



# Markovian log-supermodularity, and its applications in phylogenetics<sup>☆</sup>

Mike Steel <sup>\*</sup>, Beáta Faller

Allan Wilson Centre for Molecular Ecology and Evolution, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

## ARTICLE INFO

### Article history:

Received 21 May 2008

Accepted 17 October 2008

### Keywords:

Tree

Markov process

AD inequality

FKG inequality

Phylogenetic diversity

## ABSTRACT

We establish a log-supermodularity property for probability distributions on binary patterns observed at the tips of a tree that are generated under any 2-state Markov process. We illustrate the applicability of this result in phylogenetics by deriving an inequality relevant to estimating expected future phylogenetic diversity under a model of species extinction. In a further application of the log-supermodularity property, we derive a purely combinatorial inequality for the parsimony score of a binary character. The proofs of our results exploit two classical theorems in the combinatorics of finite sets.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Finite-state Markov processes on trees are widely used in evolutionary biology to model the way in which discrete characteristics of present-day species have evolved from the state present in some common ancestor [1,2]. We investigate a generic inequality that applies to 2-state Markov processes on trees, and provide two applications. The first application, which was the motivation for our study, is to the theory of biodiversity conservation. We consider the expected loss of ‘phylogenetic diversity’ under a model in which extinction risk is associated with an underlying state that evolves on the tree. We are interested in comparing this expected loss to simpler models in which extinction events are treated independently; we find that when extinction events reflect phylogenetic history, then the expected loss of phylogenetic diversity is always greater than or equal to that predicted by an independent extinction scenario. In a second application, we derive a new, purely combinatorial result concerning the ‘parsimony score’ of a binary character on a tree. We also briefly discuss how the generic inequality for 2-state Markov processes relates to recent work on phylogenetic invariants and inequalities for particular submodels.

## 2. Markov processes on trees

Consider a Markov random field on a tree  $T$  with state space  $\{0, 1\}$ , and for each vertex  $v$  of  $T$ , let  $\xi(v)$  be the random state (0 or 1) that  $v$  is assigned. This process is usually described as follows. We have a root vertex  $\rho$  for which we specify a probability, say  $\pi_i$ , that  $\xi(\rho) = i$ , for  $i \in \{0, 1\}$ . Direct all the edges of  $T$  away from  $\rho$  and for any arc  $(r, s)$  of the resulting directed tree  $T = (V_T, A_T)$ , let  $P^{(r,s)}$  denote the  $2 \times 2$  transition matrix for which the  $ij$ -entry (for  $i, j \in \{0, 1\}$ ) is the conditional probability that  $\xi(s) = j$  given that  $\xi(r) = i$ . Specifying  $\pi = [\pi_0, \pi_1]$  together with the transition matrices  $P^{(r,s)}$  for all the arcs  $(r, s)$  of  $T$  uniquely defines the Markov random field on  $T$  (see, for example, [3,2,4]); an explicit formula appears below (Eq. (1)). We will assume throughout that  $\pi$  is strictly positive and that  $\det P^{(r,s)} \geq 0$  holds for each transition matrix. Notice that this determinant condition automatically holds if one views the transition matrix for an arc as describing the net effect of a continuous-time Markov process operating for some duration for that arc. Note however that we are not assuming that any such process is the same between the arcs of  $T$ .

<sup>☆</sup> We thank the New Zealand Marsden Fund for supporting this work.

\* Corresponding author.

E-mail address: [m.steel@math.canterbury.ac.nz](mailto:m.steel@math.canterbury.ac.nz) (M. Steel).

For  $U \subseteq V_T$  let  $P(U)$  denote the probability that  $U$  is precisely the set of vertices of  $T$  in state 0; that is:  $P(U) = \mathbb{P}(\{v \in V_T : \xi(v) = 0\} = U)$ . To express  $P(U)$  in terms of the transition matrices and  $\pi$ , let  $\delta(U, v) = 0$  if  $v \in U$  and let  $\delta(U, v) = 1$  if  $v \in V_T - U$ . Then, the Markov property gives:

$$P(U) = \pi_{\delta(U, \rho)} \cdot \prod_{(r,s) \in A_T} P_{\delta(U,r)\delta(U,s)}^{(r,s)}. \quad (1)$$

For any subset  $W$  of the leaf set  $X$  of  $T$ , let  $p_W$  denote the probability that  $W$  is precisely the set of leaves of  $T$  that are in state 0. This marginal probability is:

$$p_W = \sum_{U \in \mathcal{A}_W} P(U), \quad \text{where } \mathcal{A}_W := \{U \subseteq V_T : U \cap X = W\}. \quad (2)$$

