

Some of these slides have been borrowed from Dr. Paul Lewis, Dr. Joe Felsenstein. Thanks!

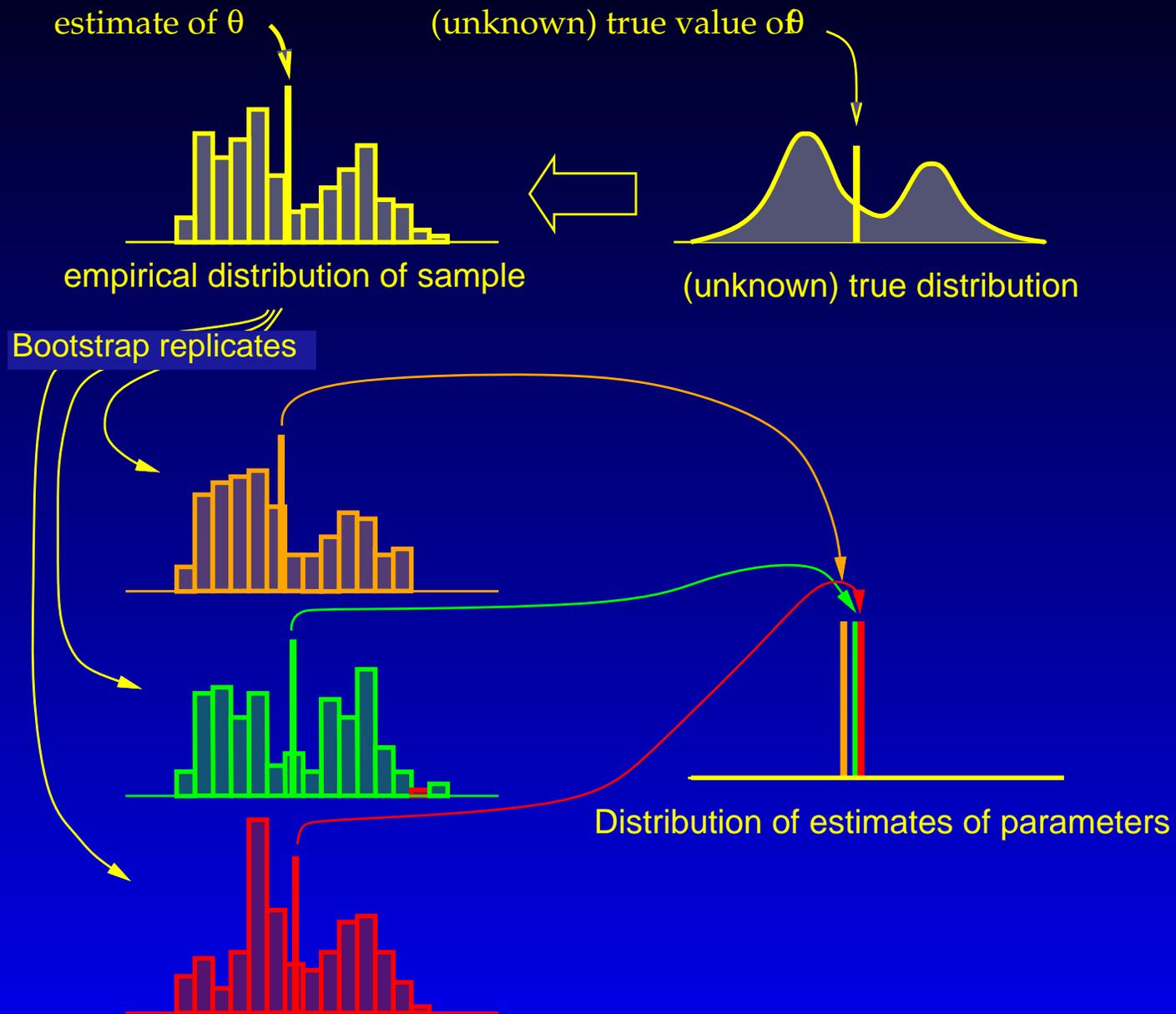
Paul has many great tools for teaching phylogenetics at his web site:

<http://hydrodictyon.eeb.uconn.edu/people/plewis>

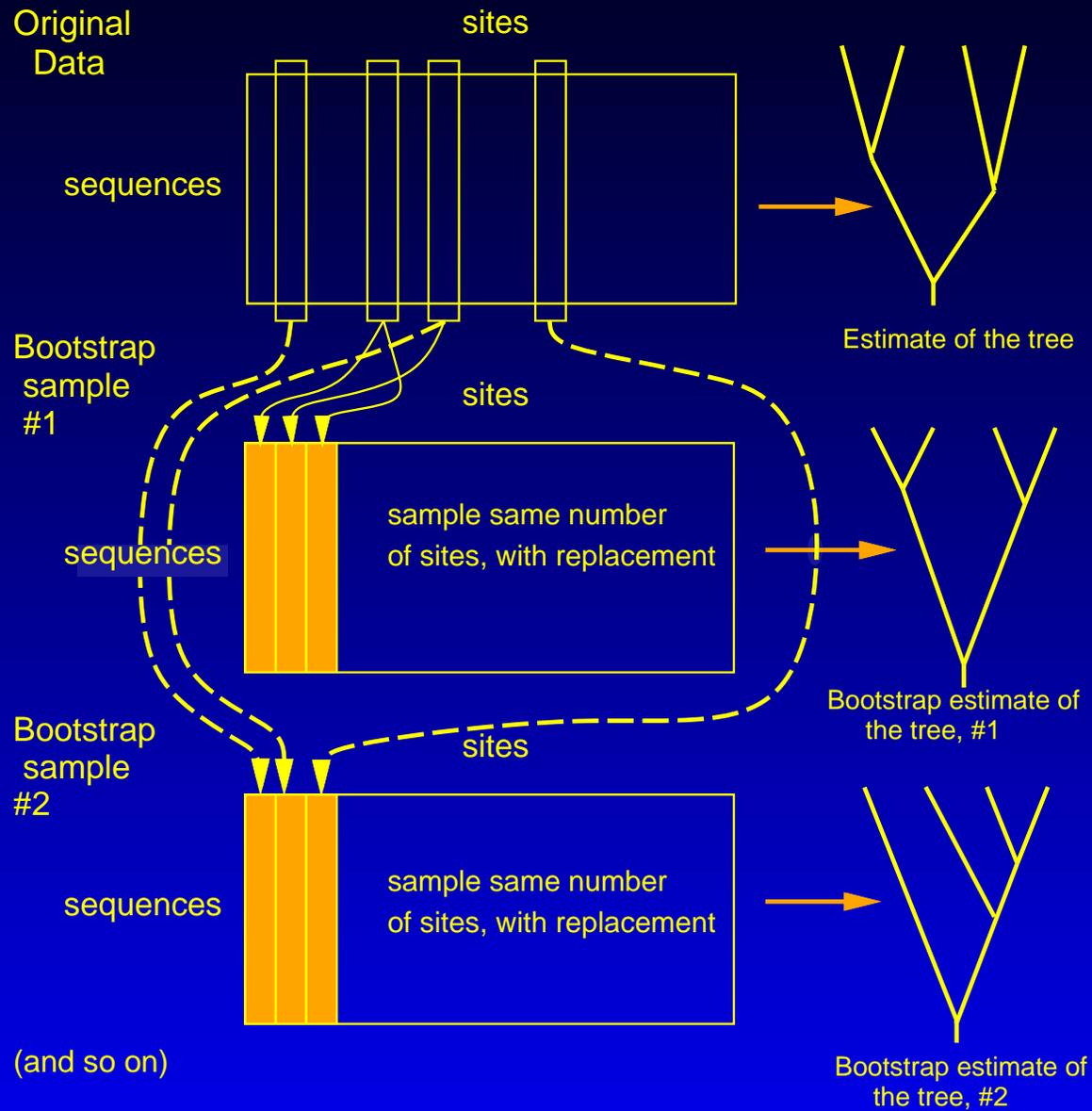
Overview:

- Bootstrapping to estimate clade support
- controversies about the interpretation of the bootstrap support
- the AU test

The bootstrap



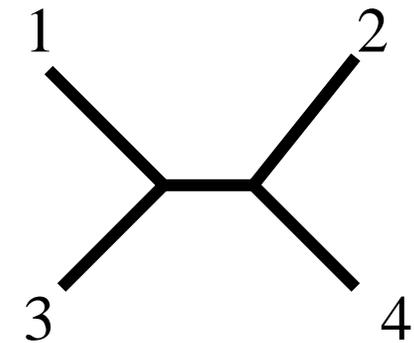
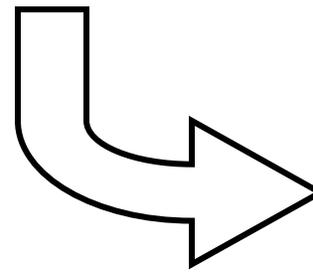
The bootstrap for phylogenies



Bootstrapping: first step

	1	2	3	4	5	6	7	...	k
1	T	A	G	T	C	G	T	...	A
2	T	C	A	T	C	G	T	...	G
3	A	T	G	T	C	A	C	...	G
4	A	T	A	T	C	G	C	...	G

From the original data, estimate a tree using, say, parsimony (could use NJ, LS, ML, etc., however)

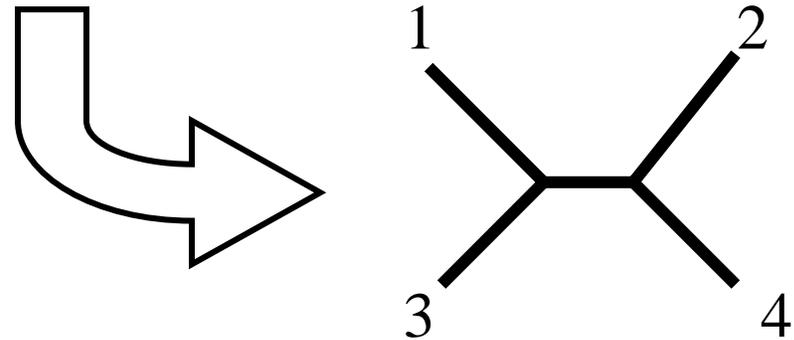


Bootstrapping: first replicate

	1	2	3	4	5	6	7	...	k
weights	1	2	0	0	1	3	1	...	2
1	T	A	G	T	C	G	T	...	A
2	T	C	A	T	C	G	T	...	G
3	A	T	G	T	C	A	C	...	G
4	A	T	A	T	C	G	C	...	G

Sum of weights equals k (i.e., each bootstrap dataset has same number of sites as the original)

From the bootstrap dataset, estimate the tree using the same method you used for the original dataset

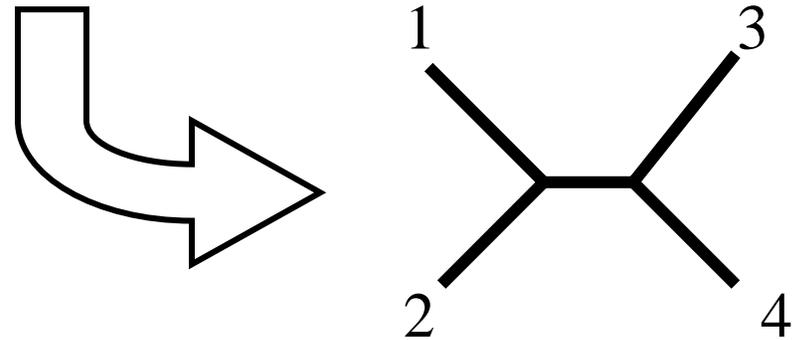


Bootstrapping: second replicate

	1	2	3	4	5	6	7	...	k
weights	0	1	1	1	1	3	0	...	0
1	T	A	G	T	C	G	T	...	A
2	T	C	A	T	C	G	T	...	G
3	A	T	G	T	C	A	C	...	G
4	A	T	A	T	C	G	C	...	G

Note that weights are different this time, reflecting the random sampling with replacement used to generate the weights

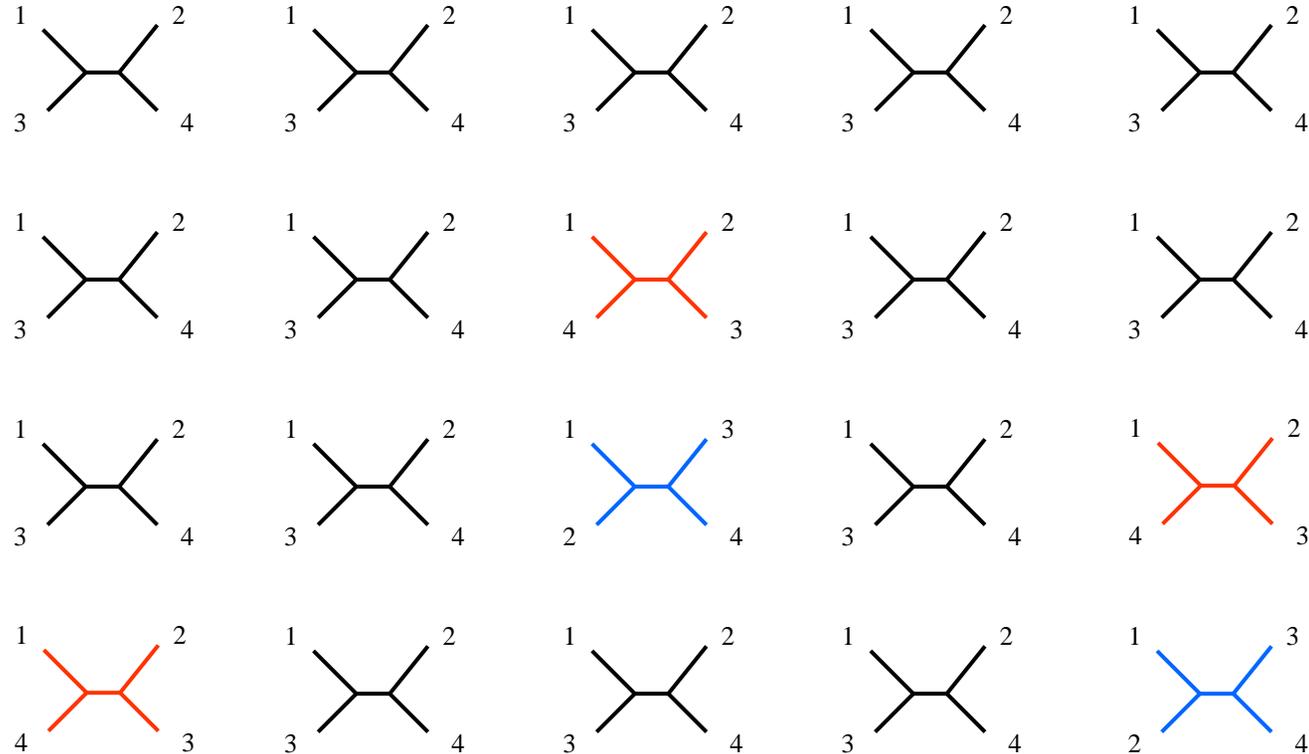
This time the tree that is estimated is different than the one estimated using the original dataset.



