Week 7: Bayesian inference

Genome 570

February, 2016

Bayes' Theorem

Conditional probability of hypothesis given data is:

$$\text{Prob}\,(\mathsf{H} \mid \mathsf{D}) \;=\; \frac{\text{Prob}\,(\mathsf{H}\ \&\ \mathsf{D})}{\text{Prob}\,(\mathsf{D})}$$

Since

$$\text{Prob}\,(\mathsf{H}\ \&\ \mathsf{D}) \;=\; \text{Prob}\,(\mathsf{H})\ \ \text{Prob}\,(\mathsf{D} \mid \mathsf{H}),$$

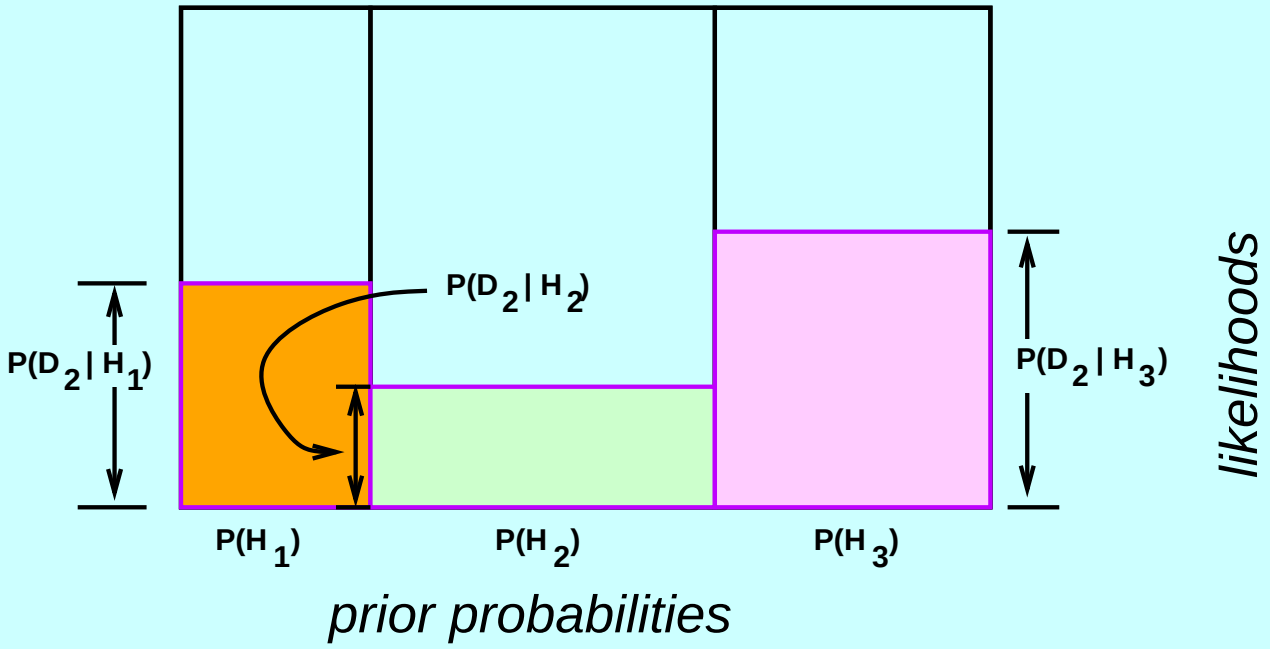
substituting this in:

$$\text{Prob}\,(\mathsf{H} \mid \mathsf{D}) \;=\; \frac{\text{Prob}\,(\mathsf{H})\ \ \text{Prob}\,(\mathsf{D} \mid \mathsf{H})}{\text{Prob}\,(\mathsf{D})}$$

The denominator $\text{Prob}\,(\mathsf{D})$ is the sum of the numerators over all possible hypotheses H, so it is the sum of the numerators over those, giving

$$\text{Prob}\,(\mathsf{H} \mid \mathsf{D}) \;=\; \frac{\text{Prob}\,(\mathsf{H})\ \ \text{Prob}\,(\mathsf{D} \mid \mathsf{H})}{\sum_{\mathsf{H}}\ \text{Prob}\,(\mathsf{H})\ \ \text{Prob}\,(\mathsf{D} \mid \mathsf{H})}$$

