

# Foundations of the Age-Area Hypothesis

Preliminary and Incomplete - Not for Citation

**Matthew J. Baker**

Department of Economics  
Hunter College and the Graduate Center, CUNY

March 19, 2018

## **Abstract**

A commonly-used tool in explaining the origins, geographical dispersion, and degree of relatedness between cultures is the Age-Area Hypothesis developed by Sapir (1916). The hypothesis posits that the origin point of a group of related cultures or a linguistic stock is most likely where the languages comprising the stock are most divergent. While compelling and often-corroborated, the Age-Area Hypothesis has never been founded in basic principles, and this lack of structure limits its use. I describe a micro-founded model of the Age-Area Hypothesis based on mass migration, develop a measure of divergence between cultures based on phylogenetic trees, and develop an Age-Area Theorem. The model allows computation of probabilities that different locations are the origin points for the constituents of a group of related cultures. The model also provides a probabilistic explanation of Occam's razor: migratory paths explaining the geographical dispersion of related cultures that are simpler are more likely. The paper concludes with applications to the origins of the Na-Dene and Afro-Asiatic Ethno-linguistic groups.

# 1 Introduction

Studying the impact of ethnic, genetic, and cultural diversity on economic outcomes has become a central part of economics. Ashraf and Galor (2013), for example, discuss the nuanced role that genetic diversity might play in driving modern economic outcomes, while Spolaore and Wacziarg (2009) describe the importance of genetic distance and economic distance between populations.

As Cavalli-Sforza and Cavalli-Sforza (1995) have pointed out, genetic, ethnic, linguistic, and cultural diversity are all closely related, and much recent economic research (reviewed in Alesina et al. (2005)) has focused on the relationship between ethnic diversity, fractionalization and economic growth. Some research, such as Spolaore and Wacziarg (2013), develops a link between ancient institutions and cultural practices on modern economic outcomes.

Other work has pushed this agenda in a different direction, with the aim of understanding the origins of ethnolinguistic and genetic diversity around the world. Michalopoulos (2012) shows that geographical diversity is an important determinant of human diversity. Ahlerup and Olsson (2012) develop a model of ethnic diversity based on genetic drift and migration. They show that, over time, populations spread out and experience cultural drift, and that populations that have been in place longer tend to be more culturally diverse. This echoes a long and rich tradition in other disciplines, as detailed in Mace et al. (2005).

A - perhaps “the” - key contributor to the world-wide distribution of cultures are historical, large-scale mass migrations. And a critical tool in understanding the timing and sequence of events behind mass migrations is the so-called Age-Area Hypothesis (henceforth AAH). A leading early development and application of the idea is Sapir (1916), in his study of the geographical origins of the North American Na-Dene language group. The AAH is often invoked when linguistic evidence is used to develop or support hypotheses about where groups of related cultures originated, how they evolved over time, and how the current geographical distribution of the cultures came to be.<sup>1</sup> The AAH provides a critical link between geography and linguistic phylogenies. While one

---

<sup>1</sup>One must be careful in invoking “the” AAH, as there are possibilities for confusion with other, sometimes unrelated, ideas. The AAH in this paper is not to be confused with, for example, the Sapir-Whorf Hypothesis, which refers to the idea that peoples’ thinking is influenced by the structure of their language. What might be called the Cultural Age-Area Hypothesis is the principle that the more widely geographically distributed a cultural trait is, the older it likely is.

can find competing and sometimes contrary definitions in the literature, briefly stated, the hypothesis says that the geographic area where a phylogeny originated is most likely the place where the component languages of the phylogeny are most divergent, or maximally differentiated.

Applications - either implicit or explicit - abound. Atkinson and Gray (2003), following Renfrew (1987) hypothesis and Dolgopolsky (1988), couple computational linguistics with the AAH to suggest that the Indo-European languages originated in Anatolia, not, as has sometimes been argued, on the steppes of Siberia. Ruhlen (1994) uses the AAH and also provides overviews of debates about the origins of the Na-Dene cultures, the Bantu expansion in Africa, and the peopling of the South Pacific. Ehret (2002) Makes extensive and efficient use in hit sweeping account of how and when the cultures of Africa found their current locations.

In spite of its widespread invocation, there is, to my knowledge, no theoretical basis for the hypothesis. The lack of a theoretical basis for the AAH is a problem not only because it leads to imprecise definition, but also because it is difficult to use in technical situations. Suppose one wished to include a migratory path in a statistical model of migration and cultural evolution, or that one wanted to integrate the timing and path of migration into a larger statistical model, perhaps with an eye toward more accurately controlling for relatedness between differing peoples. Since the hypothesis has no theoretical underpinnings, it is difficult to see how to construct a likelihood for a certain path or migratory history, or even for the likelihood of competing hypotheses.