Bootstrapping: 20 replicates

1234	Freq

-*-*	75.0
-***	15.0
--**	10.0



Note: usually at least 100 replicates are performed, and 500 is better

Bootstrap support for branches can be displayed:

- on the best estimate of the tree (on the MP, OLS, or ML tree)
- on a majority-rule consensus tree (see [Berry and Gascuel, 1996](#))

Bootstrap proportions have been characterized as providing:

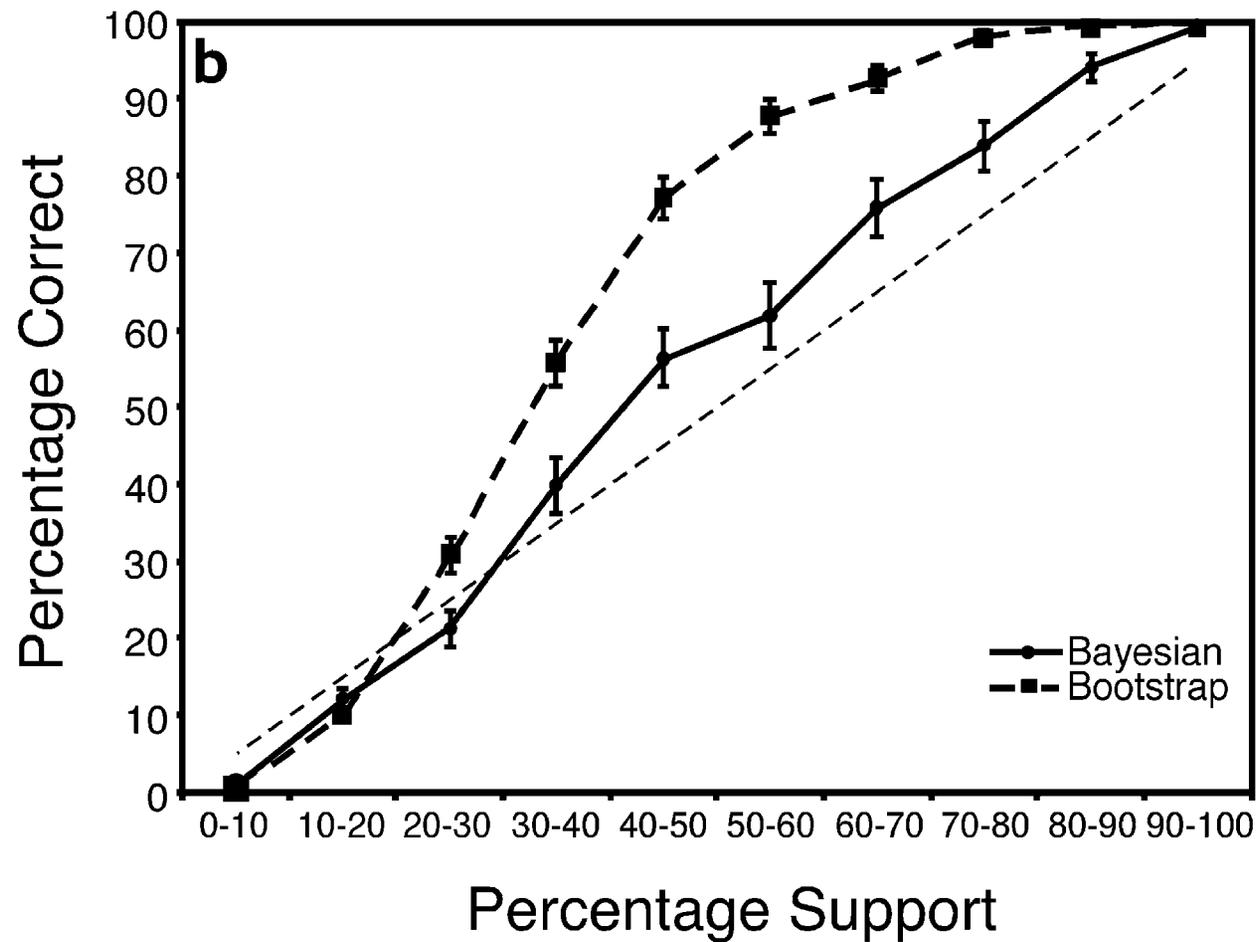
- a measure of repeatability,
- an estimate of the probability that the tree is correct (and bootstrapping has been criticized as being too conservative in this context),
- the P-value for a tree or clade

Bootstrap Proportion \neq Posterior Probability

Several studies have compared the non-parametric bootstrap proportion of clade from an ML analysis of a data set to the posterior probabilities when the same data is analyzed under the same model (Suzuki et al., 2002; Wilcox et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003).

Not all of these have implied that the measures **should** be the same, but some authors have, usually citing Efron et al. (1996).

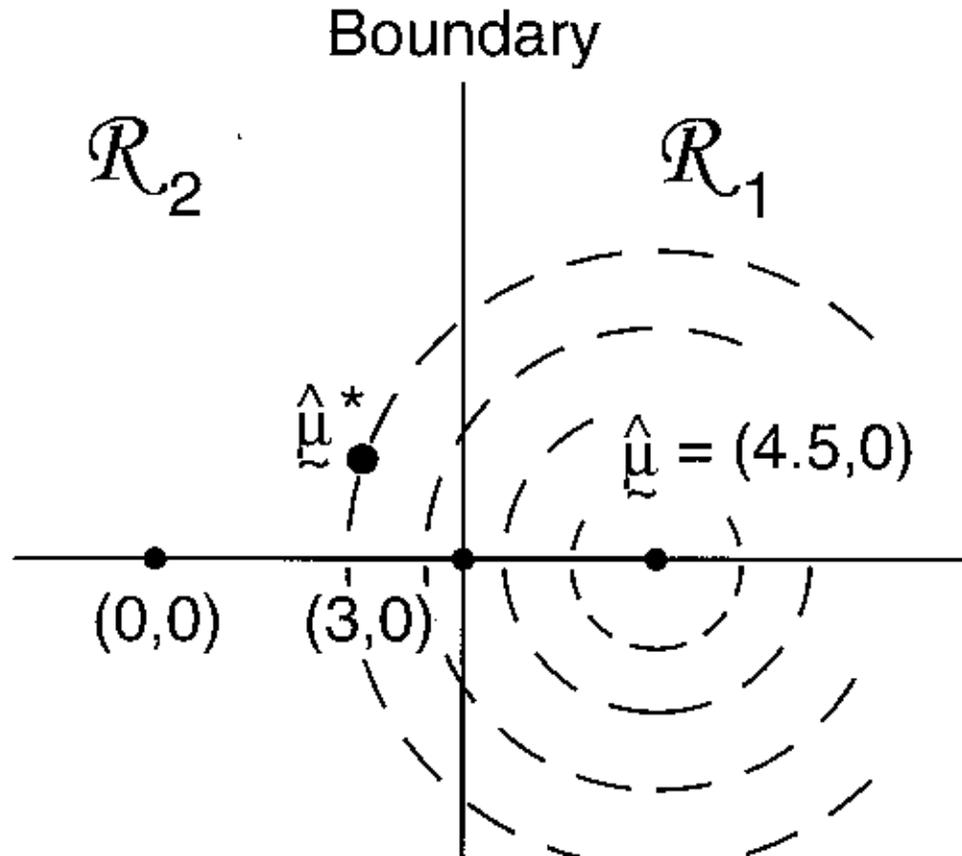
Bootstrap Proportion \neq Posterior Probability in general



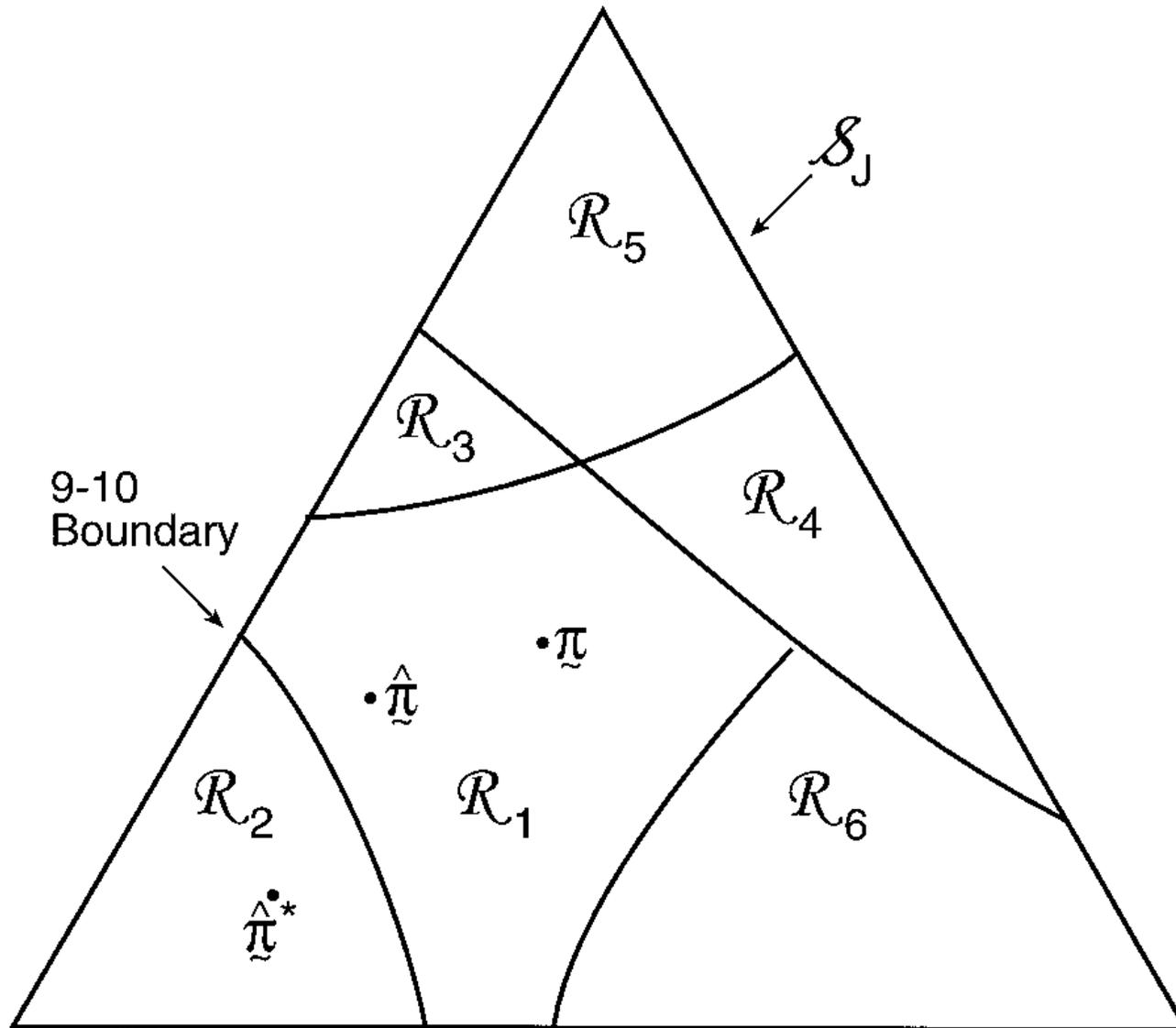
from [Wilcox et al. \(2002\)](#)

What did Efron et al. (1996) say?

