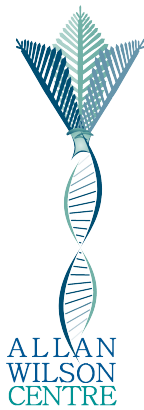


## Some mathematical aspects of Bayesian phylogenetics



Mike Steel

Allan Wilson Centre for  
Molecular Ecology and Evolution

Biomathematics Research Centre  
University of Canterbury,  
Christchurch, New Zealand



Budapest, June 28, 2008

1

## Overview

- Priors on trees
- Decision-theoretic view
- Models of sequence evolution
- The Bayesian 'star paradox'

2

## Part 1: Priors on binary trees\*

- $T_B$  - tree on labelled leaf set  $B$   $\sum_{T_B} P(T_B) = 1$

- Desirable properties:

- Exchangeable  $P(T_B^\sigma) = P(T_B)$
- Markovian

$$t_1 \perp_{(B_1, B_2)} t_2 \quad t_1 \approx T_{B_1}, t_2 \approx T_{B_2}$$

$$\Rightarrow P(T = t) = \prod_{v \in V_{\text{int}}} p(\#v_1, \#v_2)$$

- Consistent  $A \subset B \Rightarrow T_B \mid A \approx T_A$

$\Rightarrow p(i, j)$  satisfy system of equations

\*From 'Gibbs fragmentation trees', McCullagh, Pitman, Winkel, 2008 <sup>3</sup>

## Priors on trees: Aldous' beta-splitting model

$$P(T = t) = \prod_{v \in V_{\text{int}}} p(\#v_1, \#v_2)$$

$$p(i, j) = \frac{1}{Z} \int_0^1 x^{i+\beta} (1-x)^{j+\beta} dx$$

$\beta = \frac{-3}{2}$ , Uniform (PDA) distribution on binary trees

$\beta = 0$ , Yule model (=coalescent)

$\beta = -1$ , Intermediate balance (good fit to some data)

4

## What's special about the beta-splitting model

$$p(i,j) = \frac{1}{Z} \int_0^1 x^{i+\beta} (1-x)^{j+\beta} dx \quad P(T=t) = \prod_{v \in V_{\text{int}}} p(\#v_1, \#v_2)$$

$$p(i,j) \propto w(i)w(j)$$

- **Theorem** [McCullagh, Pitman, Winkel, 2008].

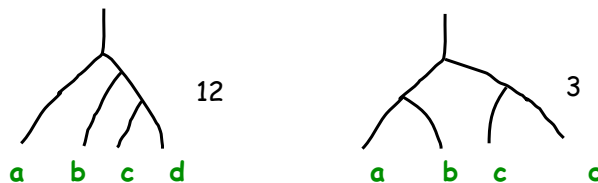
If a prior on binary trees is exchangeable, Markovian, consistent, and with factorizing  $p$ , then it is the Aldous  $\beta$ -splitting model for some  $\beta > -2$ .

5

## Strange bayes, indeed?

Pickett, K.M. and Randle, C.P., 2005. Strange bayes indeed: uniform topological priors imply non-uniform clade priors. *Mol. Phyl. Evol.* 34, 203-211.

- Clade size... a uniform prior on trees gives a uniform prior on clades (regardless of size)?



$$P(\text{bcd clade}) = 3/15 + 0/15 = 1/5$$

$$P(\text{cd clade}) = 2/15 + 1/15 = 1/5$$

- **Theorem:** For  $n > 4$ , there is no prior on trees is uniform on clade size.