A visual example of Bayes' Theorem

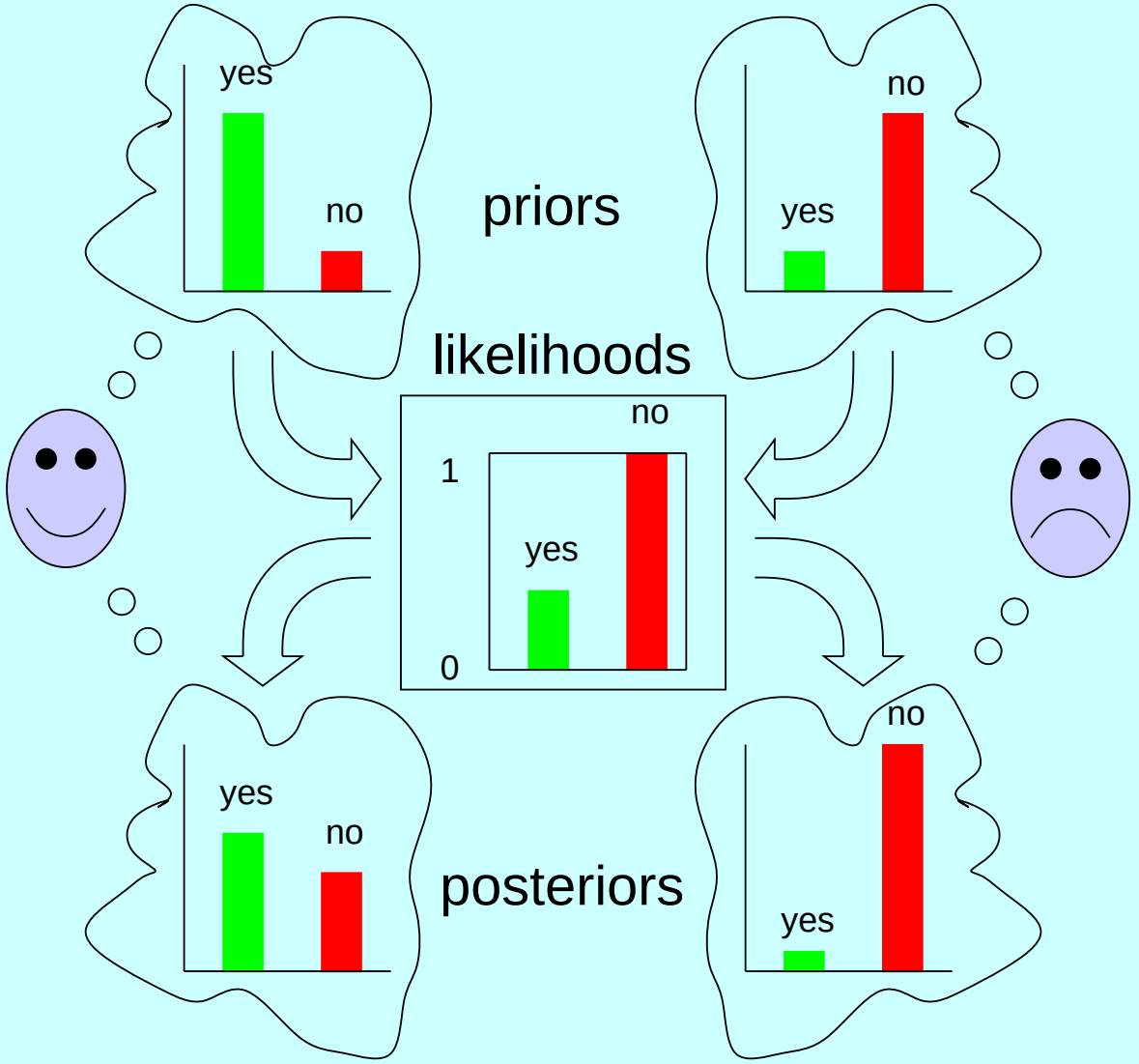


posterior probability

$$P(H_3|D_2) = \frac{\square}{\square + \square + \square}$$

A dramatic, if not real, example

Example: "Space probe photos show no Little Green Men on Mars!"



Calculations for that example

Using the Odds Ratio form of Bayes' Theorem:

$$\underbrace{\frac{\text{Prob}\,(H_1|D)}{\text{Prob}\,(H_2|D)}}_{\text{posterior odds ratio}} = \underbrace{\frac{\text{Prob}\,(D|H_1)}{\text{Prob}\,(D|H_2)}}_{\text{likelihood ratio}} \underbrace{\frac{\text{Prob}\,(H_1)}{\text{Prob}\,(H_2)}}_{\text{prior odds ratio}}$$

For the odds favoring their existence, the calculation is, for the optimist about Little Green Men:

$$\frac{4}{1} \times \frac{1/3}{1} = \frac{4/3}{1} = 4:3$$

While for the pessimist it is

$$\frac{1}{4} \times \frac{1/3}{1} = \frac{1/12}{1} = 1:12$$

With repeated observation the prior matters less

If we send 5 space probes, and all fail to see LGMs, since the probability of this observation is $(1/3)^5$ if there are LGMs, and 1 if there aren't, we get for the optimist about Little Green Men:

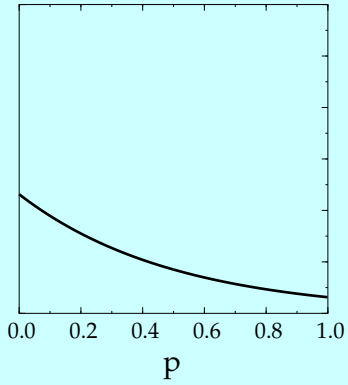
$$\frac{4}{1} \times \frac{(1/3)^5}{1} \quad = = \quad \frac{4/243}{1} \quad = \quad 4 : 243$$

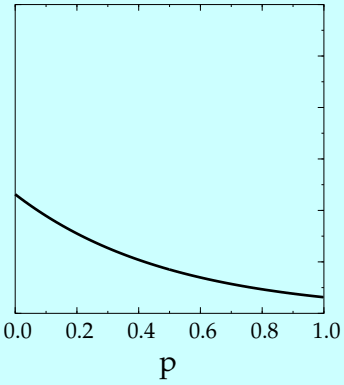
while for the pessimist about Little Green Men:

$$\frac{1}{4} \times \frac{(1/3)^5}{1} \quad = = \quad \frac{1/972}{1} \quad = \quad 1 : 972$$