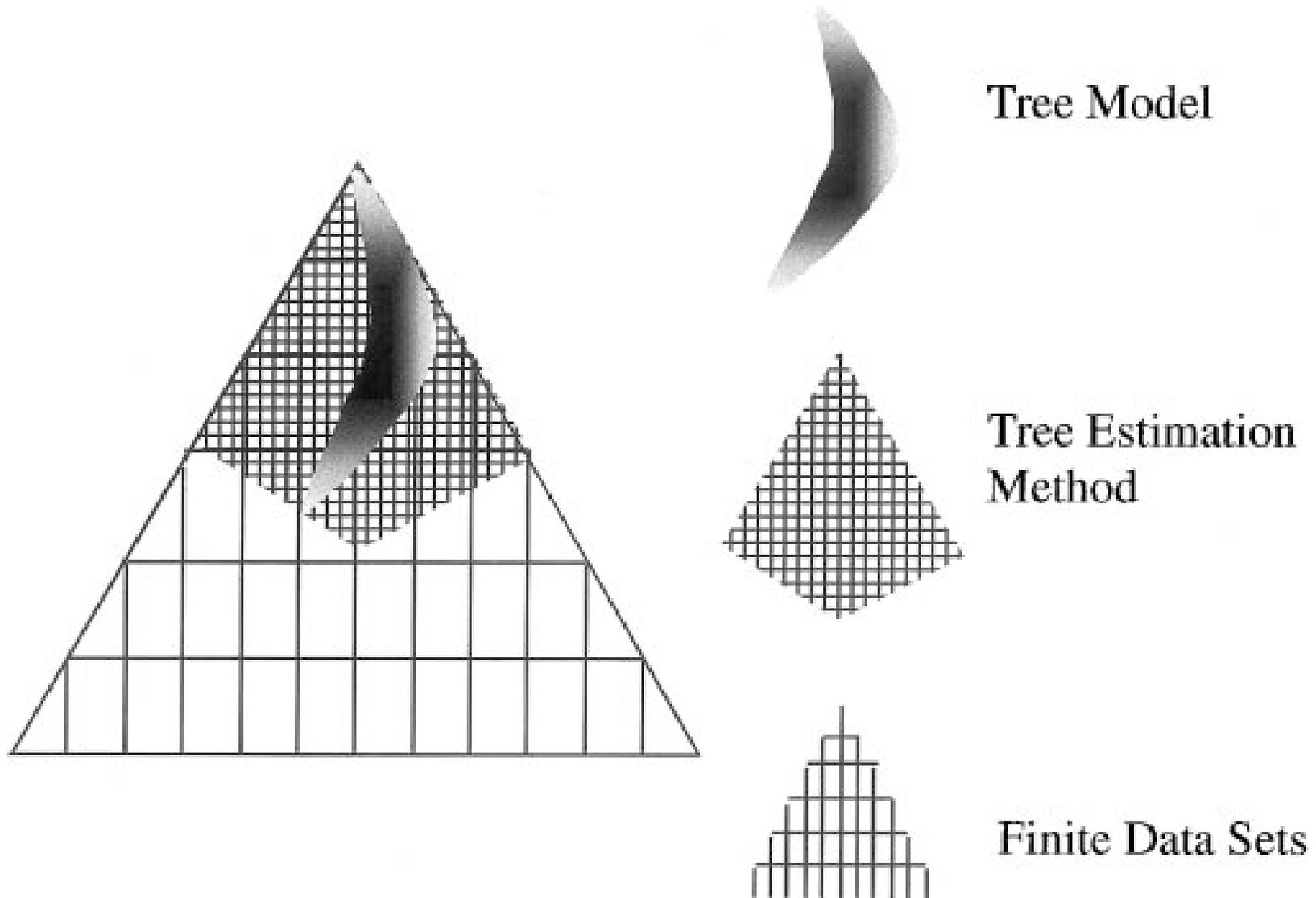
We can use a Bayesian model to show that $\tilde{\alpha}$ is a reasonable assessment of the probability that \mathcal{R}_1 contains μ . Suppose we believe *a priori* that μ could lie anywhere in the plane with equal probability. Then having observed $\hat{\mu}$, the *a posteriori* distribution of μ given $\hat{\mu}$ is $N_2(\hat{\mu}, I)$ exactly the same as the bootstrap distribution of $\hat{\mu}^*$. In other words, $\tilde{\alpha}$ is the *a posteriori* probability of the event $\mu \in \mathcal{R}_1$, if we begin with an “uninformative” prior density for μ .



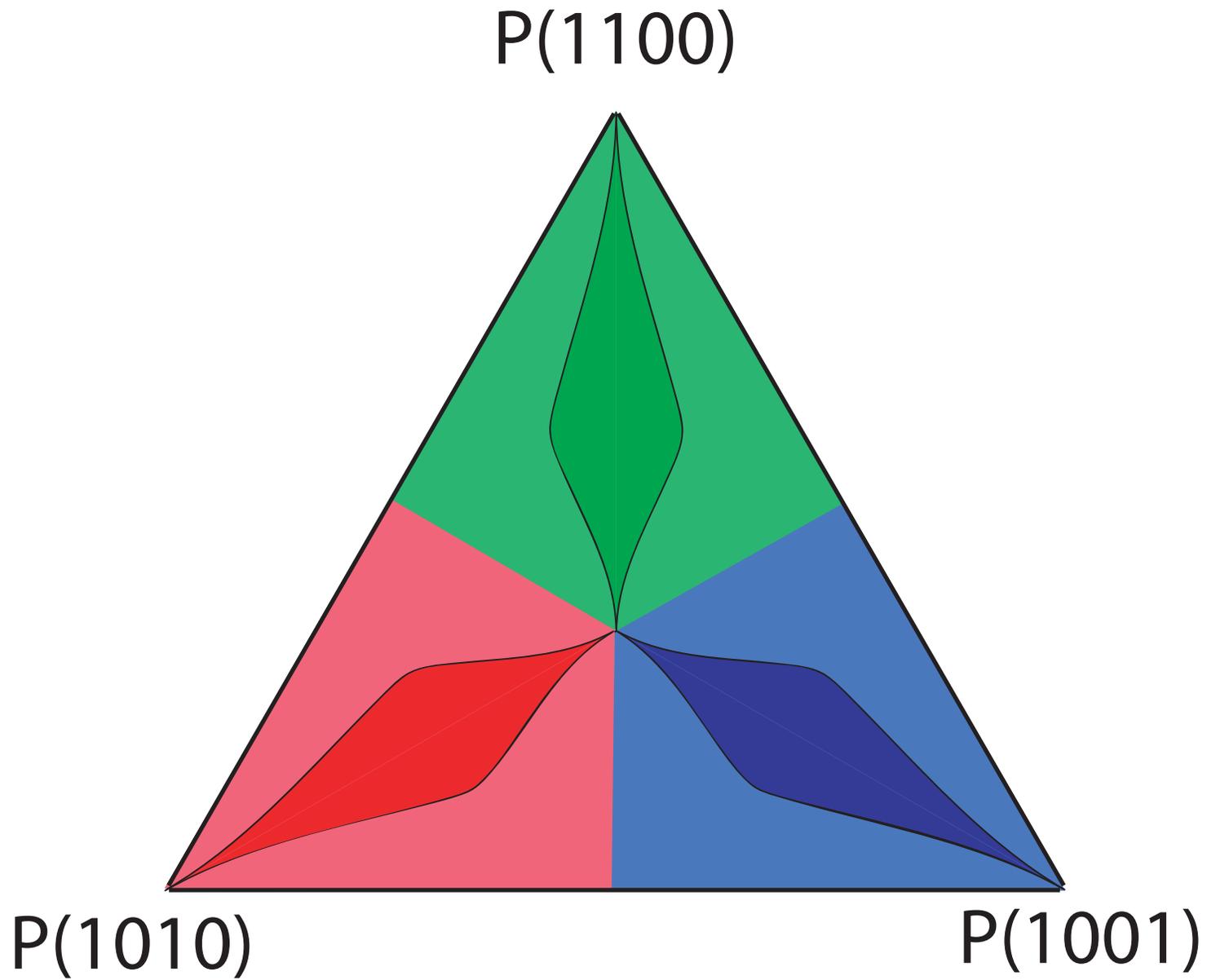
Efron et al. (1996) view of tree space



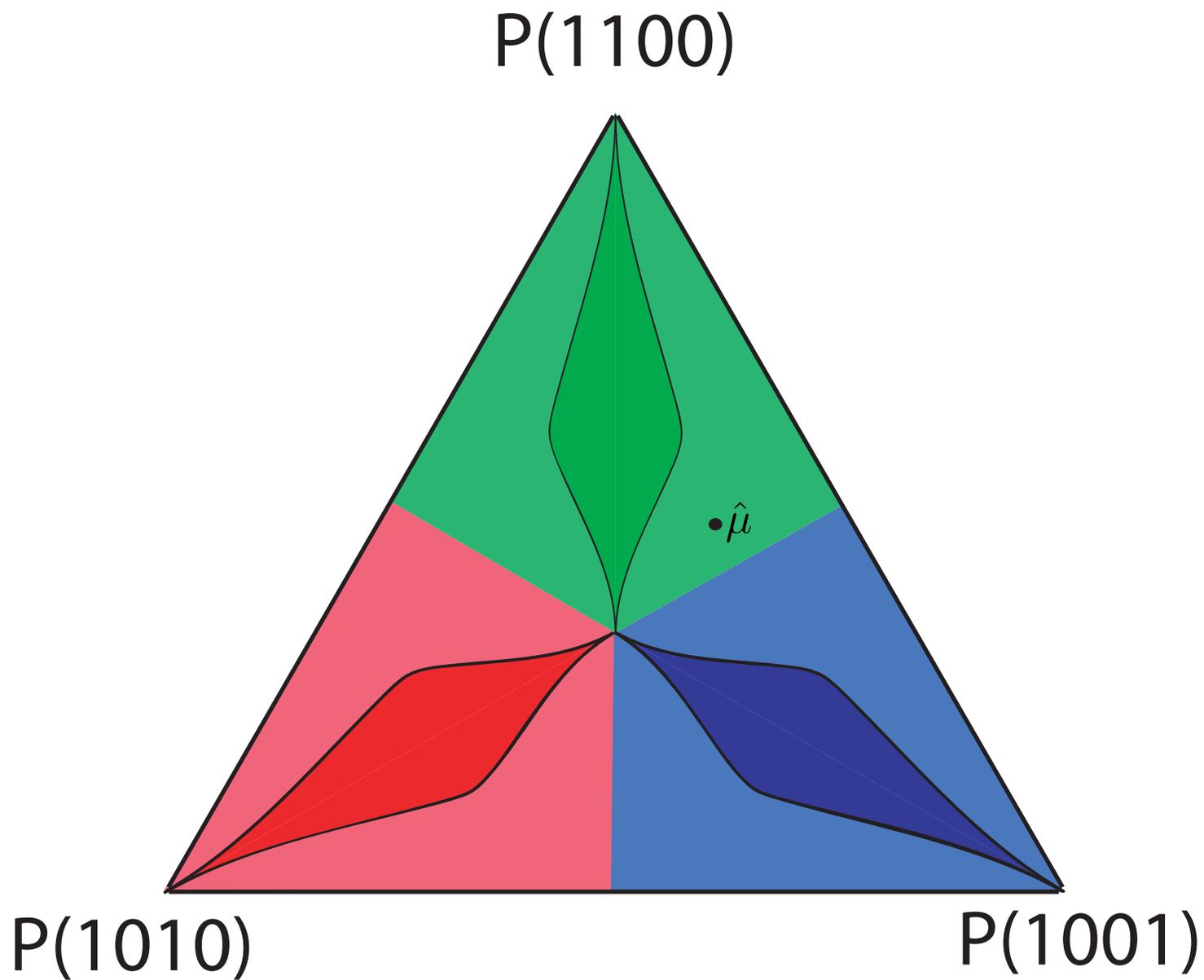
Kim (2000) view of tree space



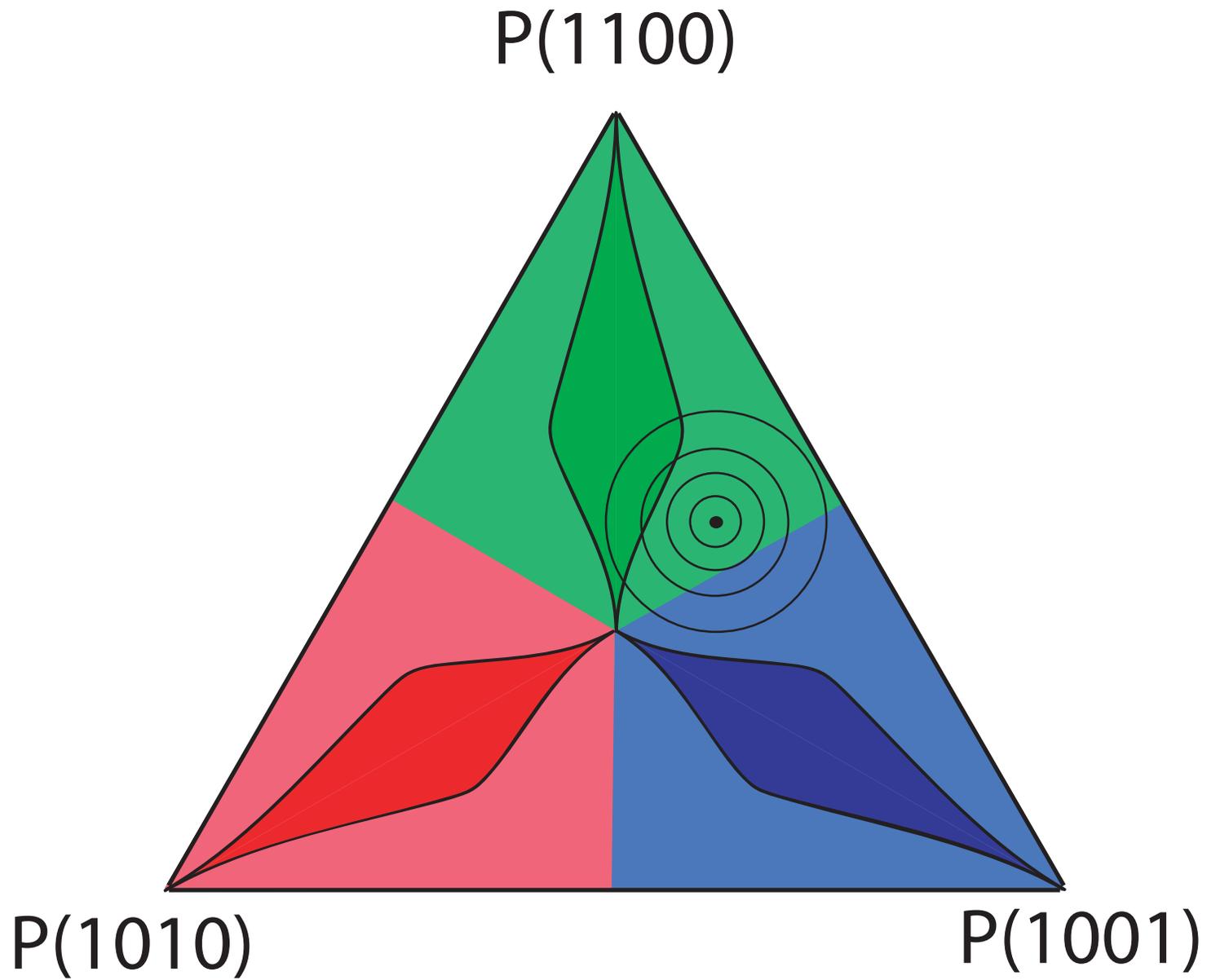
Parsimony-informative Pattern Frequency Space