On the impossibility of uniform priors on clade size  
Mike Steel and Kurt M. Pickett (MPE,2006)

Why non-uniform priors on clades are both unavoidable and unobjectionable

Joel D. Velasco (MPE,2007)

6

## Part 2: A decision-theoretic perspective (I)

$$T^* \rightarrow X \xrightarrow{\varphi} T$$

Example:  $\varphi = \varphi_{MP}$

$L(T^*, T)$  'loss function'

Expected posterior loss of a tree  $T$  (given  $X$ )

$$\rho(T) = \sum_{T^*} L(T^*, T) P(T^* | X)$$

7

## A decision-theoretic perspective (II)

- 0-1 loss

$$L_{0/1}(T^*, T) = \begin{cases} 0 & \text{if } T^*=T, \\ 1, & \text{if } T^* \text{ is different to } T. \end{cases}$$

- **Trivial observation:** The tree  $T$  that minimizes the EPL for  $L_{0/1}$  is the max posterior probability tree

$$\rho(T) = \sum_{T^*} L(T^*, T) P(T^* | X)$$

8

## A decision-theoretic perspective (II)

$$\text{Branch loss } L_\alpha(T^*, T) = \alpha \cdot |T - T^*| + (1 - \alpha) \cdot |T^* - T|$$
$$\alpha > \frac{1}{2}$$

**Question:** What tree minimizes the expected posterior loss for  $L_\alpha$ ?

Posterior probability of split  $\sigma$   $P(\sigma | X) = \sum_{T^*} P(T^* | X)$

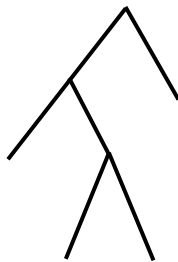
■ **Theorem** [M. Holder\*, 2008].

The tree  $T$  that minimizes the EPL for  $L_\alpha$  is precisely the tree consisting of those splits  $\sigma$  with  $P(\sigma | X) \geq \alpha$

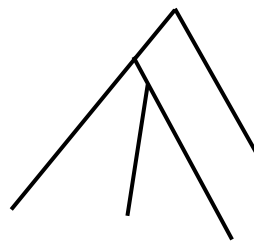
\*A justification for reporting majority-rule consensus trees in Bayesian phylogenetics, Syst. Biol. (in press).

## Part 3: Models of sequence evolution

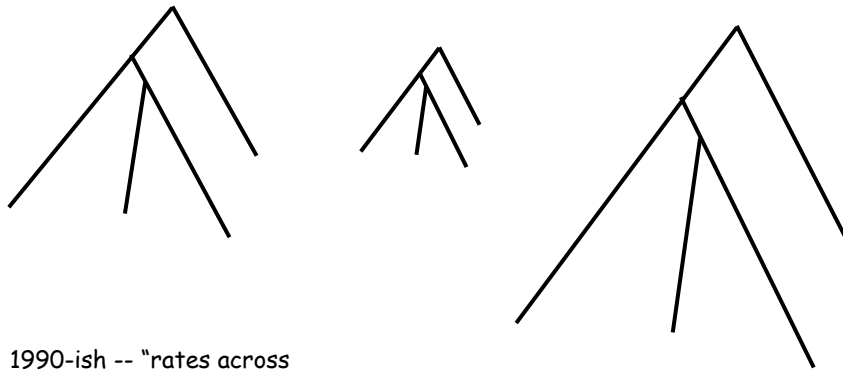
'spherical cows'



'elliptical cows'



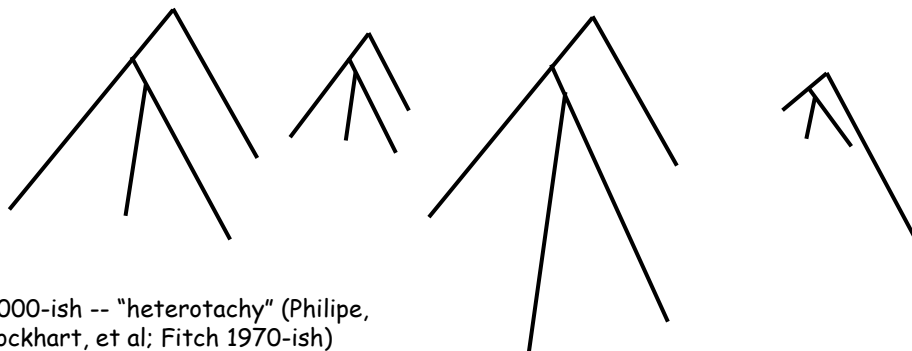
## Cows come in different sizes...



