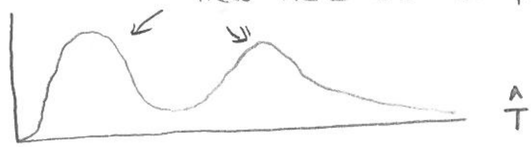
Bayesian Methods in Phylogenetics

In contrast to $\hat{T}_{MLE}$ which gives a POINT ESTIMATE for the "best" tree relating sequence data, a Bayesian analysis will give a

POSTERIOR DISTRIBUTION

fictionalized posterior

trees here are most probable



Area corresponds to probability

There are many, many considerations for constructing a posterior distribution (size of tree space, for fixed T distributions on numerical parameters, etc.). Indeed, we have a whole graduate cource in Bayesian Statistics.

Fundamental ingredients in both ML and Bayesian analyses are the likelihood function $\ell(\theta) = \ell(\theta \mid \text{data})$ and the data. [A Bayesian analysis also require a PRIOR - details forthcoming.] The key equation is derived from the simple fact relating joint and conditional probabilities

$$P(A,B) = P(A \mid B)\,P(B) = P(B \mid A)\,P(A)$$

This leads to

Bayes' Rule: $P(A \mid B) = \dfrac{P(B \mid A)\,P(A)}{P(B)}$ Graphic

↑

1701-1761

Mr. Bayes, *BEAST

Let A = parameters Θ (T, N) under some model

B = data

$$P(\theta \mid \text{data}) = \frac{P(\text{data} \mid \theta)\, P(\theta)}{P(\text{data})}$$

OR

Likelihood

$$P(\theta \mid \text{data}) = \frac{\ell(\theta)\, P(\theta)}{P(\text{data})}$$

Prior on parameters Θ

Posterior dist. of Θ

In a Bayesian framework, we must think of data as somehow generated

at random. In general computing this is HARD.

The purpose of the prior $P(\Theta)$ is to probabilistically quantify our

beliefs in the "true" values of parameters Θ [Eg. Each tree shape is equally probable; only trees which group H-C-G will have positive probability, etc.]

It is called a "prior" since we set its distribution before collecting data.

Controversial?

The numerator $\ell(\theta \mid \text{data})\, P(\theta)$ is a sort of weighted $P(\Theta, \text{data})$

which incorporates our prior beliefs about Θ.

Using Bayes Rule, we obtain the Posterior Distribution of

Θ: $P(\theta \mid \text{data})$ which takes into account both the data

and our prior beliefs about Θ.

Since the posterior assigns support to all values of Θ, it should be clear that it is hard to compute, but somehow reflects our post-data collection view of what Θ's distribution should look like.

Example: From book.

Suppose you flip a coin 3 times with $p = P(H)$ $[= \theta]$ unknown.

data: HHT then $\hat{p}_{MLE} = \frac{2}{3}$ is the point estimate for p

For a Bayesian analysis, we want

$$P(p \mid HHT) = \frac{\ell(\theta \mid HHT)\, P(p)}{P(\text{data})}$$

The likelihood is $\ell(\theta \mid HHT) = p^2(1-p)$ under the iid. assumption.

The prior $P(p)$, chosen by the user, is say $P(p) = \begin{cases} 1/3 & \text{if } p = .25, .5, .75 \\ 0 & \text{all other } p \end{cases}$

To compute $P(\text{data})$ we use the prior and the Law of Total Probability

$$P(\text{data}) = P(HHT) = P(HHT \mid p=.25)\,P(p=.25) + P(HHT \mid p=.5)\,P(p=.25)$$

$$+ P(HHT \mid p=.75)\,P(p=.75) \qquad + 0$$

prior assign no mass to any other values of p

$$= \left(\tfrac{1}{4}\right)^2\left(\tfrac{3}{4}\right)\left[\tfrac{1}{3}\right] + \left(\tfrac{1}{2}\right)^2\left(\tfrac{1}{2}\right)\left[\tfrac{1}{3}\right] + \left(\tfrac{3}{4}\right)^2\left(\tfrac{1}{4}\right)\left[\tfrac{1}{3}\right]$$

