more detailed exposition of the bias of P values in a normal case]
- Felsenstein, J. 1985c. Confidence limits on phylogenies with a molecular clock. *Systematic Zoology* **34**: 152-161. [A 3-species case where we can evaluate methods]
- Goldman, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* **36**: 182-98. [Parametric bootstrapping for testing models]
- Harshman, J. 1994. The effect of irrelevant characters on bootstrap values. *Systematic Zoology* **43**: 419-424. [Not much effect on parsimony whether or not you include invariant characters when bootstrapping]
- Hasegawa, M., H. Kishino. 1989. Confidence limits on the maximum-likelihood estimate of the hominoid tree from mitochondrial-DNA sequences. *Evolution* **43**: 672-677 [The KHT test]
- Hasegawa, M. and H. Kishino. 1994. Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. *Molecular Biology and Evolution* **11**: 142-145. [RELL probabilities]

more references

- Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**: 182-192. **Bias in P values seen in a large simulation study]**
- Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* **29**: 170-179. **[The KHT test]**
- Künsch, H. R. 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**: 1217-1241. **[The block-bootstrap]**
- Margush, T. and F. R. McMorris. 1981. Consensus n-trees. *Bulletin of Mathematical Biology* **43**: 239-244i. **[Majority-rule consensus trees]**
- Prager, E. M. and A. C. Wilson. 1988. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. *Journal of Molecular Evolution* **27**: 326-335. **[winning-sites test]**

more references

Sanderson, M. J. 1995. Objections to bootstrapping phylogenies: a critique. *Systematic Biology* **44**: 299-320. [**Good but he accepts a few criticisms I would not have accepted**]

Shimodaira, H. and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114-1116. [**SH test for multiple trees**]

Sitnikova, T., A. Rzhetsky, and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Molecular Biology and Evolution* **12**: 319-333. [**The interior-branch test**]

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**: 221-244. [**First paper on KHT test**]

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling plans in regression analysis. *Annals of Statistics* **14**: 1261-1295. [**The delete-half jackknife**]

Zharkikh, A., and W.-H. Li. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Molecular Biology and Evolution* **9**: 1119-1147. [**Discovery and explanation of bias in P values**]

How it was done

This projection produced as a PDF, not a PowerPoint file, and viewed using the Full Screen mode (in the View menu of Adobe Acrobat Reader):

- using the `prospect` style in LaTeX,
- using LaTeX to make a `.dvi` file,
- using `dvi2ps` to turn this into a Postscript file,
- using `ps2pdf` to mill it into a PDF file, and
- displaying the slides in Adobe Acrobat Reader.

Result: nice slides using freeware.