1990-ish -- "rates across sites" (Yang et al.)

11

## Some cows have horns...



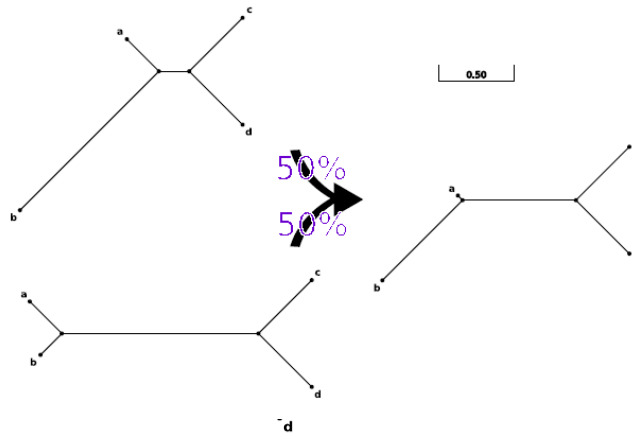
2000-ish -- "heterotachy" (Philippe, Lockhart, et al; Fitch 1970-ish)

... others are just plain weird

12

## Phylogenetic Mixtures

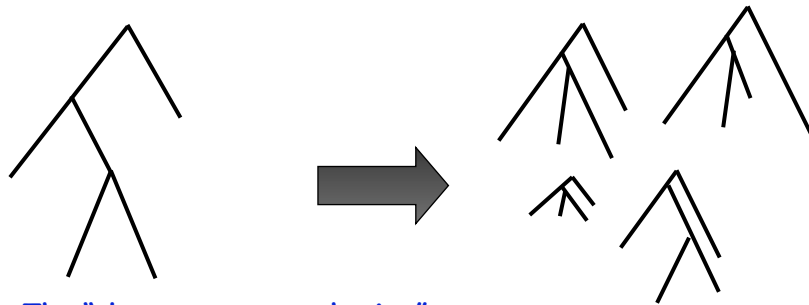
- [Matsen, E. and S, 2007. Phylogenetic mixtures on a single tree can mimic a tree of another topology. *Systematic Biology* 56(5): 767--775. ]



Two elliptical cows = elliptical cow (same or different topology)

13

## From spherical cows to... 'real' cows



UCM: The "ultra-common mechanism" model

NCM: The "no-common mechanism" model

14

## How to restore 'statistical respectability' to a NCM approach?

### Two recent approaches

- Huelsenbeck, Ane, Larget, Ronquist

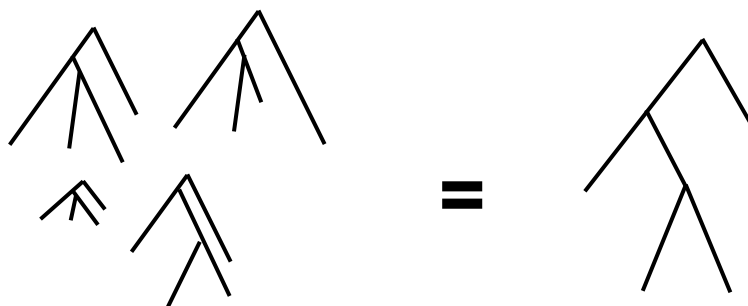
"A Bayesian perspective on a non-parsimonious parsimony model." *Syst. Biol.* 56(3): 406–419, 2008.

- Wu, Susko and Roger

"An independent heterotachy model and its implications for phylogeny and divergence time estimation." *Mol. Phyl. Evol.* ;46(2):801–806, 2008.