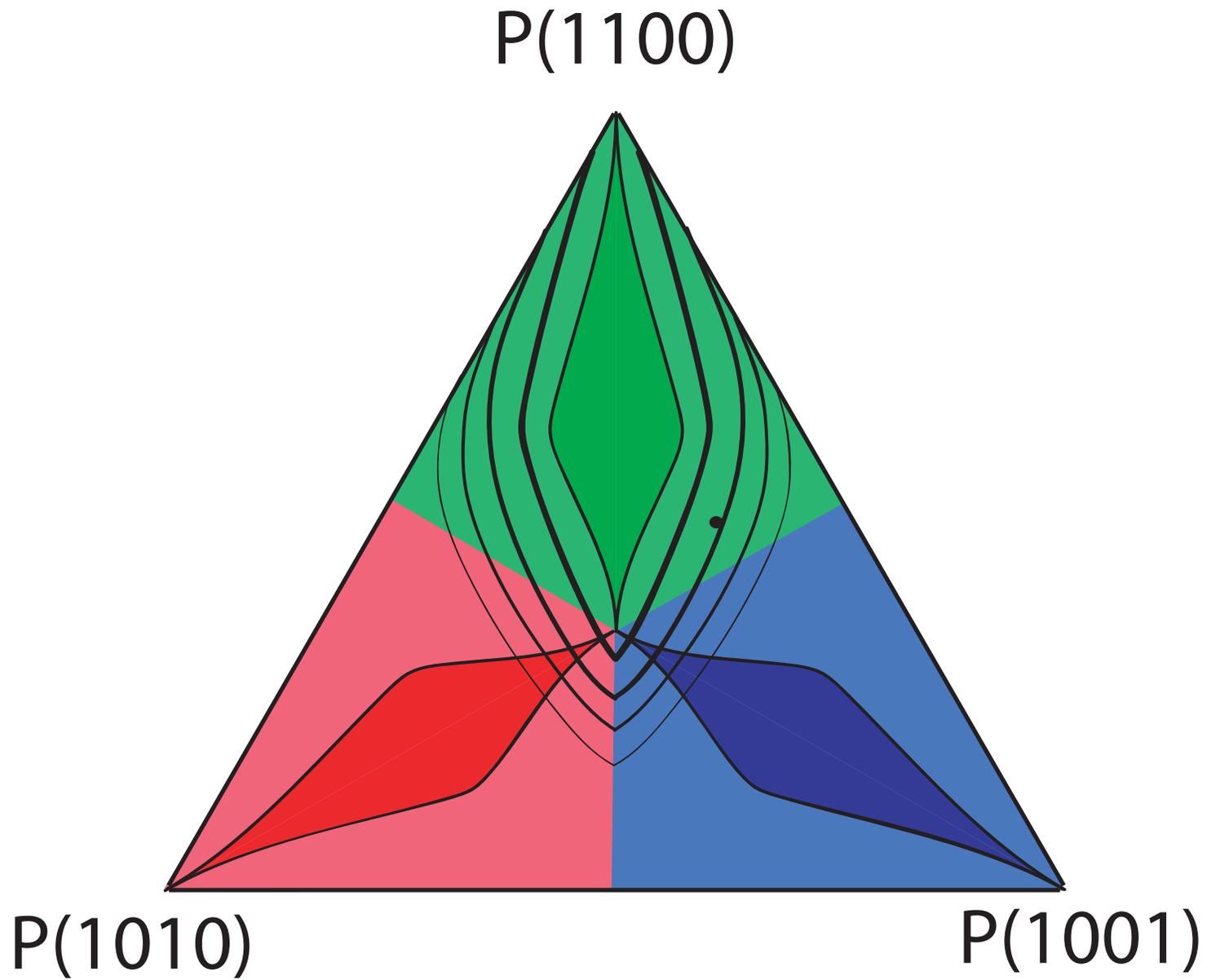
Pattern Frequency Space With Observed Data



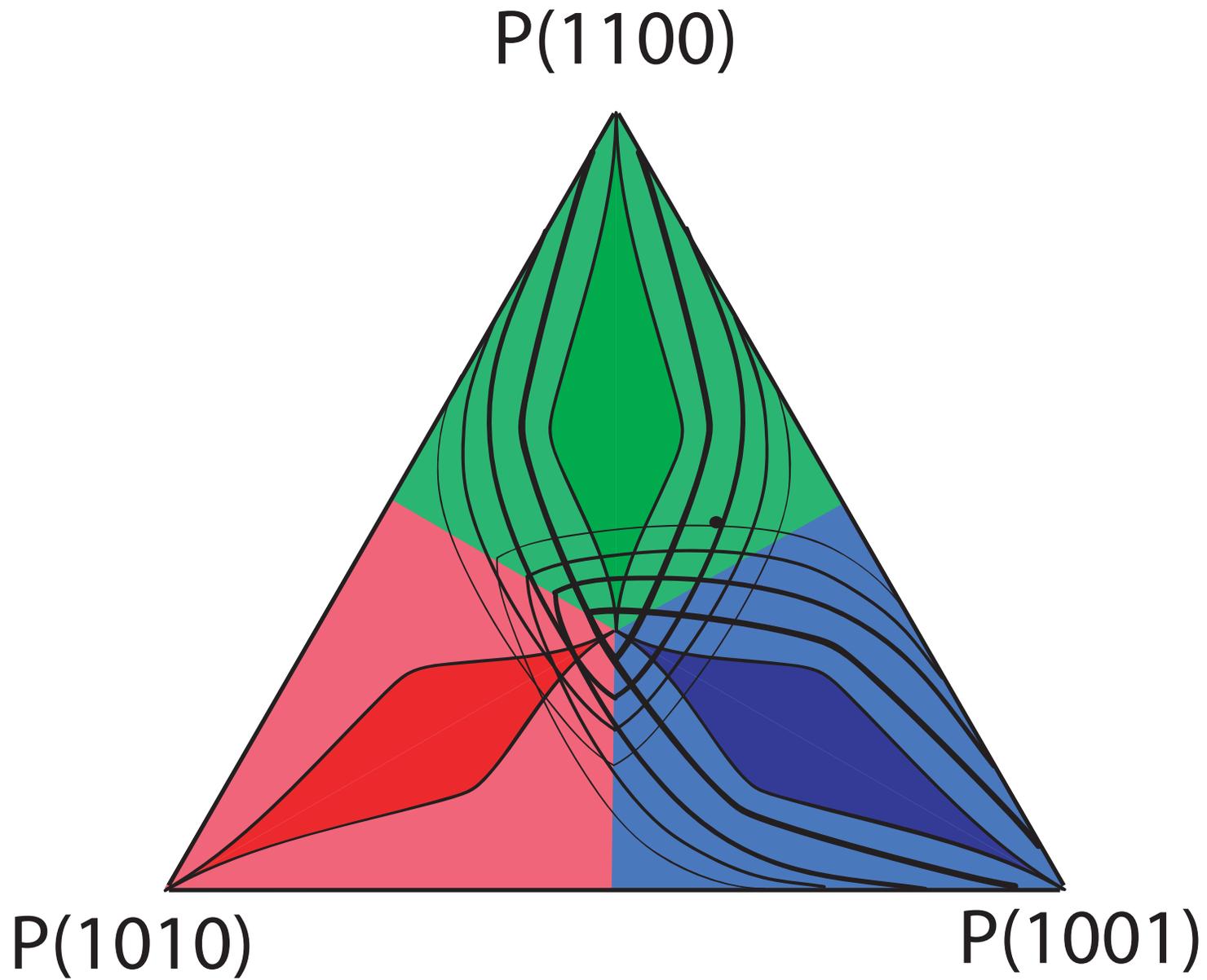
Bootstrapping in Pattern Frequency Space



Posterior Densities in Pattern Frequency Space



Posterior Densities in Pattern Frequency Space



What did Efron et al. (1996) say (and mean)?

- the “uninformative” prior density is a uniform prior over all of pattern space
- this is *not* equivalent to a prior that would be expected to yield a phylogeny (it is actually identical to the prior you would get if you assumed that all pairwise distances between taxa were ∞),
- Efron et al. (1996) were *not* predicting that the bootstrap proportions should be identical to those from a Bayesian phylogenetic analysis with real phylogenetic priors.
- Svennblad et al. (2006) have a nice paper on this subject.

Bootstrap proportions have been characterized as providing:

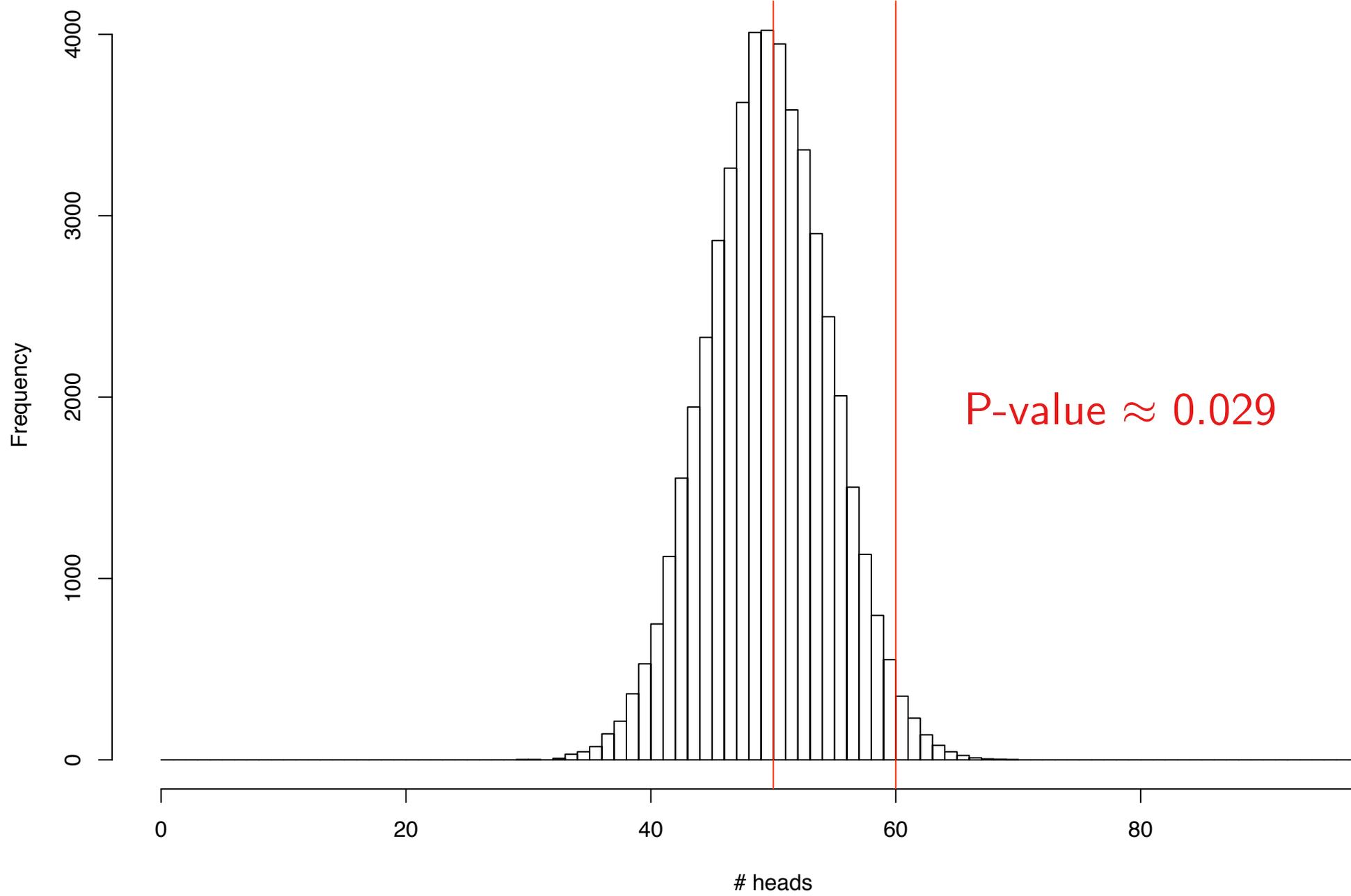
- a measure of repeatability,
- an estimate of the probability that the tree is correct (and criticized as being too conservative in this context),
- the P-value for a tree or clade

coin flipping (yet again)