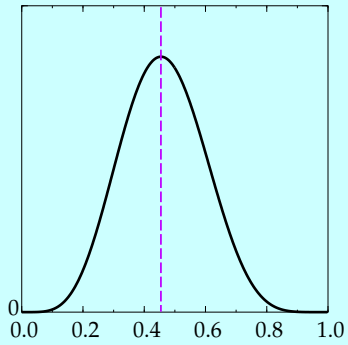
A coin tossing example

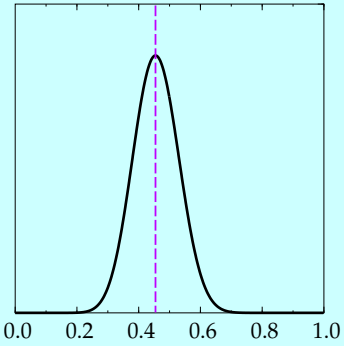
The prior



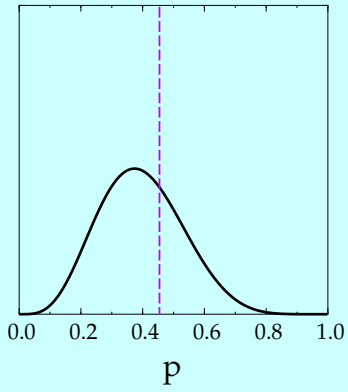


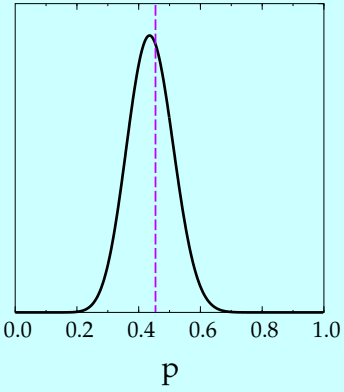
The likelihood function





The posterior





11 tosses with 5 heads 44 tosses with 20 heads

Markov Chain Monte Carlo sampling

To draw trees from a distribution whose probabilities are proportional to $f(t)$, we can use the Metropolis algorithm:

1. Start at some tree. Call this T_i.
2. Pick a tree that is a neighbor of this tree in the graph of trees. Call this the proposal T_j.
3. Compute the ratio of the probabilities (or probability density functions) of the proposed new tree and the old tree:

$$R = \frac{f(T_j)}{f(T_i)}$$

4. If $R \geq 1$, accept the new tree as the current tree.
5. If $R < 1$, draw a uniform random number (a random fraction between 0 and 1). If it is less than R, accept the new tree as the current tree.
6. Otherwise reject the new tree and continue with tree T_i as the current tree.
7. Return to step 2.

Two programs demonstrating MCMC sampling

Two excellent programs exist to demonstrate how MCMC finds peaks in a two-dimensional space. The user can place the peaks and can run both regular and “heated” chains to explore the space.

- Paul Lewis’s Windows program `MCRobot` which is available at http://www.eeb.uconn.edu/people/plewis/software.php
- John Huelsenbeck’s similar Mac OS X program `McmcApp` (also called `iMCMC`) which is available at http://cteg.berkeley.edu/software.html
  (hint: to place a peak on the run window, before running move the cursor on the window while holding down the mouse button. The peak is much wider than the initial ellipse, so keep that smallish).

We will demonstrate one of these in class.

Bayesian MCMC

We try to achieve the posterior

$$\text{Prob (T) Prob (D | T) / (denominator)}$$

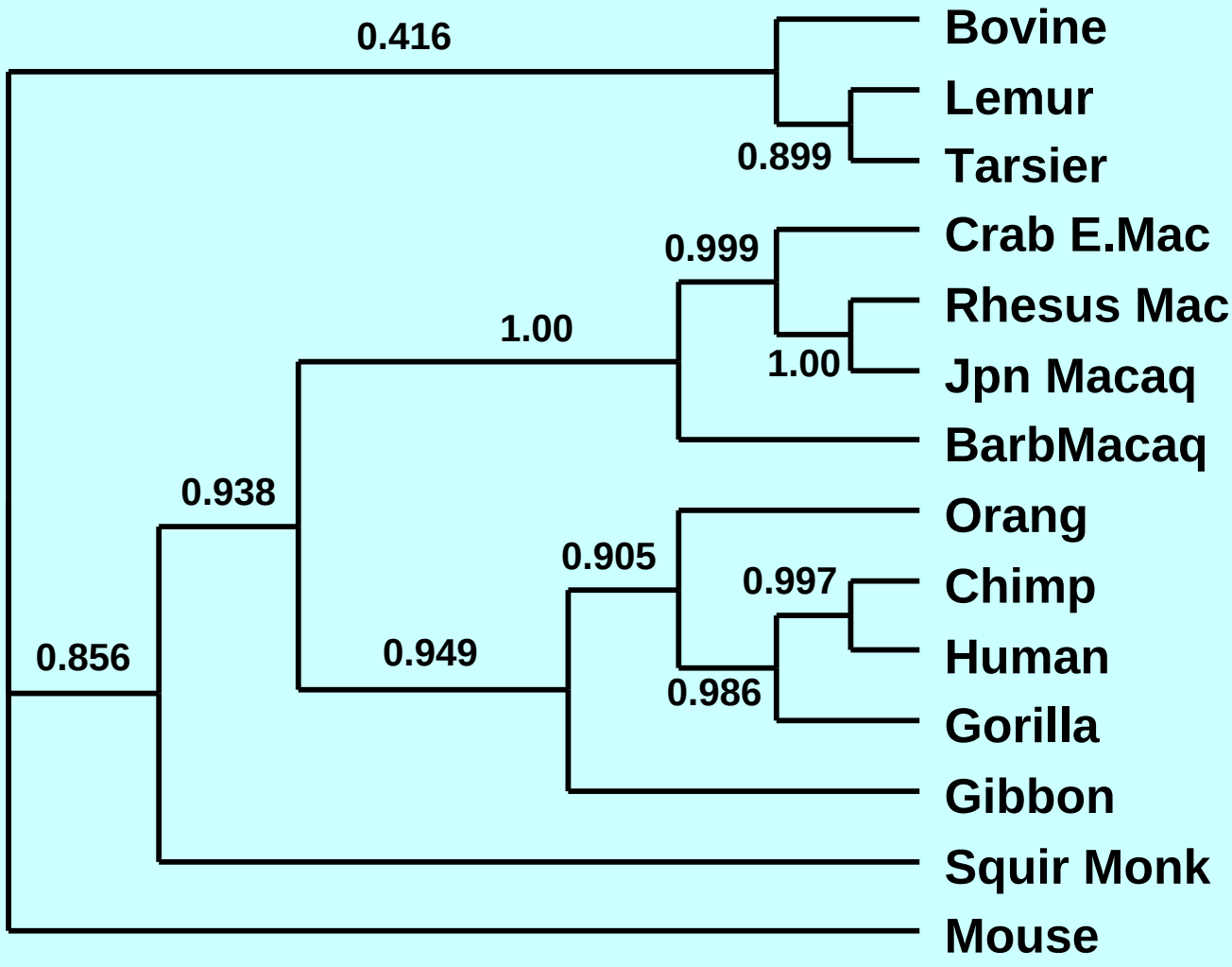
and this turns out to need the acceptance ratio