15

## From 'real' cows, to....



**BNCM:** The Bayesian "no-common mechanism" model (Neyman-model)

**UCM:** The "ultra-common mechanism" model

16



- Properties of expectation (mean)

$$E[X + Y] = E[X] + E[Y]$$

$$X \perp Y \Rightarrow E[X \cdot Y] = E[X] \cdot E[Y]$$

17

## Why?

$$P(f) = \int_{\Gamma} \sum_{F:F|L=f} P(F | \gamma) \phi(\gamma) d\gamma$$

$$= E_{\phi} \left[ \sum P(F | \gamma) \right] = \sum E_{\phi} [P(F | \gamma)]$$

$$= \sum E \left[ \prod_{e \in C} p_e(\gamma_e) \cdot \prod_{e \notin C} q_e(\gamma_e) \right]$$

$$= \sum \prod_{e \in C} p'_e \cdot \prod_{e \notin C} q'_e$$

18

## Wu *et al.* model

Branches have fixed lengths, but each has an iid variable site-specific rate

for JC, same as ER (elliptical cow) model

for other rate matrices  $R$ , eigenvalues of  $R$  matter

logdet is additive on this model.

19

## Moral of story so far?



**Pessimist:** Models of DNA evolution that ignore site position can't be too different to UCM

**Optimist:** Unrealistic models can have more realistic interpretations

...and these more 'realistic(?)' models allow T to be identified...

20

## Part 4: The Bayesian Star Paradox



### Some general comments

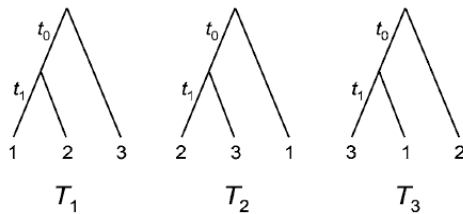
Not peculiar to phylogeny

Not peculiar to Bayesian

The phenomenon is real  
(*contra*. Kolaczkowski and Thornton)

But... its significance for phylogenetics is debatable.

21



$$P[T_1 | \text{Data}] = c \cdot \int_0^\infty \int_0^\infty P[\text{Data} | T_1, t_0, t_1] f(t_0, t_1) dt_0, dt_1$$

"Data" =  $(n_0, n_1, n_2, n_3)$ ,  $n_0 + n_1 + n_2 + n_3 = n$  (sequence length)

$$P_1 = P[T_1 | \text{Data}]$$

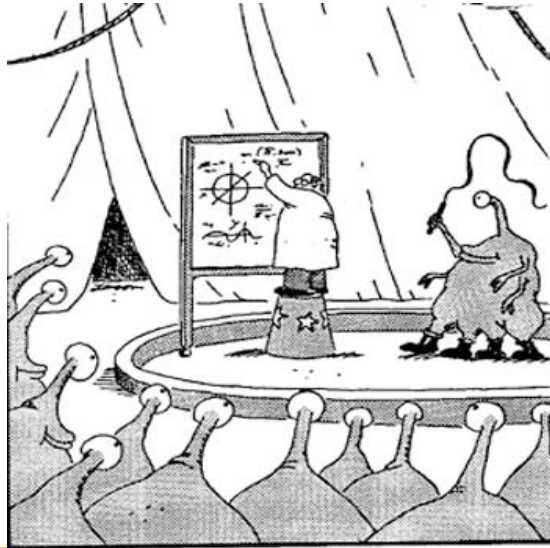
$P_1$  is a random variable.

• As  $n$  becomes large we might expect  $P_1$  to  $\rightarrow 1/3$

• What's distribution of  $P_1$  or  $(P_1, P_2, P_3)$  for  $n$  large?

