

Can we avoid SIN in the House of 'No Common Mechanism'?

Mike Steel



Portland, June 26th, 2010

1

Original SIN



$$\text{Set } p(i) = \begin{cases} 0.5 - \alpha_i & \text{if } i \nabla S \\ 0 & \text{if } i \Delta S \end{cases}$$

for $\alpha_i > 0$ and denote the associated family of transition matrices $E_i(\alpha_i)$. Then, as $\alpha_i \rightarrow 0^+$

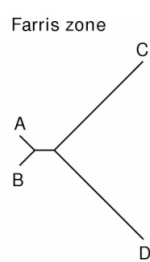
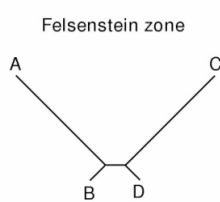
$$P(\alpha_i) | T, F_i(B) \rightarrow 2^{-i}$$

thus establishing (i). Combining this with (3) gives (2). Now, from equation (1), we have

$$\begin{aligned} \sup \{L\} &= \sum_{i=1}^t \sup_{\alpha_i} \{P(\alpha_i) | T, F_i\} \\ &= \sum_{i=1}^t (2^{-i} \alpha_i) \\ &= 2 \sum_{i=1}^t \alpha_i \end{aligned}$$

and so we know the tree(s) which maximizes L is the tree(s) which minimizes

$$\sum_{i=1}^t \alpha_i$$



a

b

Parsimony and explanatory power

James S. Farris*

...23 pages later:

“Many, no doubt, would consider such cases depressing, signs of waning integrity among scientists.

But there is a silver lining.

If these are the strongest criticisms that my view of parsimony has to face, then it has a bright future indeed.”

Models



- Parameters (t, β)
 t in T (a finite set), β in $B(t)$ = open subset of Euclidean space
 - Model: $(t, \beta) \mapsto P_{(t, \beta)}$ (=probability distribution on finite set S)
 - N_r -model: $r = 2, 3, 4, \dots, \infty$
-
- ↑
↑
- Jukes-Cantor
Kimura-Crow infinite allele model

Why is tree reconstruction fraught with peril?

- “Kissing” condition:

$$\overline{p(B(t))} \cap \overline{p(B(t'))} \neq \emptyset$$

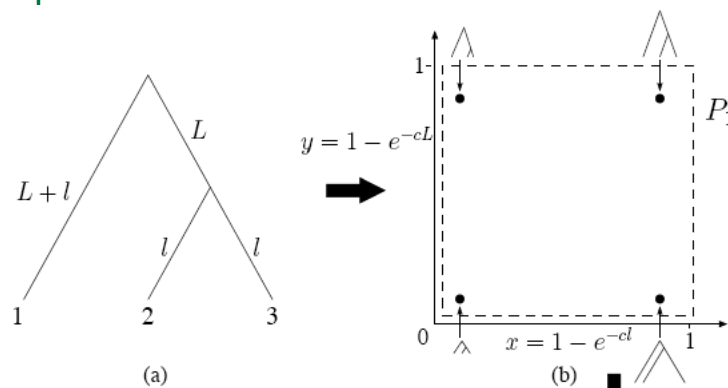
but still possible?