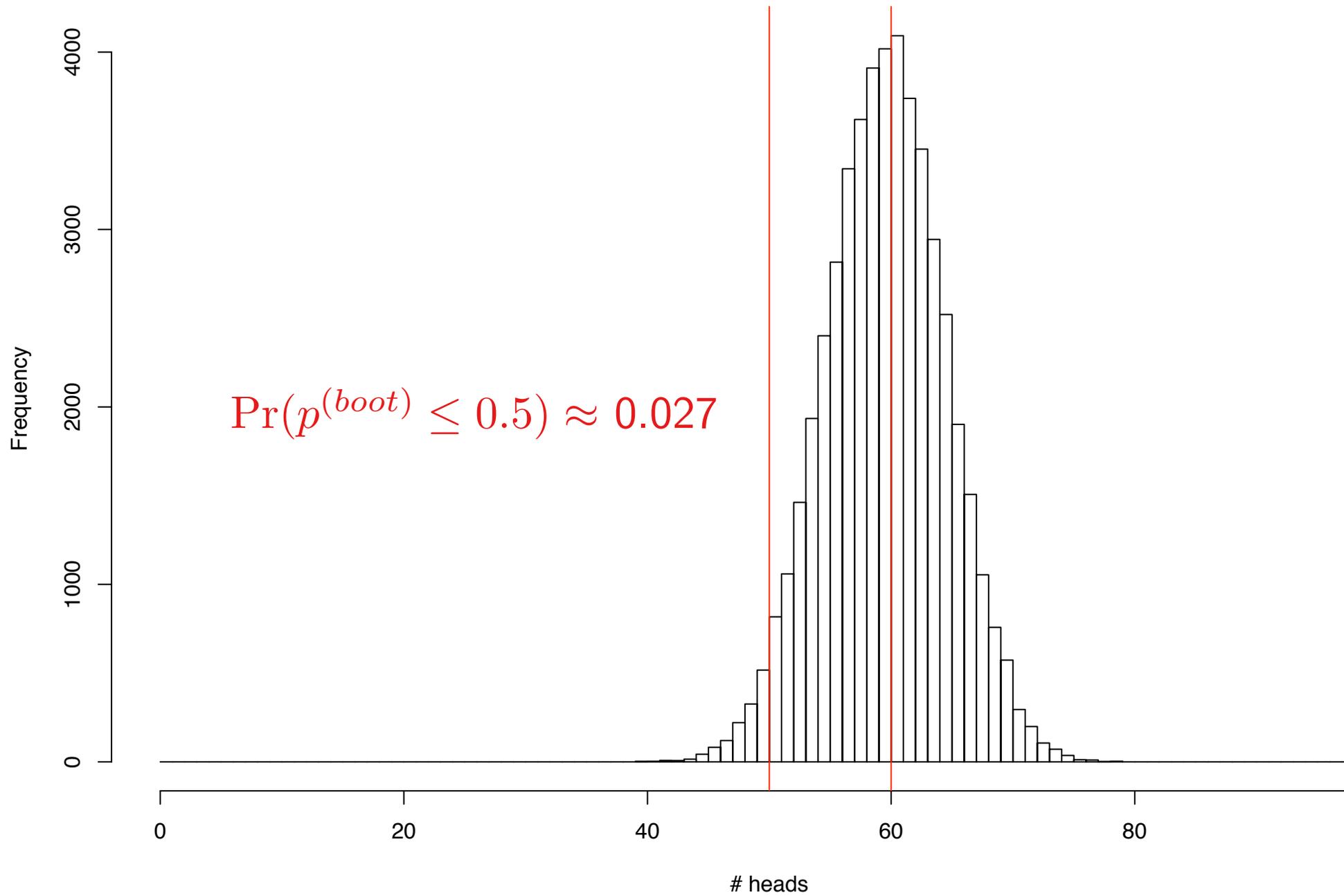
$N = 100$ and $H = 60$

Can we reject the hypothesis of a fair coin?

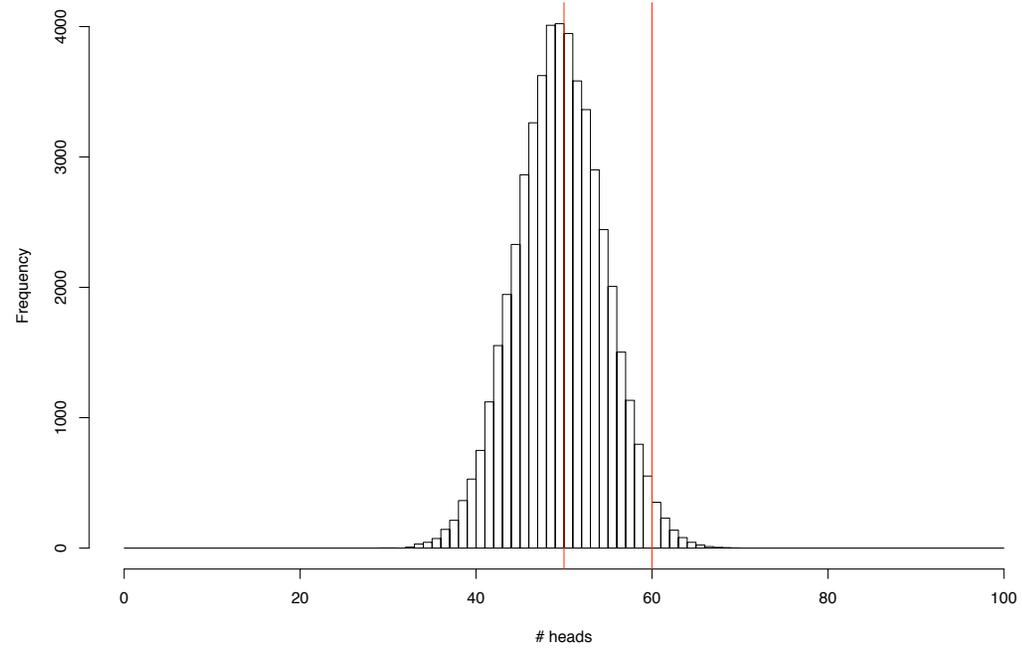
We can use simulation to generate the null distribution (we could actually use the binomial distribution to analytically solve this one)...



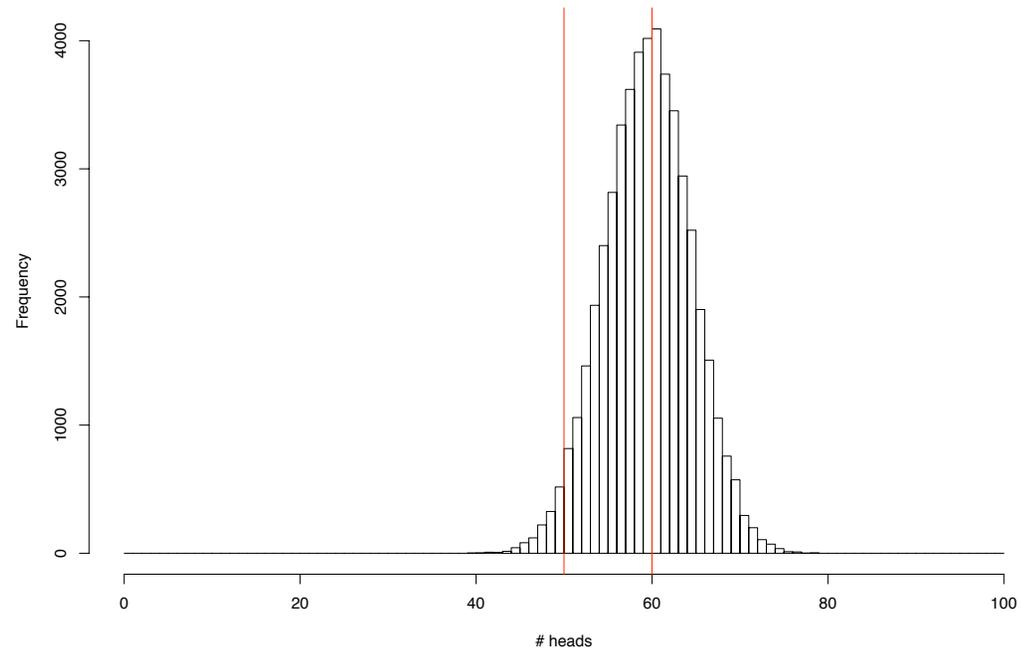
If we bootstrap we get a sense of the variability in our estimate, but we can also get a tail probability for $\Pr(p^{(boot)} \leq 0.5)$



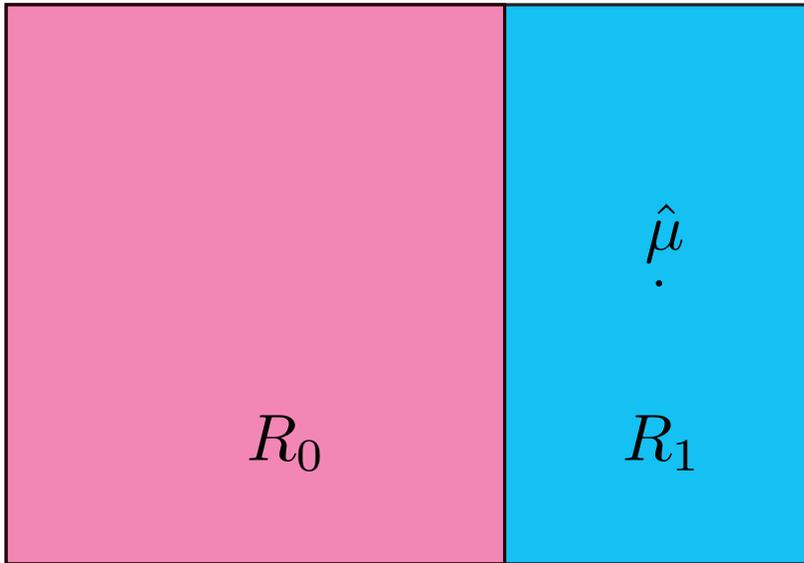
A simulation of the null distribution of the # heads



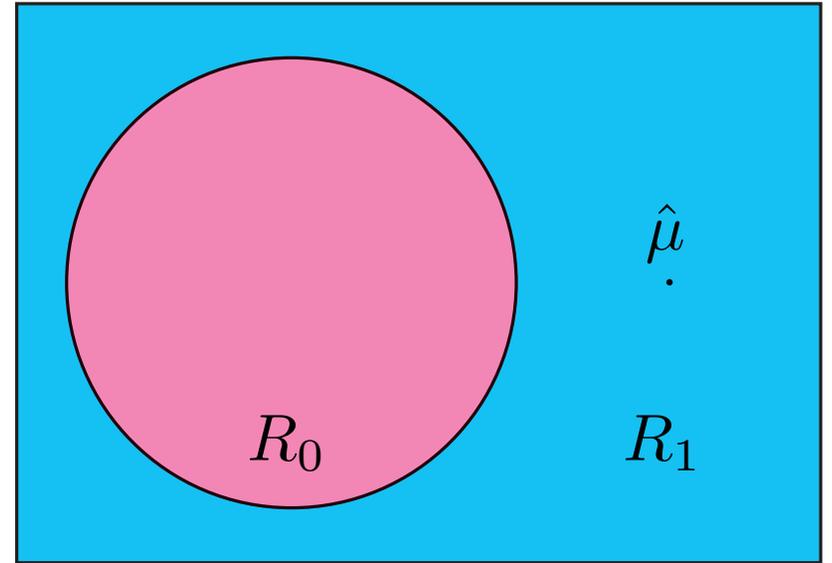
Distribution of the # heads in bootstrap resampled datasets