$$R = \frac{\text{Prob}\,(T_{\text{new}})\ \text{Prob}\,(D \mid T_{\text{new}})}{\text{Prob}\,(T_{\text{old}})\ \text{Prob}\,(D \mid T_{\text{old}})}$$

(the denominators are the same and cancel out. This is a great convenience, as we often cannot evaluate the deonominator, but we can usually evaluate the numerators).

Note that we could also have a prior on model parameters too, and as we move through tree space we could also be moving through parameter space.

Using MrBayes on the primates data



Frequencies of partitions (posterior clade probabilities)

Issues to think about with Bayesian inference

- Where do you get your prior from?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?
  - Or is the tree drawn from a birth-death process? If so, what are the rates of birth and death?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?
  - Or is the tree drawn from a birth-death process? If so, what are the rates of birth and death?
  - Or is there also a stage where the species studied are chosen are selected from all extant species of the group? How do you model that? Are you modelling biologists' decision-making processes?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?
  - Or is the tree drawn from a birth-death process? If so, what are the rates of birth and death?
  - Or is there also a stage where the species studied are chosen are selected from all extant species of the group? How do you model that? Are you modelling biologists' decision-making processes?
- Is your prior the same as your reader's prior?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?
  - Or is the tree drawn from a birth-death process? If so, what are the rates of birth and death?
  - Or is there also a stage where the species studied are chosen are selected from all extant species of the group? How do you model that? Are you modelling biologists' decision-making processes?
- Is your prior the same as your reader's prior?
  - If not, what do you do?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?
  - Or is the tree drawn from a birth-death process? If so, what are the rates of birth and death?
  - Or is there also a stage where the species studied are chosen are selected from all extant species of the group? How do you model that? Are you modelling biologists' decision-making processes?
- Is your prior the same as your reader's prior?
  - If not, what do you do?
  - Use several priors?

Issues to think about with Bayesian inference

- Where do you get your prior from?
  - Are you assuming each branch has a length drawn independently from a distribution? How wide a distribution?
  - Or is the tree drawn from a birth-death process? If so, what are the rates of birth and death?
  - Or is there also a stage where the species studied are chosen are selected from all extant species of the group? How do you model that? Are you modelling biologists' decision-making processes?
- Is your prior the same as your reader's prior?
  - If not, what do you do?
  - Use several priors?
  - Just give the reader the likelihood surface and let them provide their own prior?