A number of authors have noticed that certain inequalities hold for quadratic functions of the  $p_W$  values. For example, for any  $x, y \in X$  with  $x \neq y$ , it is well known that:  $p_{\{x\}} \cdot p_{\{y\}} \leq p_{\{x,y\}} \cdot p_{\emptyset}$ . Moreover, in [5] the following inequality was described: for subsets  $\{x, y\}$  and  $\{x, z\}$  of  $X$  where  $x, y, z$  are distinct, we have  $p_{\{x,y\}}p_{\{x,z\}} \leq p_{\{x,y,z\}}p_{\{x\}}$ . We now provide a much more general inequality.

**Proposition 2.1.** *For any 2-state Markov process on a tree with leaf set  $X$ , and any two subsets  $Y, Z$  of  $X$ , we have  $p_Y \cdot p_Z \leq p_{Y \cup Z} \cdot p_{Y \cap Z}$ .*

**Proof.** Let  $A, B$  be arbitrary subsets of  $V_T$ . We first establish the following:

$$P(A) \cdot P(B) \leq P(A \cup B) \cdot P(A \cap B). \quad (3)$$

Applying Eq. (1) to  $U \in \{A, B, A \cup B, A \cap B\}$ , the product  $P(A) \cdot P(B)$  and the product  $P(A \cup B) \cdot P(A \cap B)$  can each be written as a product of two entries of  $\pi$  multiplied by a product over the arcs  $(r, s)$  of  $T$  of two entries of  $P^{(r,s)}$ . Moreover, regardless of where  $r$  and  $s$  lie in relation to the sets  $A, B$ , the product of the two  $\pi$  terms agree in  $P(A) \cdot P(B)$  and  $P(A \cup B) \cdot P(A \cap B)$  (i.e., we have  $\pi_{\delta(A,\rho)}\pi_{\delta(B,\rho)} = \pi_{\delta(A \cup B,\rho)}\pi_{\delta(A \cap B,\rho)}$ ), while the product of the two  $P^{(r,s)}$  terms agree in  $P(A) \cdot P(B)$  and  $P(A \cup B) \cdot P(A \cap B)$ , except for the cases in which either (i)  $r \in A - B$  and  $s \in B - A$ , or (ii)  $r \in B - A$  and  $s \in A - B$ . However, in both cases (i) and (ii), the product  $P_{01}^{(r,s)}P_{10}^{(r,s)}$  appears in the term for  $P(A) \cdot P(B)$  while  $P_{00}^{(r,s)}P_{11}^{(r,s)}$  appears in the term for  $P(A \cup B) \cdot P(A \cap B)$ , and the former term is less than or equal to the second since  $P_{00}^{(r,s)}P_{11}^{(r,s)} - P_{01}^{(r,s)}P_{10}^{(r,s)} = \det P^{(r,s)}$  and  $\det P^{(r,s)} \geq 0$  by assumption. Consequently, all the terms in  $P(A) \cdot P(B)$  are either less than or equal to (in cases (i) and (ii)), or equal to (in all the remaining cases) the corresponding terms in  $P(A \cup B) \cdot P(A \cap B)$ . This establishes (3).

We now invoke a classical result of Ahlswede and Daykin (1978) [6], sometimes called the ‘four functions theorem’. A particular form of this theorem that suffices for our purposes is the following (we follow [7]). Suppose we have a finite set  $S$  and a function  $\alpha$  that assigns a non-negative real number to each subset of  $S$ . Suppose that  $\alpha$  satisfies the property that for all subsets  $A, B$  of  $S$ :

$$\alpha(A)\alpha(B) \leq \alpha(A \cup B)\alpha(A \cap B).$$

For a collection  $\mathcal{C}$  of subsets of  $S$ , let  $\alpha(\mathcal{C}) := \sum_{C \in \mathcal{C}} \alpha(C)$ . Then for any two collection of subsets of  $S$ ,  $\mathcal{A}$  and  $\mathcal{B}$ , say, we have:

$$\alpha(\mathcal{A})\alpha(\mathcal{B}) \leq \alpha(\mathcal{A} \vee \mathcal{B})\alpha(\mathcal{A} \wedge \mathcal{B}), \quad (4)$$