Case 1

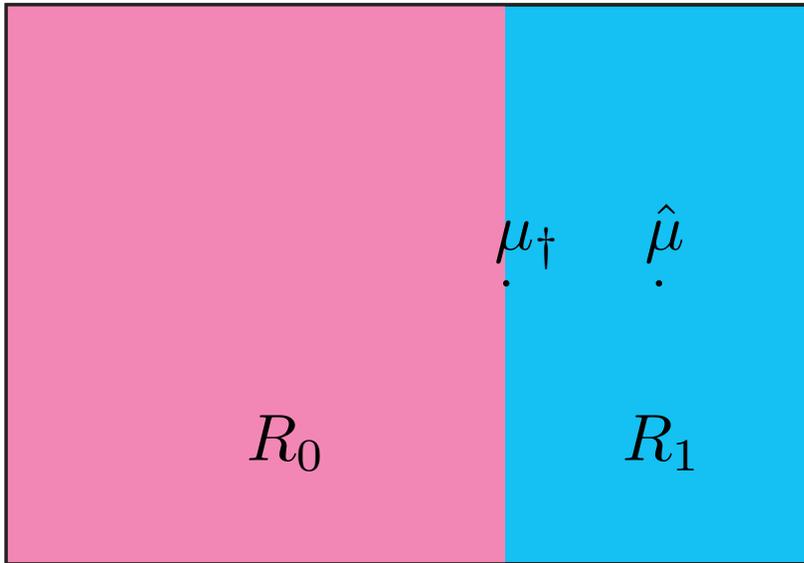


Case 2

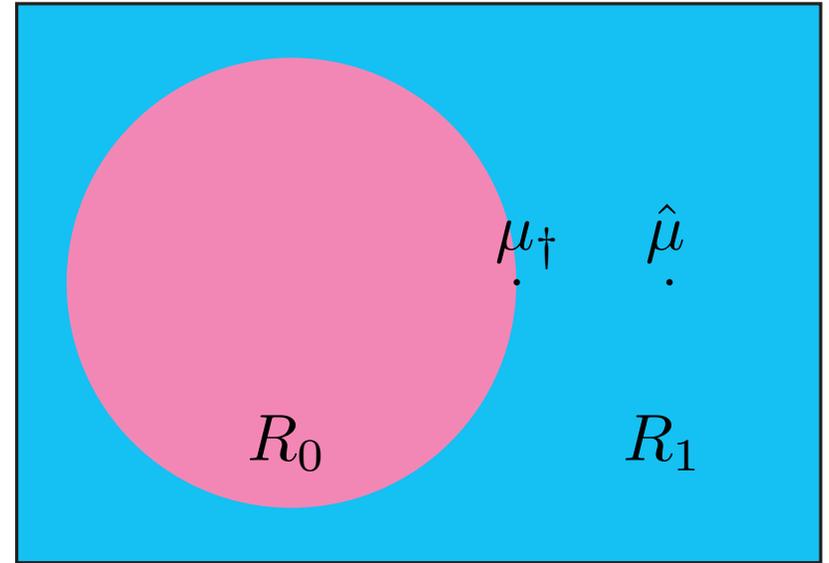


$\hat{\mu}$ is the best point calculated from the data

Case 1

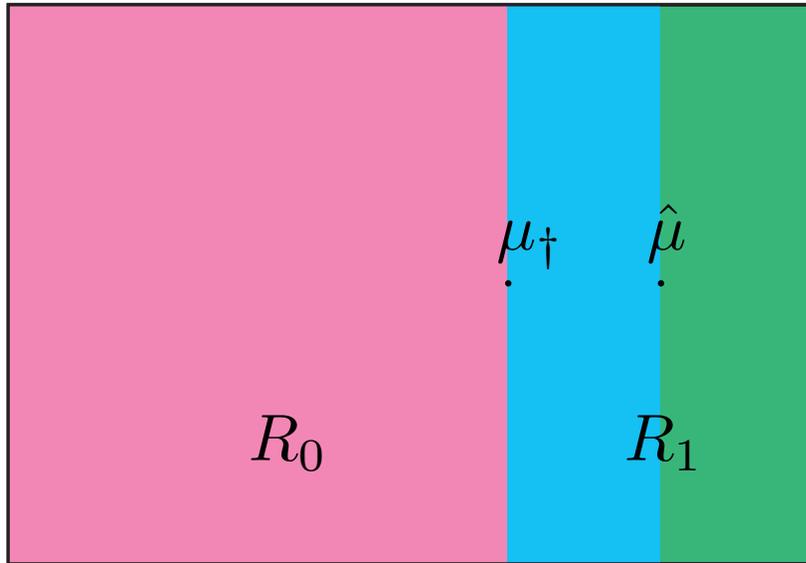


Case 2

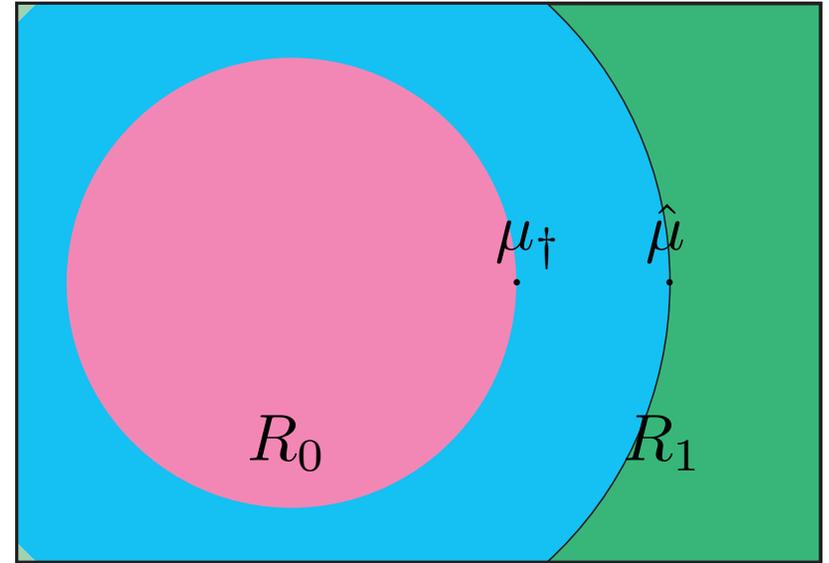


$\hat{\mu}$ is the best point calculated from the data
 μ_{\dagger} is least-favorable condition (LFC) point in R_0

Case 1



Case 2

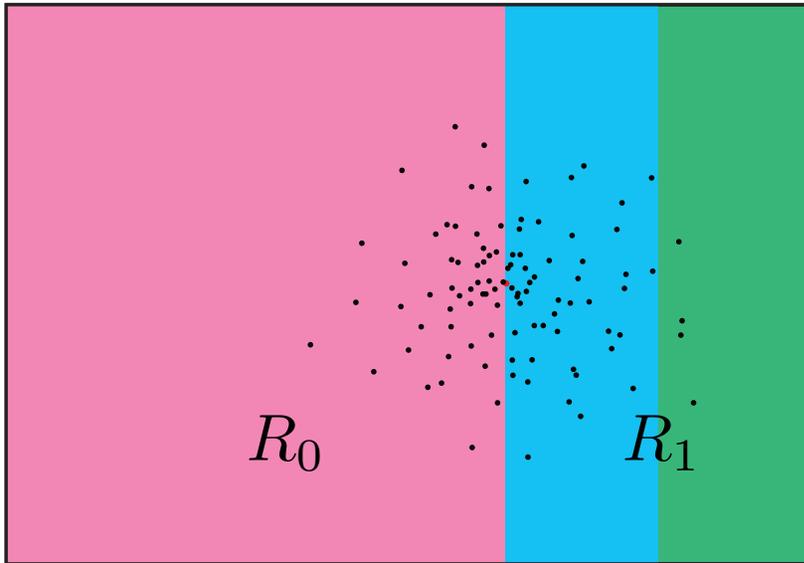


$\hat{\mu}$ is the best point calculated from the data

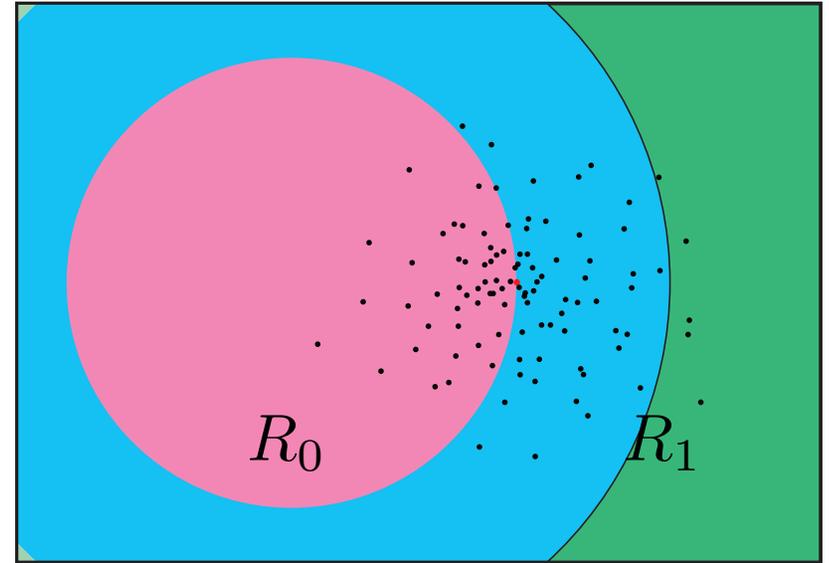
μ_{\dagger} is least-favorable condition (LFC) point in R_0

green areas are the tails - they correspond to values of the test statistic more extreme than $\hat{\mu}$ (relative to that $\mu \in R_0$)

Case 1

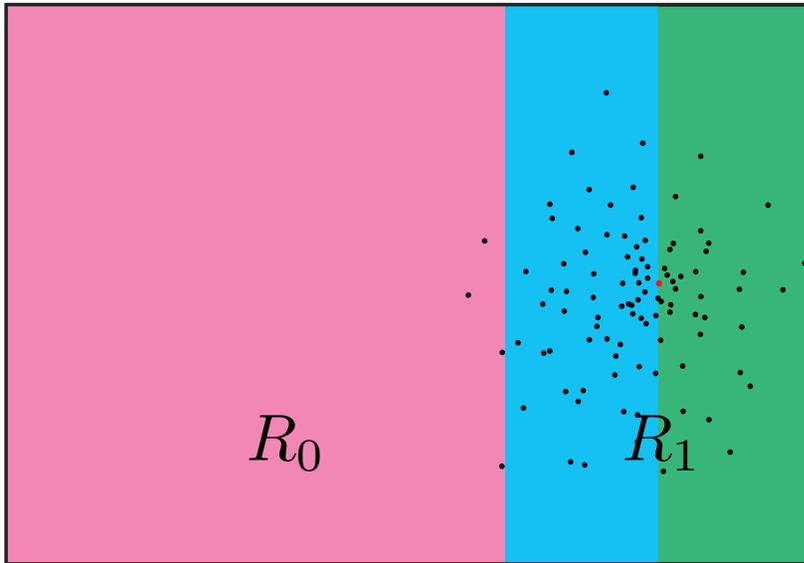


Case 2

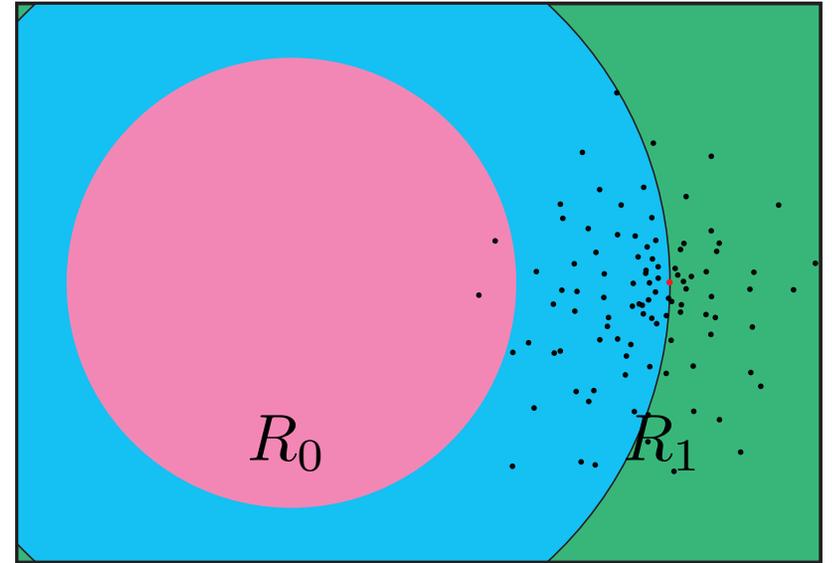


$\hat{\mu}$ is the best point calculated from the data
 μ_{\dagger} is least-favorable condition (LFC) point in R_0
Case 1 P-value < the P-value in Case 2

Case 1

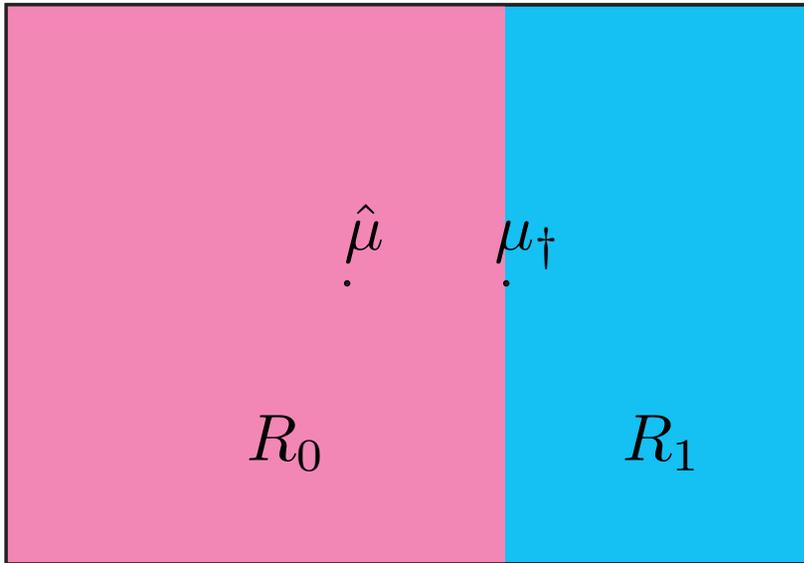


Case 2

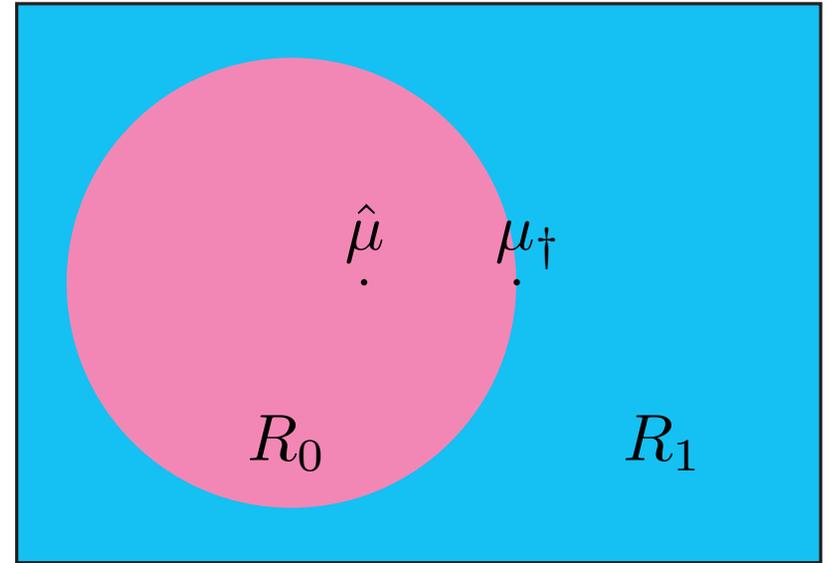


In case 1 - the bootstrap proportion is a good estimate of the P-value
In case 2 - the bootstrap proportion underestimates the P-value