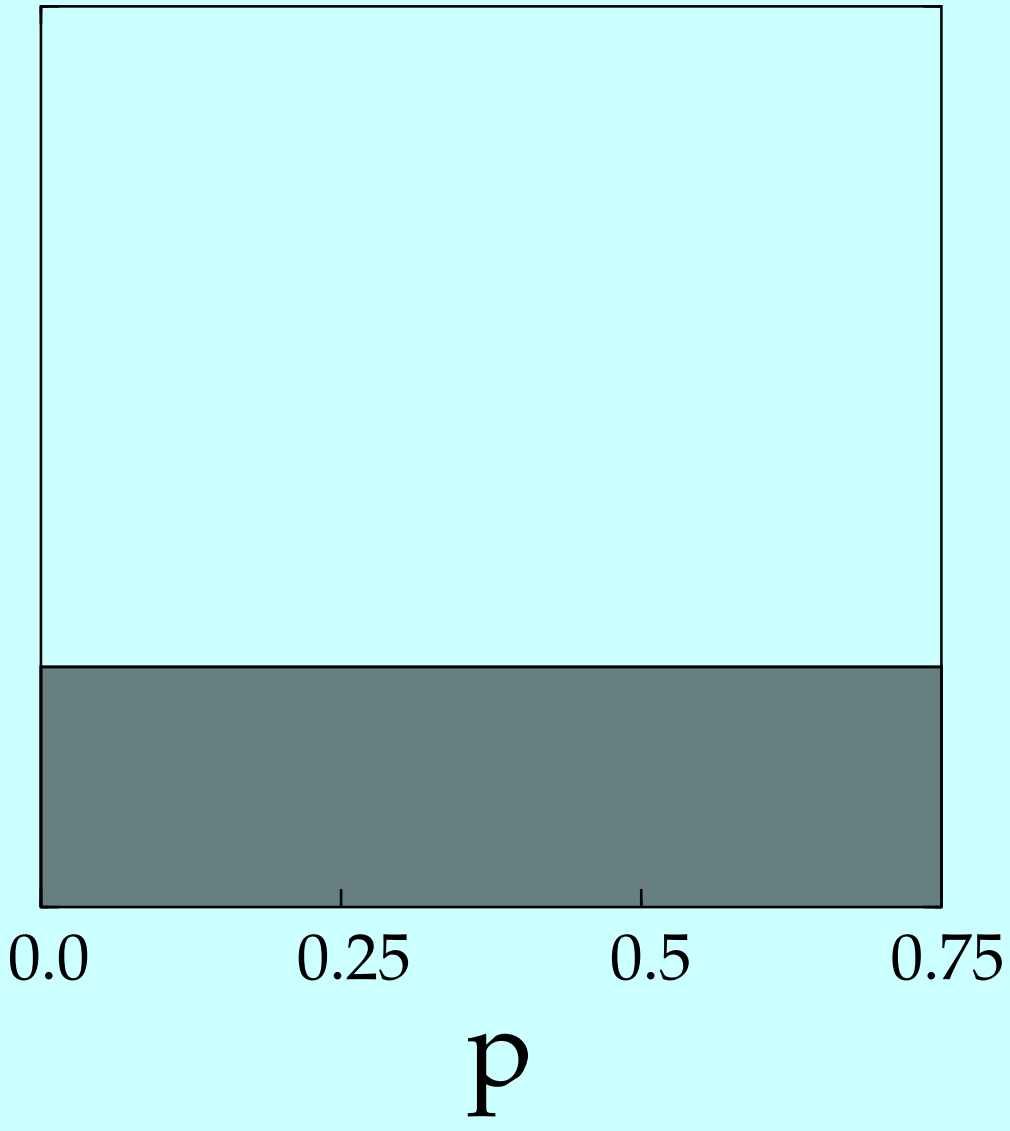
An example

Suppose we have two species with a Jukes-Cantor model, so that the estimation of the (unrooted) tree is simply the estimation of the branch length between the two species.

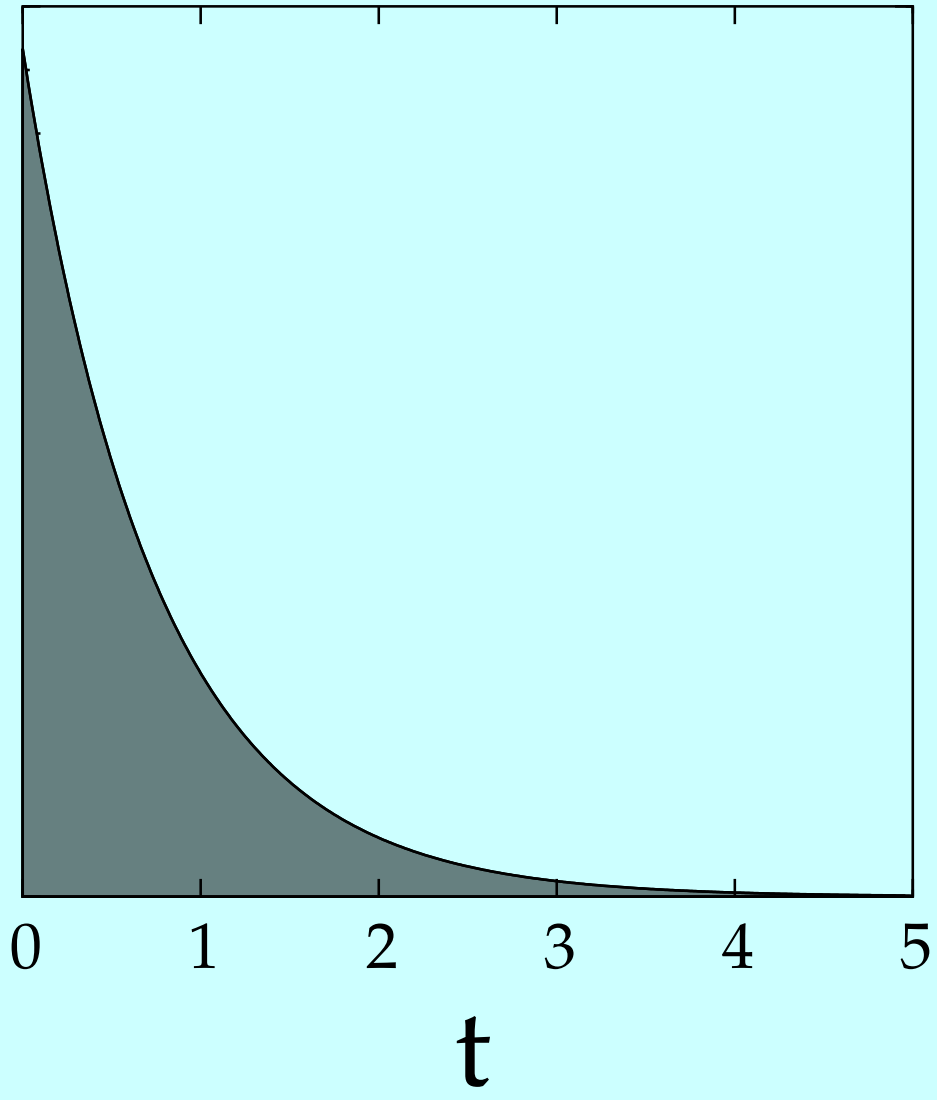
We can express the result either as branch length t, or as the net probability of base change