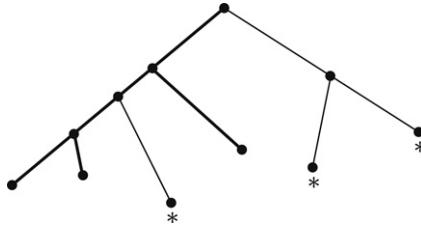
where  $\mathcal{A} \vee \mathcal{B} := \{E \subseteq S : E = A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\}$ , and where  $\mathcal{A} \wedge \mathcal{B} := \{E \subseteq S : E = A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$ . We will apply this to our problem by taking  $S = V_T$ ,  $\alpha(U) = P(U)$  and noting that  $\alpha$  satisfies the required hypothesis by (3). By the definition of  $\mathcal{A}_W$  in (2),  $\mathcal{A}_Y \vee \mathcal{A}_Z = \mathcal{A}_{Y \cup Z}$  and  $\mathcal{A}_Y \wedge \mathcal{A}_Z = \mathcal{A}_{Y \cap Z}$ . Thus taking  $\mathcal{A} = \mathcal{A}_Y$  and  $\mathcal{B} = \mathcal{A}_Z$  in (4) we have  $\alpha(\mathcal{A}_Y)\alpha(\mathcal{A}_Z) \leq \alpha(\mathcal{A}_{Y \cup Z})\alpha(\mathcal{A}_{Y \cap Z})$ . The proposition now follows by observing that  $p_W = \alpha(\mathcal{A}_W)$  for all subsets  $W$  of  $X$ , in particular the subsets  $Y, Z, Y \cup Z$  and  $Y \cap Z$ .  $\square$

### 3. Applications in phylogenetics

We first show how Proposition 2.1, together with another inequality, provides a general inequality concerning the loss of expected future biodiversity under species extinction models. Suppose that  $T$  is a rooted tree with leaf set  $X$ , and with each arc  $e = (u, v)$  of  $T$  there is an associated length  $\lambda_e$ . Given a subset  $Y$  of  $X$ , the *phylogenetic diversity* (PD) of  $Y$ , denoted  $\varphi_Y$ , is the sum of the lengths of the edges of the minimal subtree of  $T$  connecting the root and the leaves in  $Y$ . Under various possible interpretations of the  $\lambda$  values, PD has been widely used as a measure for quantifying present and expected future biodiversity [8–10].

For each species  $x \in X$  let  $E_x$  denote the event that species  $x$  is extinct at some future time  $t$ . Then the expected phylogenetic diversity of the species that are extant at time  $t$ , referred to as *expected future PD* and denoted  $\mathbb{E}[\varphi]$ , is given by:

$$\mathbb{E}[\varphi] = \sum_{e=(u,v) \in A_T} \lambda_e \cdot \left( 1 - \mathbb{P} \left( \bigcap_{x \in C_v} E_x \right) \right) = \varphi_X - \sum_{e=(u,v) \in A_T} \lambda_e \cdot \mathbb{P} \left( \bigcap_{x \in C_v} E_x \right), \quad (5)$$



**Fig. 1.** If the species indicated by \* become extinct, then the remaining PD is the sum of the lengths of the bold edges indicated.

where  $C_v$  denotes the subset of  $X$  that is separated from the root by  $v$ . A simple model, referred to as the *generalized field of bullets model* (g-FOB) in [9] (generalizing an earlier model from [10]), assumes that the events  $E_x$  are independent. Then, if we let  $p_x = \mathbb{P}(E_x)$ , the value of  $\mathbb{P}(\bigcap_{x \in C_v} E_x)$  in (5) (the probability of the extinction of all the species descended from  $v$ ) is given by:

$$\mathbb{P}\left(\bigcap_{x \in C_v} E_x\right) = \prod_{x \in C_v} p_x. \quad (6)$$

An example to illustrate this concept is provided in Fig. 1. The assumption that the events  $E_x$  are independent is likely to be unrealistic in most settings (see, for example, [11,12]). For example, species ‘close together’ in  $T$  are more likely to share attributes that may put them at risk in a hostile future environment. As one topical scenario, consider extinction risk due to climate change. Suppose that the extinction risk of each species in  $X$  is partially influenced by some associated binary state (0 or 1) where state 0 confers an elevated risk of extinction under climate change. We suppose that these states are not known in advance for the species in  $X$ , and that this state has evolved under some Markovian model on  $T$ . Once the states are determined at the leaves, then extinction proceeds according to the g-FOB model, where species  $x$  is extinct at time  $t$  with probability  $p_x^i$  if it is in state  $i \in \{0, 1\}$ . We call this a *state-based field of bullets model* (s-FOB). Note that this includes the g-FOB model as a special case where  $p_x^0 = p_x^1$  for all  $x$ . Moreover, once we condition on the state for each leaf, an s-FOB model is just a g-FOB model with modified extinction probabilities, but we are assuming that these states are unknown (in line with the uncertainty over what features may be helpful for an organism in a future climate).