Case 3

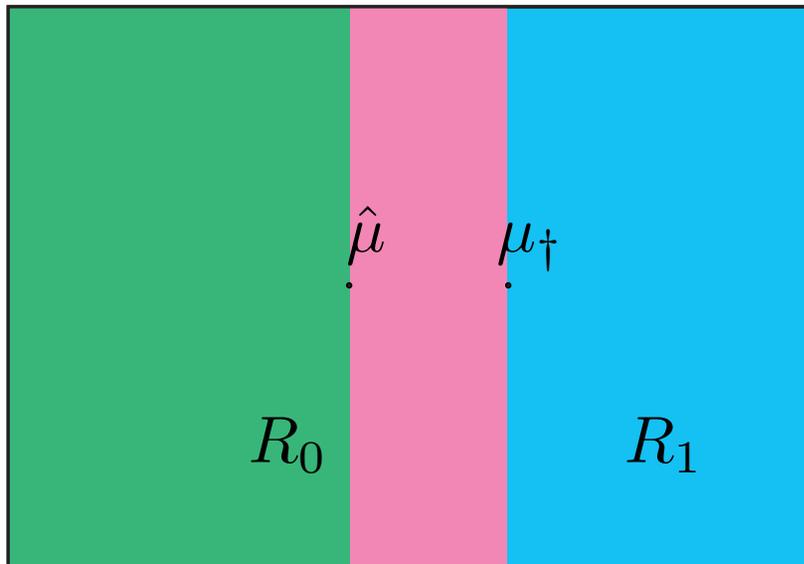


Case 4

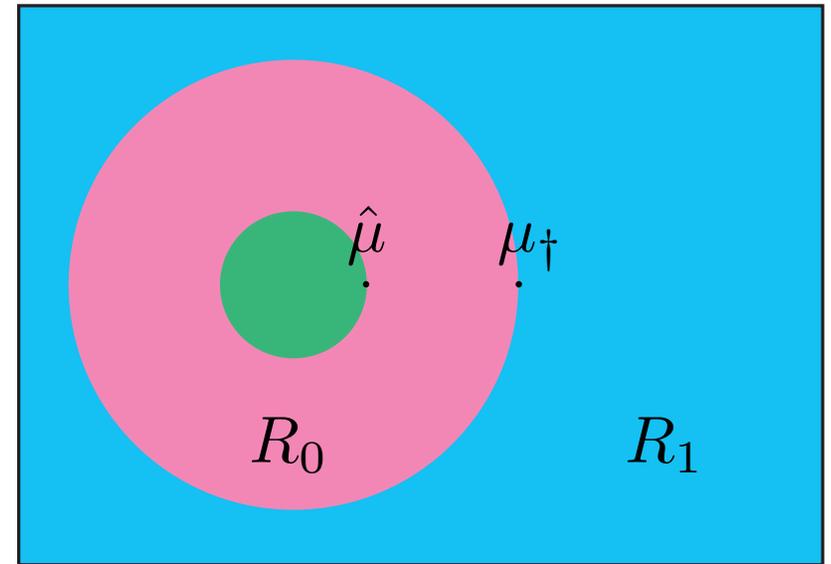


$\hat{\mu}$ is the best point calculated from the data
 μ_{\dagger} is least-favorable condition (LFC) point in R_1

Case 3



Case 4

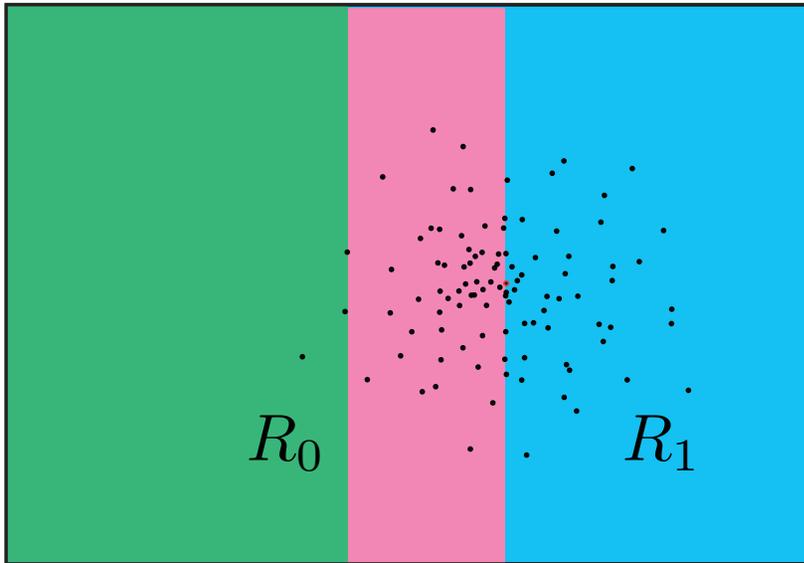


$\hat{\mu}$ is the best point calculated from the data

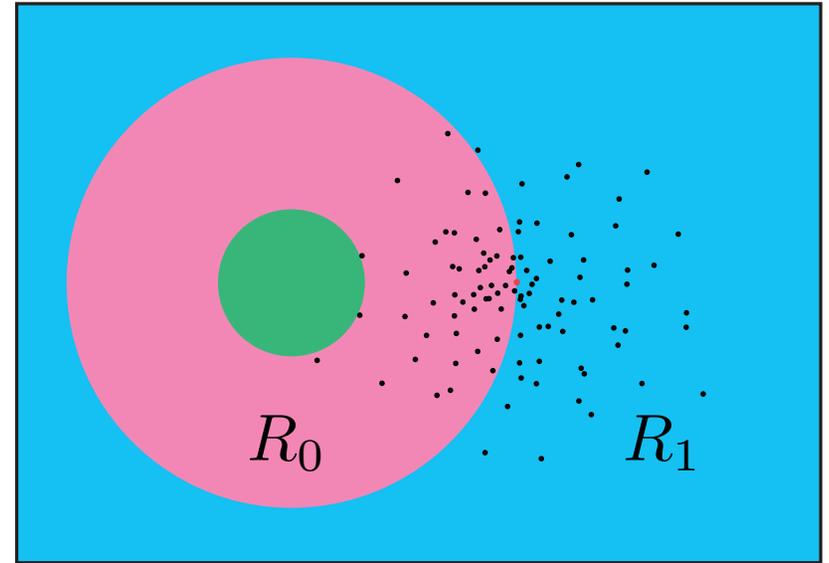
μ_{\dagger} is least-favorable condition (LFC) point in R_0

green areas are the tails - they correspond to values of the test statistic more extreme than $\hat{\mu}$ (relative to that $\mu \in R_1$)

Case 3

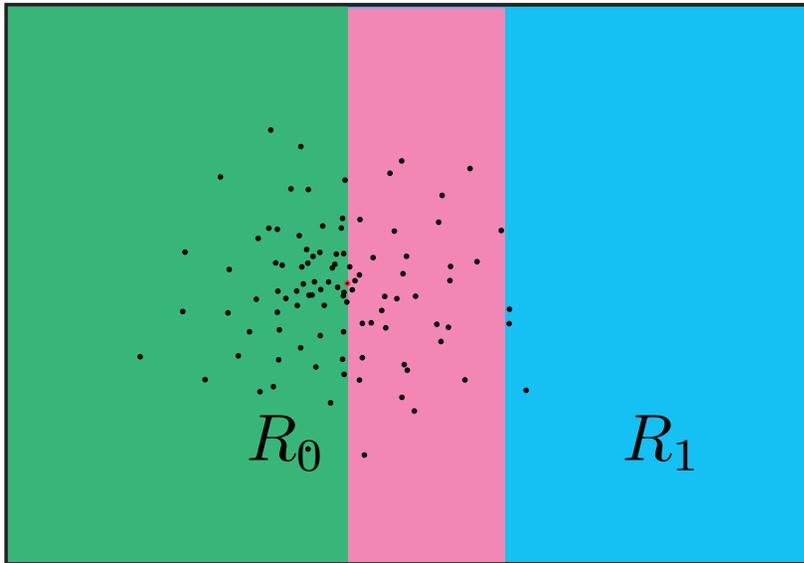


Case 4

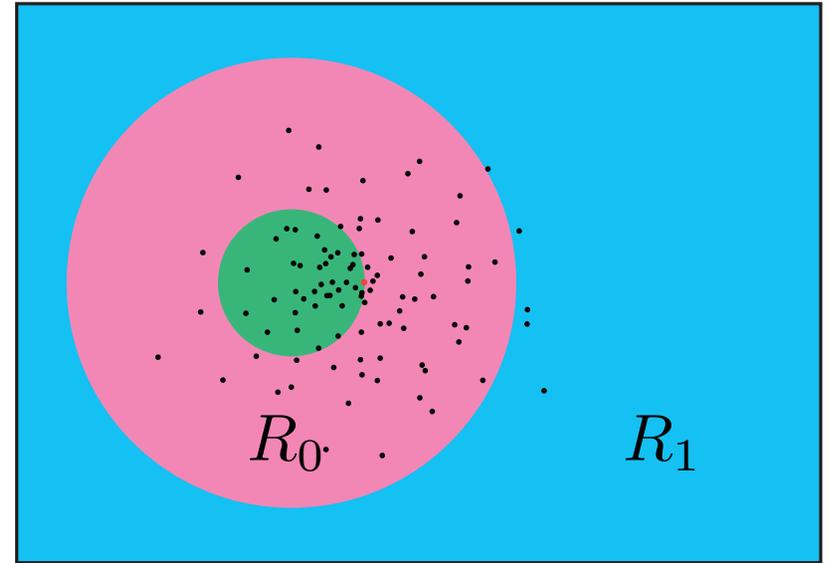


Case 3 P-value $>$ the P-value in Case 4

Case 3



Case 4

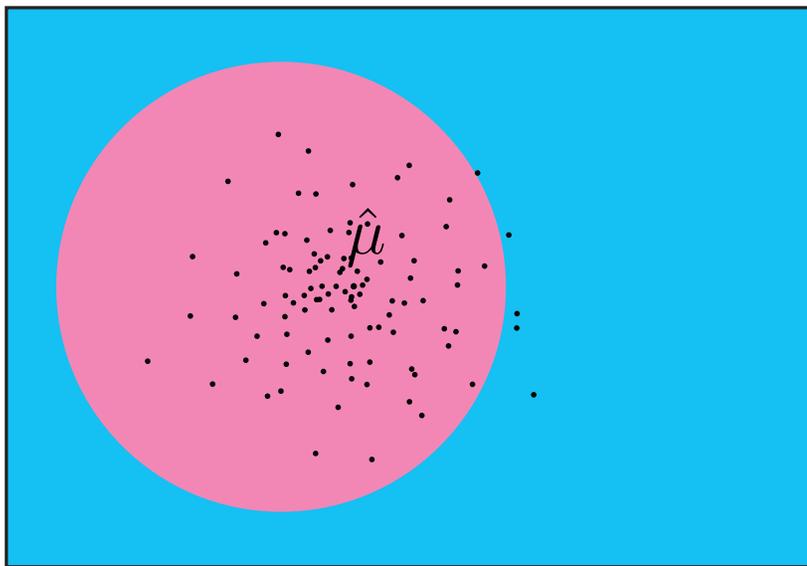


In case 3 - the bootstrap proportion is a good estimate of the P-value
In case 4 - the bootstrap proportion overestimates the P-value

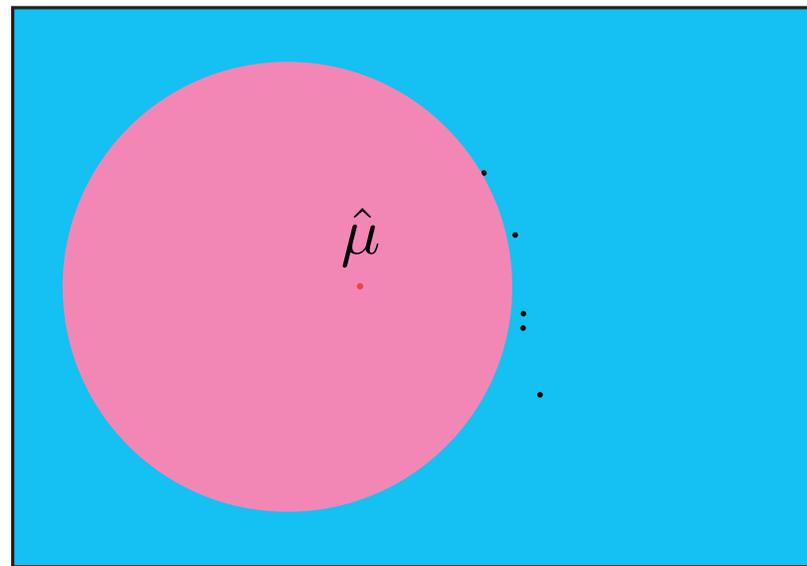
Efron et al. (1996) pointed out these issues of curvature of the boundaries between tree hypotheses.

We cannot see the boundaries in tree space, so it is hard to know how to correct for the biases so that we can use bootstrapping procedures as a means of getting a P-value for a clade – the probability that we would see this much support (or stronger support) for a clade if it were *not* present in the true tree.