$$p = \frac{3}{4}\left(1 - \exp(-\frac{4}{3}t)\right)$$

A flat prior on p

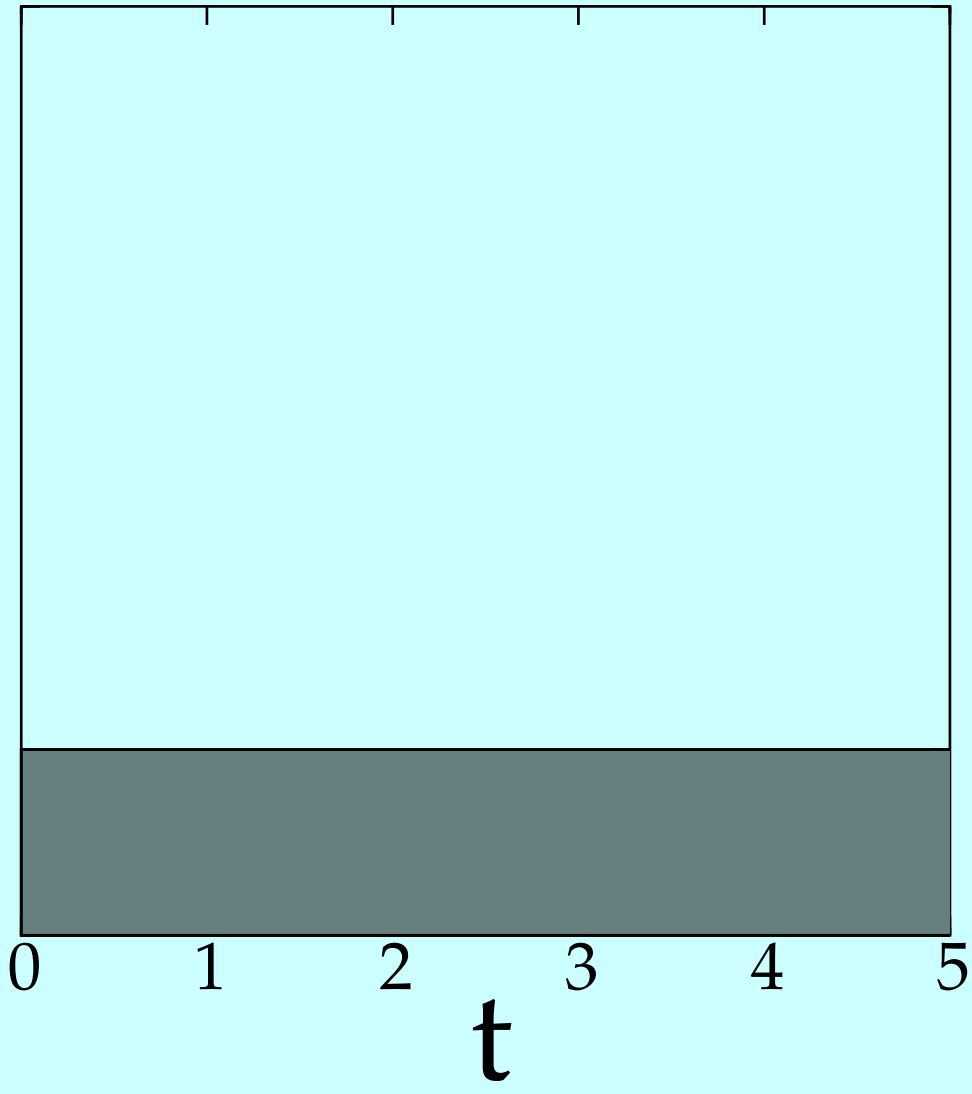


The corresponding prior on t

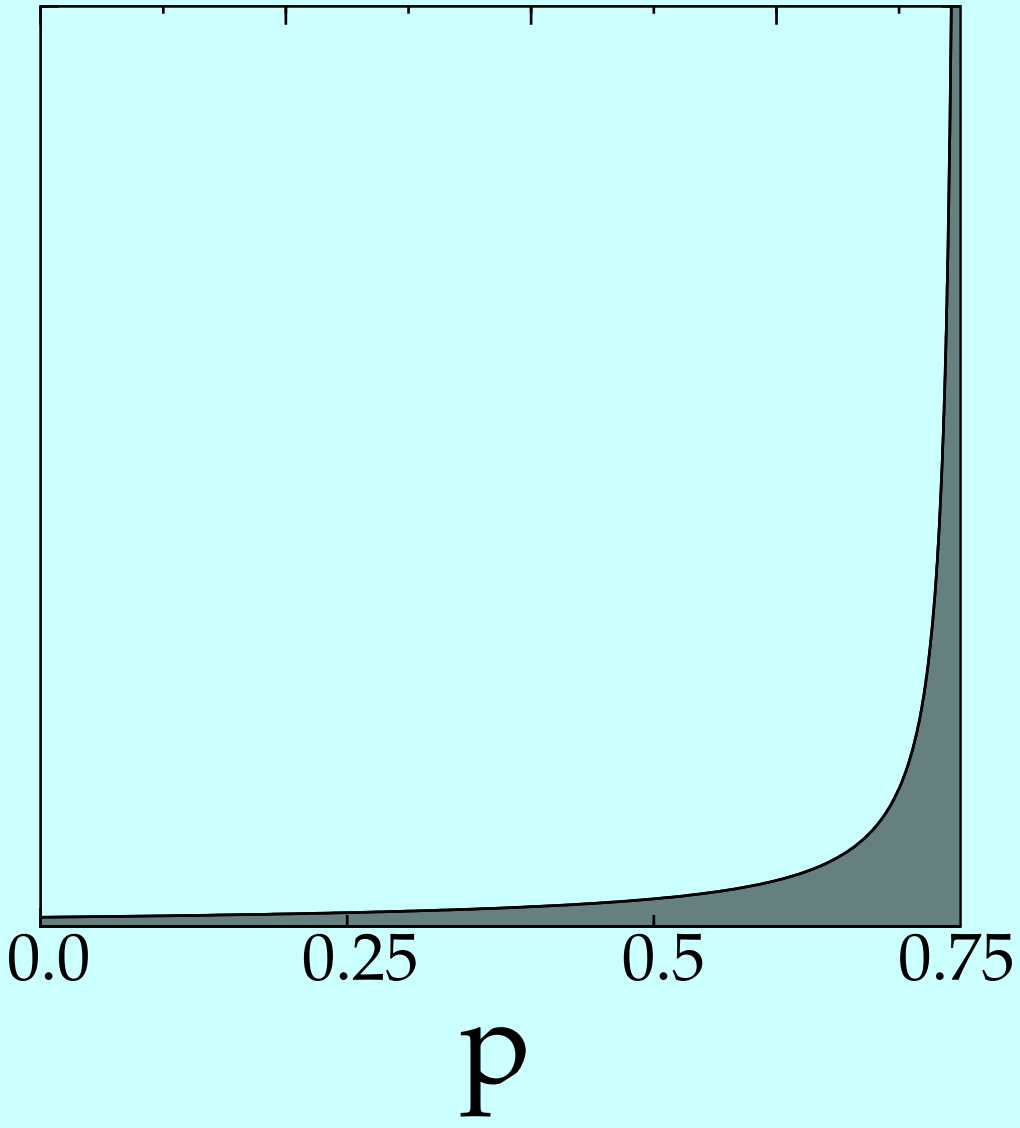


So which is the “flat prior”?