With any s-FOB model we also have an associated g-FOB model in which the extinction probability of each species  $x$  is the same as in the s-FOB model. That is, in the g-FOB model we set:

$$p_x = p_x^0 \mathbb{P}(\xi(x) = 0) + p_x^1 \mathbb{P}(\xi(x) = 1), \quad (7)$$

where  $\xi$  describes the Markov process for the binary character. A natural question arises: how does the future expected PD of an s-FOB model compare with that of its associated g-FOB model? The following result provides a general inequality.

**Theorem 3.1.** Consider a fixed tree with branch lengths and leaf set  $X$ . Consider an s-FOB model, in which state 1 is advantageous for each species, i.e.,  $p_x^1 \leq p_x^0$  for all  $x \in X$ . Then the expected future PD of this model is less than or equal to the expected future PD of the associated g-FOB model.

**Proof.** In view of (5) and (6), it suffices to show that:

$$\prod_{x \in C_v} p_x \leq \mathbb{P}\left(\bigcap_{x \in C_v} E_x\right), \quad (8)$$

where  $p_x$  is defined by Eq. (7). For each subset  $W$  of  $C_v$  let  $p_W$  denote the probability that the set of elements of  $C_v$  in state 0 is precisely  $W$ . Then,  $\mathbb{P}(\bigcap_{x \in C_v} E_x) = \sum_{W \subseteq C_v} p_W \prod_{x \in W} p_x^0 \prod_{x \in C_v - W} p_x^1$ . Thus, if we let  $f_x(W) = p_x^0$  if  $x \in W$ , and  $f_x(W) = p_x^1$  if  $x \in C_v - W$  then  $\mathbb{P}(\bigcap_{x \in C_v} E_x) = \sum_{W \subseteq C_v} p_W \prod_{x \in C_v} f_x(W)$ . Moreover,  $p_x = p_x^0 \mathbb{P}(\xi(x) = 0) + p_x^1 \mathbb{P}(\xi(x) = 1) = \sum_{W \subseteq C_v} p_W f_x(W)$ , where the second equality arises by considering in the summation those  $W$  containing  $x$  and those not containing  $x$ . Consequently, (8) is equivalent to the requirement that:

$$\prod_{x \in C_v} \left( \sum_{W \subseteq C_v} p_W f_x(W) \right) \leq \sum_{W \subseteq C_v} p_W \prod_{x \in C_v} f_x(W). \quad (9)$$

The proof of (9) involves combining Proposition 2.1 with the FKG inequality of Fortuin, Kasteleyn and Ginibre (1971) [13], a particular (and multivariate) form of which we now recall. Given a finite set  $S$ , suppose that  $f_1, f_2, \dots, f_n$  are functions from the power set of  $S$  into the non-negative real numbers, and that these satisfy the condition:

$$A \subseteq B \Rightarrow f_i(A) \leq f_i(B). \quad (10)$$

Furthermore, suppose that  $\mu$  is a probability measure on the subsets of  $S$  which satisfies the log-supermodularity condition:

$$\mu(A)\mu(B) \leq \mu(A \cup B)\mu(A \cap B). \quad (11)$$

Then:

$$\prod_{i=1}^n \left( \sum_A \mu(A) f_i(A) \right) \leq \sum_A \mu(A) \prod_{i=1}^n f_i(A), \quad (12)$$

where the summations are over all subsets of  $S$ .

We apply this form of the FKG inequality by taking  $S = \{1, \dots, n\} = C_v$ ,  $\mu(W) = p_W$ , and  $f_x$  as defined above. Then  $f_x$  satisfies (10) by the hypothesis that  $p_x^1 \leq p_x^0$  for all  $x$ , while  $\mu$  satisfies (11) by Proposition 2.1. Then inequality (12) provides the required inequality (9). This completes the proof.  $\square$

### 3.1. Combinatorics of parsimony

We now provide a second application of Proposition 2.1 to phylogenetics. Given a function  $f : X \rightarrow \{0, 1\}$ , the *parsimony score* of  $f$  on a tree  $T$  with leaf set  $X$ , denoted  $l(f, T)$ , is the minimum number of edges that have different states assigned to their endpoints, across all extensions  $F : V_T \rightarrow \{0, 1\}$  of  $f$  (for further details see [2]). For  $W \subseteq X$ , let function  $f_W$  assign state 0 to the elements of  $W$ , and state 1 to the elements of  $X - W$ . We show that the parsimony score function for a given tree is submodular.