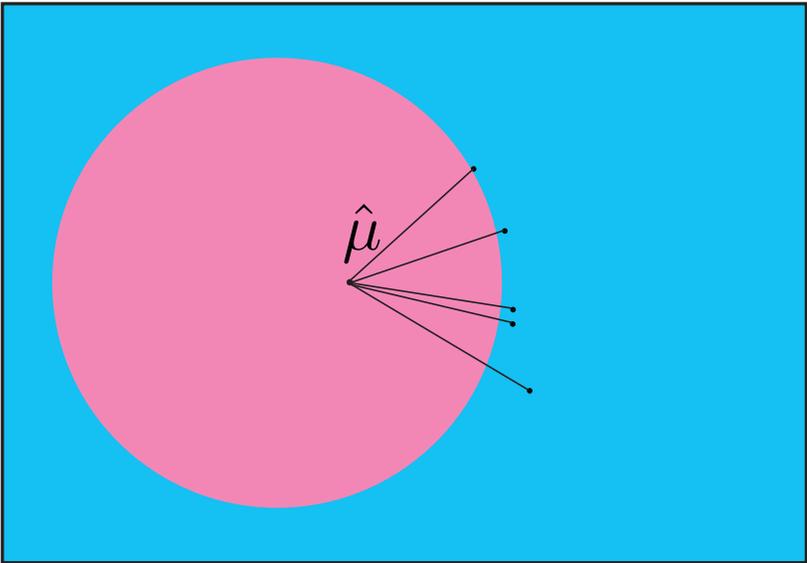
Initial bootstrap



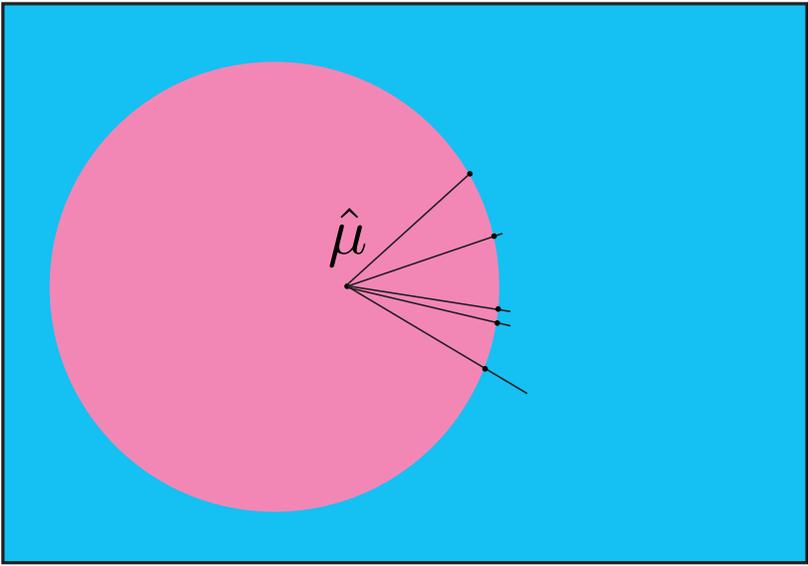
Find replicates that return a tree without the clade



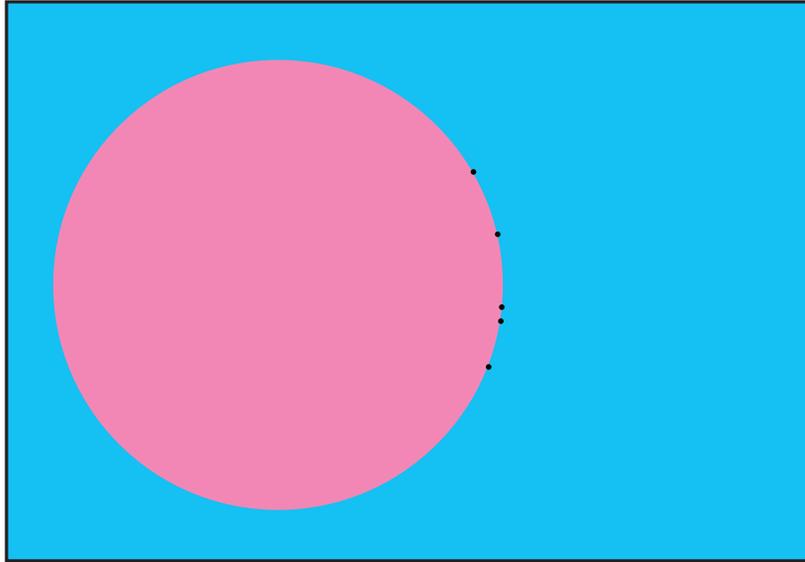
Find replicates that return a tree without the clade



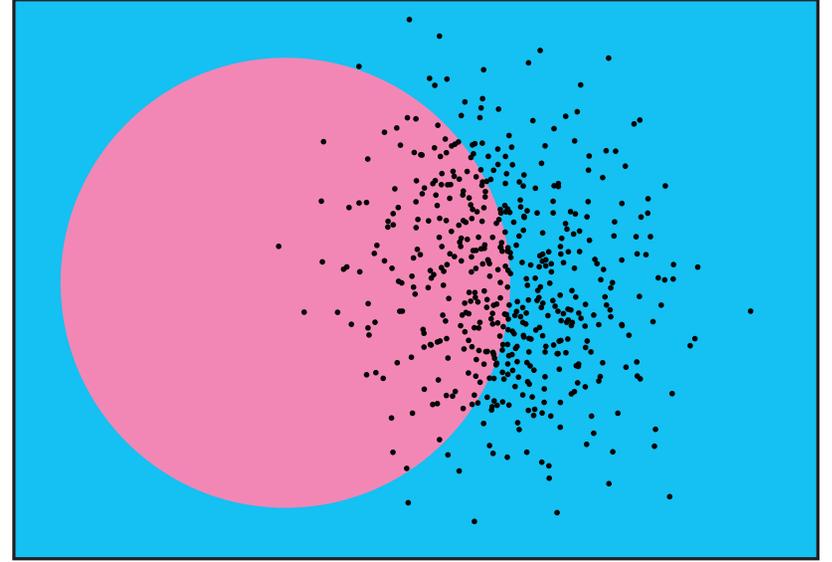
Find boundary points between regions



Find boundary points between regions



Bootstrap from these boundary points to check curvature of the boundary



The corrected bootstrap procedure of Efron et al. (1996) requires a very large number of bootstrap replicates because you need very accurate estimates of the curvature in order to apply the correction. Shimodaira (2002) expanded on this work:

- d is the distance from the point that corresponds to the data and the closest point on the boundary between another tree
- $\Phi(\cdot)$ denotes the cumulative density function of the standard Normal(0,1) distribution.
- c denotes the curvature of the boundary
- the P-value for the KH test is given by $KH = \Phi(d)$

- Shimodaira argues (from an early Efron paper) that the appropriate P-value for tree selection is:

$$AU = 1 - \Phi(d - c)$$

- In “standard” non parametric bootstrapping proportions are:

$$BP = 1 - \Phi(d + c)$$

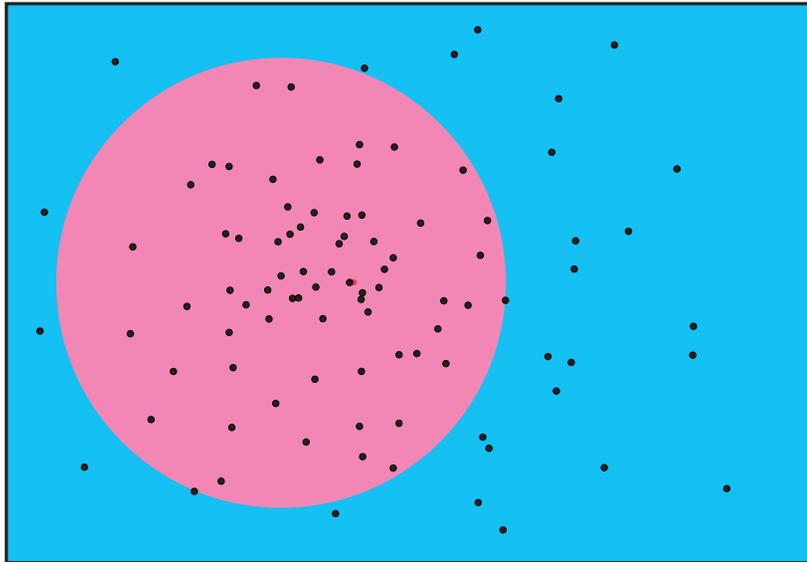
Note the incorrect sign with respect to the curvature term causes BP (and recall how on the curved boundary examples, the curvature caused the P-value to change in one direction and the BP to go in the other).

How can we find c so that we can correct for it?

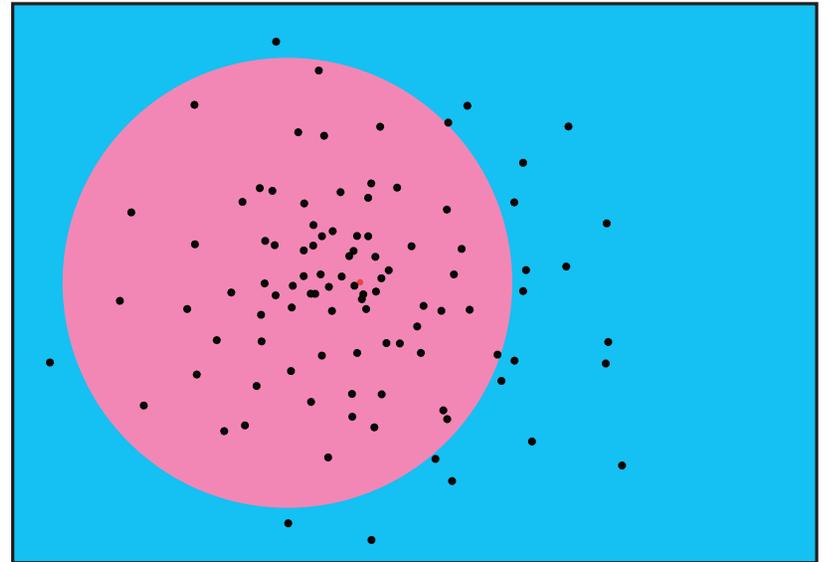
- N is the number of characters in the real data set
- N' is the number of characters in each bootstrap data set
- $r = \frac{N'}{N}$
- If you do a bootstrap in which $r \neq 1$, Shimodaira determined the expected effect on the bootstrap proportion as a function of d and c :

$$BP(r) = 1 - \Phi \left(d\sqrt{r} + \frac{c}{\sqrt{r}} \right)$$

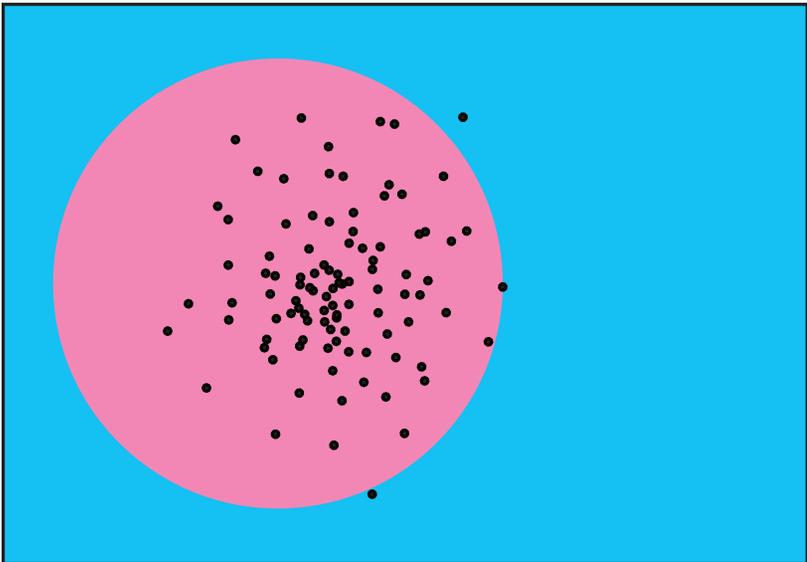
$r = 0.5$



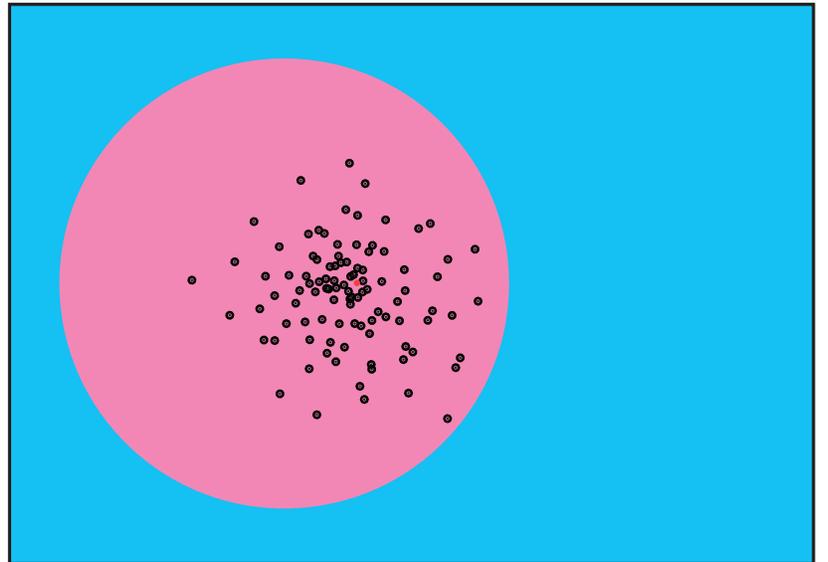
$r = 0.8$



$r = 1.0$



$r = 1.2$



AU Test

1. conduct a sweep of bootstraps with r varying (for instance $r = 0.5, r = 0.6, r = 0.7, \dots r = 1.4$, to get a set of $BP(r)$ for a tree.
2. Use weighted least squares to estimate c and d from the set of $BP(r)$
3. Calculate

$$AU = 1 - \Phi(d - c)$$

This lets you calculate a P-value for any tree of interest, and then you can construct a confidence set of trees.

References

- Alfaro, M. E., Zoller, S., and Lutzoni, F. (2003). Bayes or bootstrap? a simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, 20(2):255–266.
- Berry, V. and Gascuel, O. (1996). On the interpretation of bootstrap trees: Appropriate threshold of clade selection and induced gain. *Molecular Biology and Evolution*, 13(7):999–1011.
- Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A., and Winka, K. (2003). Comparing bootstrap and

posterior probability values in the four-taxon case. *Systematic Biology*, 52(4):477–487.

Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F., and Douzery, E. J. P. (2003). Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution*, 20(2):248–254.

Efron, B., Halloran, E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Science, U. S. A.*, 93:13429–13434.

Kim, J. (2000). Slicing hyperdimensional oranges: The geometry of phylogenetic estimation. *Molecular Phylogenetics and Evolution*, 17(1):58–75.

Shimodaira, H. (2002). An approximately unbiased test of

phylogenetic tree selection. *Systematic Biology*, 51(3):492–508.

Suzuki, Y., Glazko, G. V., and Nei, M. (2002). Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *PNAS*, 99:16138–16143.

Svennblad, B., Erixson, P., Oxelman, B., and Britton, T. (2006). Fundamental differences between the methods of maximum likelihood and maximum posterior probability in phylogenetics. *Systematic Biology*, 55(1):116–121.

Wilcox, T. P., Zwickl, D. J., Heath, T., and Hillis, D. M. (2002). Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution*, 25:361–371.