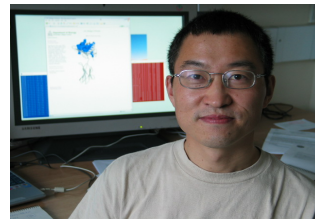
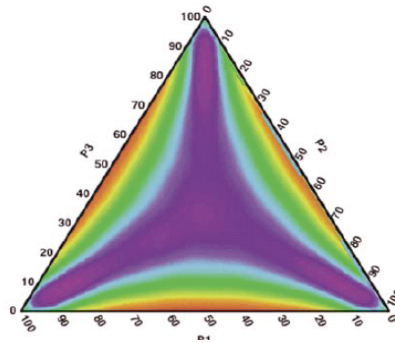
22

■ Brute force calculation?



Abducted by an alien circus company, Professor Winters is forced to write Linear Algebra equations in center ring. 23

(c)  $n = 1000$



Yang and Rannala, MBE

“We encountered problems with numerical integration using *Mathematica* for large  $n$ , and it is unclear what the limiting distribution  $f(P_1, P_2, P_3)$  is when  $n \rightarrow \infty$  .”

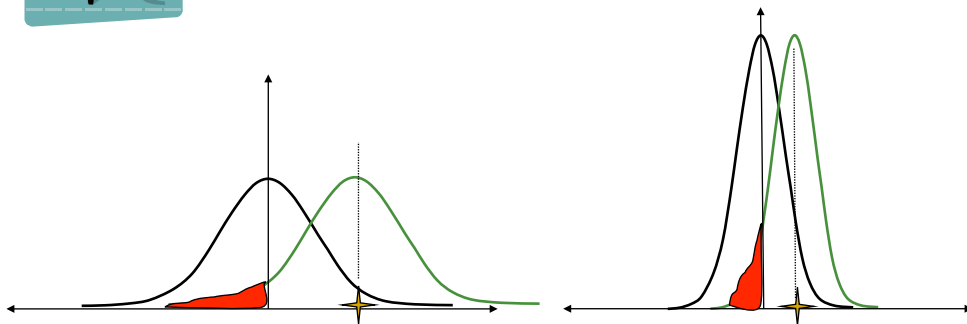
## Simpler problem: coin tossing (Bootstrap)



$H_1: p > \frac{1}{2}$

$H_2: p < \frac{1}{2}$

'Truth':  $H^* p = \frac{1}{2}$

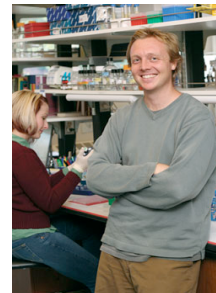


*Mol. Biol. Evol.* 23(10):1819–1823. 2006

### LETTERS

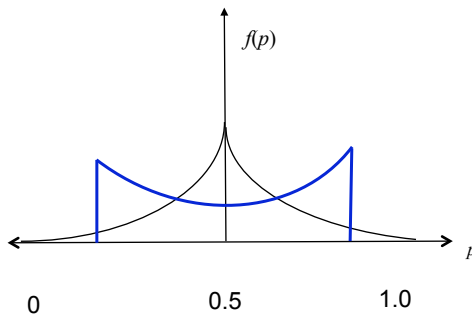
#### Is There a Star Tree Paradox?

Bryan Kolaczkowski\* and Joseph W. Thornton†



"...With infinite data, posterior probabilities give equal support for all resolved trees, and the rate of false inferences falls to zero. **We conclude that there is no star tree paradox caused by not sampling unresolved trees.**"

## Simpler problem: coin tossing (Bayesian)



$H_1: p > 1/2$

$H_2: p < 1/2$

'Truth':  $H^* p = 1/2$

'Tame' prior

27

## Chicken scratchings...



$$P(H_1 | D) = \frac{P(D | H_1)P(H_1)}{P(D)} = (1 + Y^{-1})^{-1}$$

$$Y = \frac{P(D | H_1)}{P(D | H_2)}$$

Show: Under  $H^*$  the probability that  $D$  is such that  $Y$  is large ( $>K$ ) tends to  $e^{-K} > 0$ , as  $n$  grows.