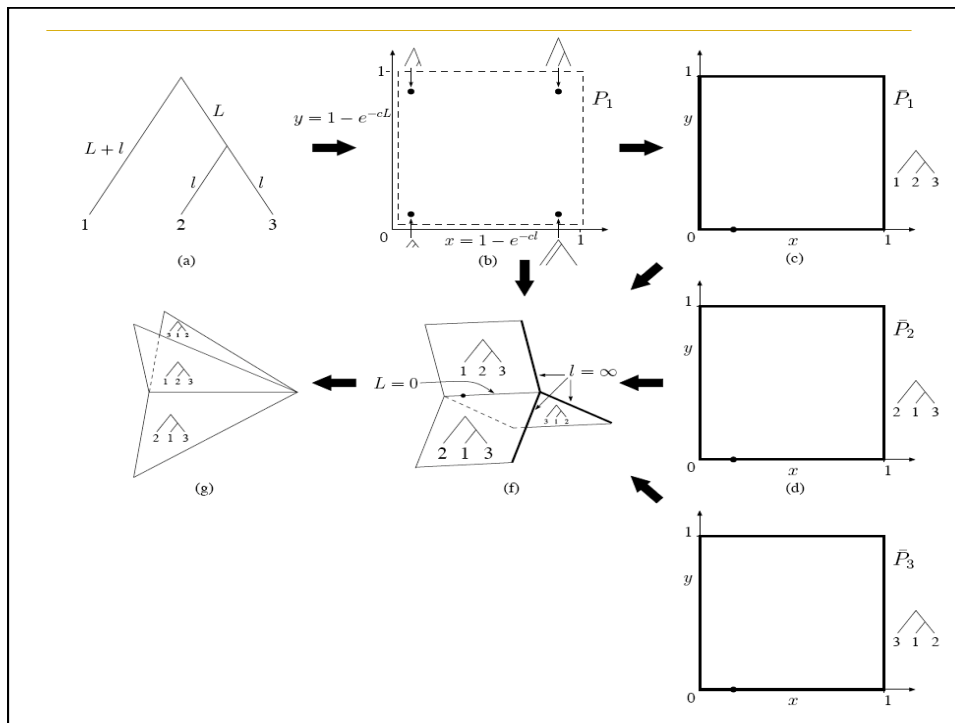
- “No touching” condition:

$$p(B(t)) \cap \overline{p(B(t'))} = \emptyset$$

$$p(B(t)) = \{p(t,b) : b \in B(t)\}$$

Example: 3-taxon molecular clock tree space





SIN in CM

$$(t, \beta) \rightarrow (u_1, \dots, u_k) \rightarrow t'$$

$$\lim_{k \rightarrow \infty} \Pr(t' = t) \rightarrow 1$$

- Violation of 'no touching' in model leads to SIN by any method
- Enforcing 'no touching' ensures MLE avoids SIN

Known violations of 'no touching':

Non-binary trees (polytomies)

Rates across sites (with too many parameters)

Phylogenetic mixtures

What is 'No Common Mechanism'?

- "This assumption [of common mechanism] can and should be removed. It is unacceptable biologically because it says, for example, that an insect species is just as likely to lose (or acquire) wings as a spot of color."

J. Cavender (1981) p.222

Isn't this what standard molecular methods do with our sequence data?

- Rates across sites (GTR+Gamma+I+.....)
- Covarion models, Mixture models, etc.

No!

these are (mostly) CM methods - sites evolve i.i.d. So what if sites evolve independently but according to different processes?

Set-up

- Data: CM-M vs NCM-M
- Max. Likelihood: CM-ML vs NCM-ML

Some questions:

- Does NCM-ML lead to SIN for NCM-M data? (or for CM-M data?)
- If so, can any method avoid SIN?

Under NCM does kissing always cause any method to SIN?

- Yes

- But only if you are very naughty...

- and even then it's not 'mortal SIN'

(Re)-defining SIN...

$$(t, \beta_i) \rightarrow u_i : (u_1, \dots, u_k) \rightarrow t'$$

- **Definition:** [Statistical consistency of a method M on NCM- model data]

For each $t \in T$, and every compact subset C of $B(t)$, the probability that M correctly estimates t from (u_1, \dots, u_k) , when each u_i is generated independently by the model with parameters (t, β_i) , where $\beta_i \in C$, converges to 1 as k grows.

Some results

- ★ *ML estimation of tree topology under the NCM- N_r model applied to NCM- N_r data leads to SIN.*
- ★ *No tree reconstruction method for NCM- N_2 data can avoid SIN*
- ★ *There is a method for inferring tree topology from NCM- N_4 (=NCM-Jukes Cantor) model that avoids SIN*
- ★ *ML estimation of tree topology under the NCM- N_∞ model of NCM- N_∞ data avoids SIN*

What does it all mean?

- ★ **In CM, if any method is consistent then ML is; but in the NCM world, this is no longer true**
- ★ **Avoiding SIN in the house of NCM requires a model that walks the 'straight and narrow'**
- ★ **A 'silver lining' for Farris?**
- ★ **Is there any hope? - it's not SIN if...**

c.f. Stefankovic and Vigoda 2007

SIN without kissing!

$$\inf\{d(p_{(t,\beta)}, p_{(t',\beta')})\} \geq q > 0.$$

- Yes, it's possible, but to learn how, and other tantalizing titbits... see [2]



- [2] Steel, M. Can we avoid 'SIN' in the House of 'No Common Mechanism'? *Syst Biol.* (in press)