$$= \frac{20}{192} = \frac{5}{48} \approx .1042$$

The posterior distribution

$P(p \mid HHT)$ has support (non-zero values) only at $p = .25, .5, .75$

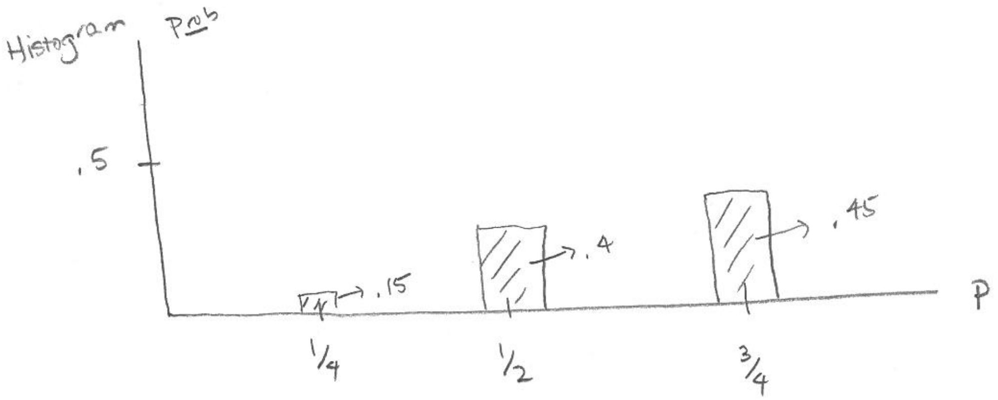
Thus,

$$P(p=.25 \mid HHT) \approx \frac{\ell(p=.25 \mid HHT)\, P(p=.25)}{.1042} = \frac{(\frac{1}{4})^2(\frac{3}{4})\left[\frac{1}{3}\right]}{.1042} \approx .15$$

$$P(p=.5 \mid HHT) \approx \frac{\ell(p=.5 \mid HHT)\, P(p=.5)}{.1042} = \frac{(\frac{1}{2})^2(\frac{1}{2})\left[\frac{1}{3}\right]}{.1042} \approx .40$$

$$P(p=.75 \mid HHT) \approx \frac{\ell(p=.75 \mid HHT)\, P(p=.75)}{.1042} = \frac{(\frac{3}{4})^2(\frac{3}{4})\left[\frac{1}{3}\right]}{.1042} \approx .45$$

Graphically,



Updated posterior dist of $p = p(H)$.

Redo: Using a <u>flat</u> continuous prior $P(p) = \begin{cases} 1 & \text{for all } p \in [0,1] \\ 0 & \text{o.w} \end{cases}$

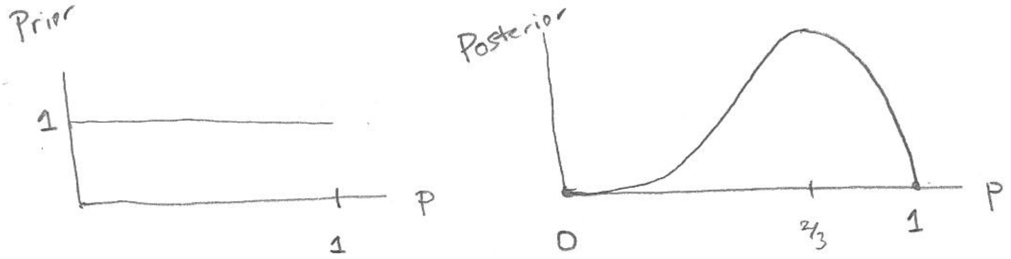
The only changes are that "sums" become "integrals."

$$P(\text{data}) = \int_0^1 \underset{\substack{\uparrow \\ 2H}}{p^2} \underset{\substack{\uparrow \\ 1T}}{(1-p)} \left[\underbrace{P(p)\,dp}_{} \right] = \int_0^1 p^2 - p^3\, dp = \frac{1}{3} - \frac{1}{4} = \underline{\underline{\frac{1}{12}}}$$

prior density weight the possible values of p

The posterior $P(p \mid HHT)$ is a continuous density function

$$P(p \mid HHT) = \frac{\ell(p \mid HHT)\, P(p)}{1/12} = \frac{p^2(1-p)[1]}{1/12} = 12p^2(1-p) \qquad 0 \le p \le 1$$



i.e. lots of probability that p is close to $2/3$.

Incorporate data and prior belief