$$Y = \frac{E[(1+X)^h(1-X)^l]}{E[(1-X)^h(1+X)^l]}$$

$$h = n \frac{1}{2} + C\sqrt{n}$$

$$Y = \frac{E[(1-X^2)^{n/2} Z^{C\sqrt{n}}]}{E[(1-X^2)^{n/2} Z^{-C\sqrt{n}}]}$$

$$Z = \frac{1+X}{1-X}$$

28

## Chicken scratchings 2:



$$Y = \frac{E[(1-X^2)^{n/2} Z^{c\sqrt{n}}]}{E[(1-X^2)^{n/2} Z^{-c\sqrt{n}}]} \quad Z = \frac{1+X}{1-X}$$

Let  $W := 1 - X^2$

Under  $H^*$ ,  $C \geq c$ , w.p.  $d = d(c) > 0$ .

$$Y \geq \frac{E[W^{n/2} c^2 n (1-W)]}{E[W^{n/2}]} = 2c^2 \left( k \frac{E[W^k] - E[W^{k+1}]}{E[W^k]} \right) \rightarrow f(c)$$

$k = n/2$

29

## Exercise for the 'mathmos'

- Show that, under  $H^*$  and any 'tame' prior that the distribution of  $P(H_1|D)$  converges to uniform  $[0,1]$  as  $n \rightarrow \infty$ .

30

## What about trees?

### Theorem

Consider sequences of length  $n$  generated by a star tree  $T_0$  with (positive, finite) edge length  $t_1$ .

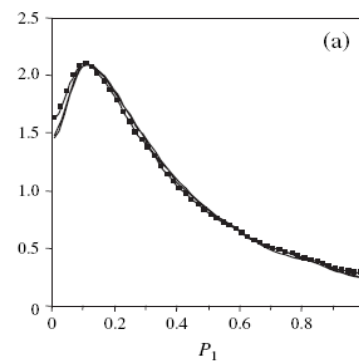
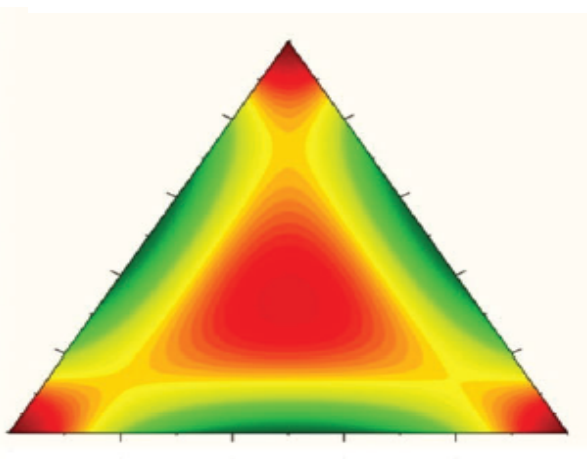
For any 'tame' prior on the three resolved trees and their branch lengths, and any  $\varepsilon > 0$  the following holds: The probability that the Data has the property that

$$P(T_1 | \text{Data}) > 1 - \varepsilon,$$

does not converge to 0 as  $n$  tends to infinity.

31

Ziheng Yang, Fair-Balance Paradox, Star-tree Paradox, and Bayesian Phylogenetics  
*Molecular Biology and Evolution* 2007 **24(8)**:1639-1655



Not uniform!

Ed Susko – forthcoming paper describes  $f(P_1)$  and extension to larger trees with different branch lengths

32



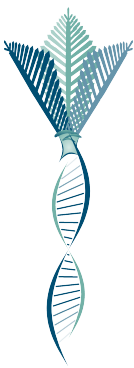
## Remedies

Allow  $> 0$  prior probability to unresolved trees

Let prior on branch lengths depend on  $n$   
(sequence length)  
--i.e. shrink to zero at correct rate.

33

## Thanks to....



- Alexander von-Humboldt Stiftung
- NZ Marsden Fund

- Reprints  
[www.math.canterbury.ac.nz](http://www.math.canterbury.ac.nz)

ALLAN  
WILSON  
CENTRE