The *Penny Ante*

Last updated 8 July, 2011

A prize of your choice between $100 or a bottle of single walk whisky (for medicinal purposes only). This is for the solution to any well-specified problem (mathematical or otherwise) hosted here that is announced by the end of the 2012 NZ phylogenetics conference. Nominations for suitable problems can be sent to us at any time. The *Penny Ante* is named in honour of Professor David Penny.

Combinatorial Challenge I: *Walking through trees*

Submitted by David Bryant

An "NNI-walk" is a sequence T_1, T_2, …, T_k of unrooted binary phylogenetic trees where each consecutive pair of trees differ by a single NNI.

i. [**Question**] What is the shortest NNI walk that passes through all binary trees on *n* leaves?
ii. [**Question**] Suppose we are given a tree *T*. What is the shortest NNI walk that passes through all the trees that lie at most one SPR (subtree prune and regraft) move from *T*?

Combinatorial Challenge II: *How little can two trees agree on?*

Submitted by Mike Steel

Given two unrooted binary phylogenetic trees T, T', each with leaf set *X* of size *n*, an *agreement set* for T, T' is a subset Y of X for which T|Y= T'|Y.
[**Question**] Is there a constant *c*, so that for any two trees T, T' have an agreement subtree of size $c \cdot \log n$?

Enumeration challenge I: *Can't see the trees for the … network*

Submitted by Luay Nakhleh

Given a phylogenetic $\underline{X}$-network *N* (a rooted binary DAG leaf labeled bijectively by set *X*), how many unique trees are displayed by *N*?

Comment: The number is bounded above by 2^h, here *h* is the number of nodes with two incoming arcs; moreover, for certain classes of networks (e.g. Normal networks) Stephen J. Willson has shown this upper limit is achieved (i.e. all the trees are distinct). In general, however, the number appears to be arbitrarily smaller than the upper bound for some networks.

Enumeration challenge II: *Counting nets (up to isomorphism or other equivalence).* Submitted by Luay Nakhleh

Question: What is the number of unique (up to digraph isomorphism) rooted phylogenetic networks on *n* taxa and with *h* reticulation nodes?

Comment: Related questions (based on different notions of equivalence) are:

What is the number of rooted phylogenetic networks on *n* taxa and with *h* reticulation nodes, where two networks are regarded as equivalent if they do not induce the same set of:

a. trees?
b. softwired clusters?
d. hardwired clusters?

Computational Complexity Challenge: *Perfect taxon sampling*
Submitted by Mike Steel

A collection *S* of subsets of *X* is said to be *phylogenetically decisive*[1] if for every binary phylogenetic *X*-tree *T*, the collection *T*|*Y*: *Y* \in *S* defines *T* (i.e. *T* is the only tree that displays these trees).

[**Question**] What is the computational complexity of determining whether *S* is phylogenetically decisive?

This may be related to the 'No-rainbow-colouring' problem of Manuel Bodirsky, whose complexity appears still open.

[1] Steel, M. and Sanderson, M.J. (2010). Characterizing phylogenetically decisive taxon coverage. *Applied Mathematics Letters* 23, 82-86.

Mathematical Challenge: *Does hill-climbing for ML always work for 'perfect data' fitted to the true tree?*
Posed by David Bryant, Mike Steel, Olivier Gascuel and others.

Consider a simple model of DNA substitution (e.g. Jukes-Cantor, or the symmetric 2-state model). Let *T* be a binary tree, θ_0 a selection of branch lengths and $s(T,\theta_0)$ the exact probability distribution on characters that the model would produce with these parameters. Consider the log-likelihood function $L(\theta)=\sum_i s_i(T,\theta_0)\log(s_i(T,\theta))$.

[**Question**] Is the only local maxima of $L(\theta)$ at the value $\theta=\theta_0$?

Statistical Challenge: *The 'Inadmissibility conjectures'*
Submitted by Mike Steel

Consider a finite-state Markov process on a phylogenetic tree T which generates an i.i.d. sequence X of k sites (specifically, consider the symmetric Poisson process, a.k.a. the Jukes-Cantor model, when there are just four states). Given a method M for reconstructing phylogenetic trees from sequences let $\Pr(M(X) = T|T,\lambda)$ denote the probability that M correctly returns the tree T when X is generated by the Markov process on T with branch lengths λ. Following decision-theoretic terminology, we say that a method M is *inadmissible* if there exists another method M′ for which $\Pr(M'(X) = T|T,\lambda) \geq \Pr(M(X) = T|T,\lambda)$ for all fully resolved phylogenetic trees T and choice of (strictly positive, but finite) branch lengths λ, and for at least one such pair (T,λ) we have strict inequality.

[**Question**] Prove (or disprove) the following conjectures:

- Maximum Parsimony is inadmissible
- Maximum Likelihood is inadmissible

Note that ML is known to be admissible in case there is a fixed known value of λ for each tree, and ML is performed subject to this constraint.

An 'evolutionary' problem from graph theory

A collection of sets has an *evolutionary ordering* if the sets can be ordered so that (i) each set contains an element that appears in at least one earlier set, and (ii) each set contains an element that does not appear in an earlier set[1].

[**Question**] What is the complexity of determining whether a finite collection of sets has an evolutionary ordering?

[1] Little CH, Campbell AE, (2000) **Evolutionary Families of Sets** *Electronic Journal of Combinatorics* **7**