**Theorem 3.2.** *For any tree  $T$  with leaf set  $X$  and subsets  $Y, Z$ , of  $X$  we have:*

$$l(f_Y, T) + l(f_Z, T) \geq l(f_{Y \cup Z}, T) + l(f_{Y \cap Z}, T).$$

**Proof.** Consider the 2-state Markov random field on  $T$  with  $\pi_0 = \pi_1 = 0.5$ , and set each transition matrix  $P^{(r,s)}$  to be the symmetric  $2 \times 2$  matrix with off-diagonal entry  $\epsilon > 0$ . Then, for any  $W \subseteq X$  a straightforward calculation shows that:

$$p_W = C_W \epsilon^{l(f_W, T)} (1 + o(\epsilon)), \quad (13)$$

for a constant  $C_W$  that depends only on  $W$  and  $T$  and not  $\epsilon$  (specifically,  $C_W$  is the number of minimal extensions of  $f_W$  to the vertices of  $T$  multiplied by  $\frac{1}{2}$ ). Now Proposition 2.1, expressed using logarithms, states that:

$$-\log(p_Y) - \log(p_Z) \geq -\log(p_{Y \cup Z}) - \log(p_{Y \cap Z}). \quad (14)$$

Applying (13) (and noting that  $\log(1 + o(\epsilon)) = o(\epsilon)$ ), the left-hand side of (14) is:

$$(l(f_Y, T) + l(f_Z, T)) \log\left(\frac{1}{\epsilon}\right) - \log(C_Y C_Z) + o(\epsilon),$$

while the right-hand side of (14) is:

$$(l(f_{Y \cup Z}, T) + l(f_{Y \cap Z}, T)) \log\left(\frac{1}{\epsilon}\right) - \log(C_{Y \cup Z} C_{Y \cap Z}) + o(\epsilon).$$

Theorem 3.2 now follows by letting  $\epsilon$  tend to zero.  $\square$

## 4. Concluding remark

As a further phylogenetic application, we note that Proposition 2.1 provides a collection of polynomial inequalities on the  $p_W$  values, which have recently been studied for a particular class of Markov 2-state models in [14]. These polynomial inequalities complement the much-studied ‘phylogenetic invariants’ (polynomial identities in the  $p_W$  values), which hold under various restrictions on the Markov model.

## References

- [1] J. Felsenstein, Inferring Phylogenies, Sinauer Press, 2003.
- [2] C. Semple, M. Steel, Phylogenetics, Oxford University Press, Oxford, 2003.
- [3] J.T. Chang, Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency, *Math. Biosci.* 137 (1996) 51–73.
- [4] M. Steel, Recovering a tree from the leaf colourations it generates under Markov model, *Appl. Math. Lett.* 7 (1994) 19–23.
- [5] J. Pearl, M. Tarsi, Structuring causal trees, *J. Complexity* 2 (1986) 60–77.
- [6] R. Ahlsweide, D.E. Daykin, An inequality for the weights of two families of sets, their unions and intersections, *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* 43 (1978) 183–185.
- [7] I. Anderson, Combinatorics of Finite Sets, Dover Publications, New York, 1987.
- [8] D.P. Faith, Conservation evaluation and phylogenetic diversity, *Biol. Conserv.* 61 (1992) 1–10.
- [9] B. Faller, F. Pardi, M. Steel, Distribution of phylogenetic diversity under random extinction, *J. Theoret. Biol.* 251 (2008) 286–296.
- [10] S. Nee, R.M. May, Extinction and the loss of evolutionary history, *Science* 278 (5338) (1997) 692–694.
- [11] S.B. Heard, A.O. Mooers, Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions, *Proc. Roy. Soc. London B* 267 (2000) 613–620.
- [12] H. Simianer, Accounting for non-independence of extinction probabilities in the derivation of conservation priorities based on Weitzman’s diversity concept, *Conserv. Genet.* 9 (2008) 171–179.
- [13] C.M. Fortuin, P.W. Kasteleyn, J. Ginibre, Correlation inequalities on some partially ordered sets, *Comm. Math. Phys.* 22 (1971) 89–103.
- [14] F.A. Matsen, Fourier transform inequalities for phylogenetic trees, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 6 (2009) 89–95.