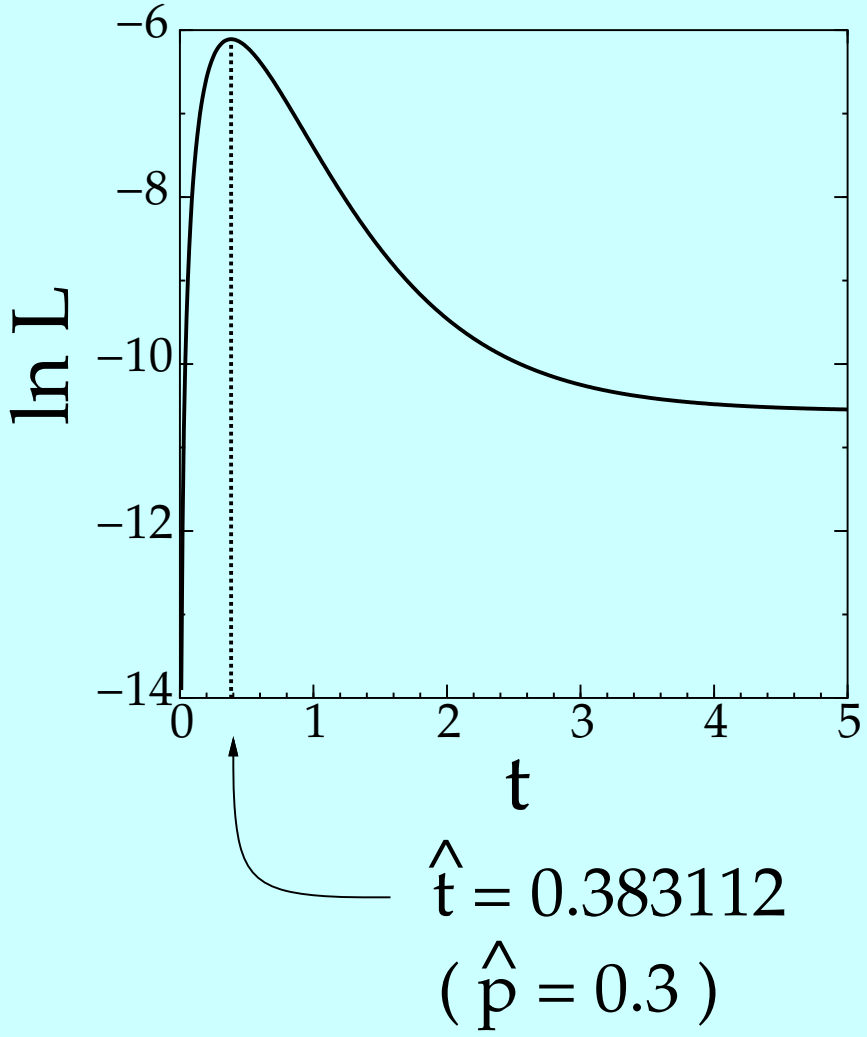
Flat prior for t between 0 and 5



The corresponding prior on p

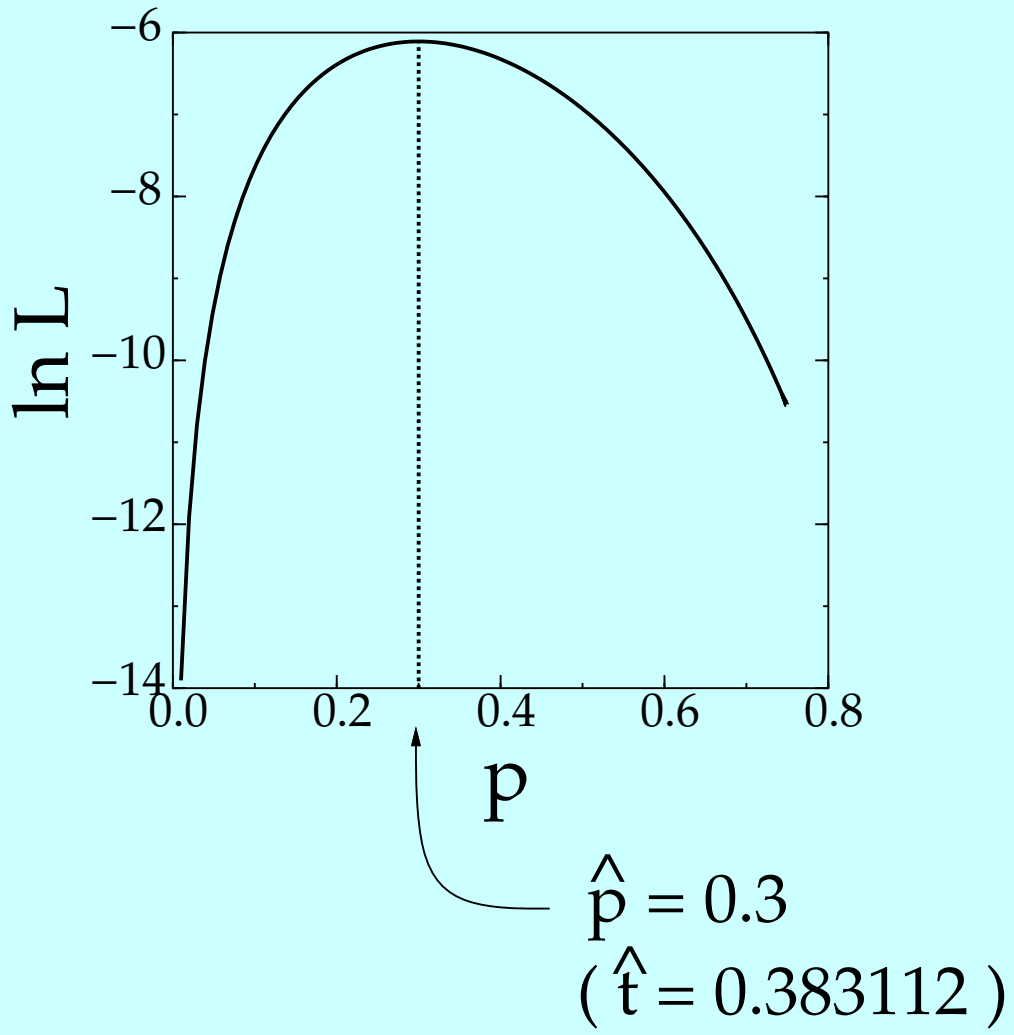


The invariance of the ML estimator to scale change



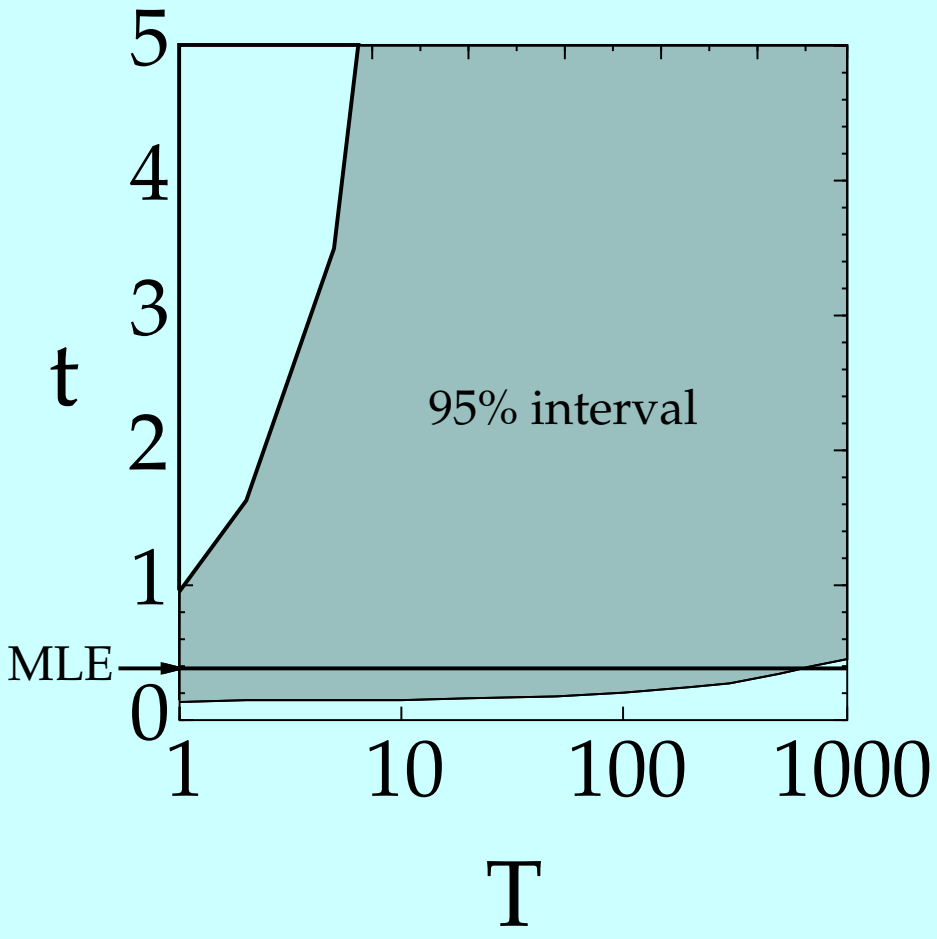
Likelihood curve for t
when 3 sites differ out of 10

The invariance of the ML estimator to scale change



Likelihood curve for p
when 3 sites differ out of 10

When t has a wide flat prior



The 95% two-tailed credible interval for t with various truncation points on a flat prior for t

What is going on in that case is ...