I develop a theoretical foundation for the age-area hypothesis. The model is micro-founded in the sense that it can be based on microeconomic principles and an associated back story. The theory shows how probabilities that different locations as being the point of origin and Occam's razor are interdependent. That is, it turns out that migratory histories that are simpler in a well defined sense are also more likely. The model indicates that migratory paths that originate at deeper points in a phylogenetic tree can be thought of as simpler in a very precise fashion, in that inclusion of such paths in a migratory history also results in a model with fewer parameters. These ideas also lead to a particular definition of divergence or dissimilarity for each constituent culture in the phylogeny.

I then develop an Age-Area Theorem using these ideas. The theorem presents assumptions under which a culture that is more linguistically divergent from the others in the stock is also more likely to reside at the point of origin of the stock. I then

present a computational algorithm which relies upon backwards traversal of a phylogenetic tree, along with some extensions of the model. I conclude by discussing some applications. The theorem, when applied recursively, also yields what can be construed as a maximum-likelihood migratory path.

It is worth emphasizing that the model can be used to formulate an likelihood function for migratory patterns, and hence can be included as part of a larger estimation problem. Maximum likelihood methods are commonly used in linguistics to estimate linguistic divergence times. So, using the results of this paper, one could in principle estimate or include the implied migratory path as part of a larger phylogenetic estimation problem. Likelihood methods are also amenable to Bayesian techniques, which expand estimation possibilities tremendously, and provide an avenue for including archaeological and genetic evidence in estimation as well.

## 2 Background Literature

One of the dominant means of describing how closely-related cultures are is through language and linguistic drift, which turns out to be a fairly reliable guide to how far in the past two similar cultures began to drift apart after some event - often a large-scale migration - lead to their geographical separation. The AAH is a primary tool linking geography with cultural drift. Trask (2000, p.12) attributes the (Linguistic) Age-Area Hypothesis - also called the center-of-gravity principle, the genetic diversity principle, and Sapir's principle - to the work of Latham (1851) and Sapir (1916). Trask (2000) further notes that Mallory (1997) and Nichols (1997) have employed and qualified the AAH in various ways. Dimmendaal (2011, p.336) expresses some doubts about the exact origins of the AAH, referring to it as the "principle of least effort," while noting that "This principle probably was applied first by the scholars working on Amerindian Languages, e.g. Sapir (1916) and Dyen (1956)".

In what seems to be the initial application of the idea, Sapir (1916) writes:

As is well known, [Athabaskan] languages are spoken in three geographically isolated areas, a very large northern area (interior of Alaska to near Hudson Bay), a Pacific area (southwestern Oregon and Northwestern California), and a southern area (Arizona, New Mexico, and western Texas)...it would seem that the historical center of gravity lies rather in the north

than in either of the other two regions and that the occupation of these latter was due to a southward movement of Athabaskan-speaking tribes. It is important to observe that the argument is not in any way dependent on the fact that the northern tribes cover a much vaster territory than those of the other two groups or even directly on the fact that probably a larger number of distinct dialects are spoken in the north than elsewhere. The argument for the northern provenience of the Athabaskan tribes is clinched by a further linguistic fact, namely that the Athabaskan dialects from one of the three major divisions of the Na-dene stock, the other two being Haida and Tlingit. The fact that the latter are spoken in the northwest coast area so emphatically locates the historical center of gravity of the stock in the north that it becomes completely impossible to think of the Athabaskan tribes as having spread north from California or the southwest.

Dyen (1956, p. 623) also cites this passage, and goes a bit further in cleaning up the theory and defining how things work in practice, describing concepts such as migrations and even stating two basic postulates to be employed in assessing homelands for linguistics. Her first postulate is that the area of origin of related languages is continuous. The second postulate, which is more important for the purposes of this paper, is that “the probabilities of different reconstructed migrations are in inverse relation to the number of language movements that is required.” (Dyen, 1956, p.613) To some degree, my principle objective in this paper is to clarify and develop Dyen’s second postulate, rather than simply asserting that it is true.

As a final note, consider some of the comments made in Greenhill and Gray (2005), who discuss the peopling of the South Pacific and also present a detailed discussion of quantitative methods in historical linguistics. They develop statistical tests comparing different hypotheses for how the South Pacific came to be settled. They write the following in describing the need for formal modeling and associated hypothesis tests in resolving disputes about migratory routes:

“...many expansion scenarios are little more than plausible narratives. A common feature of these narratives is the assertion that a particular line of evidence (archaeological, linguistic, or genetic) is ‘consistent with’ the scenario. ‘Consistent with’ covers a multitude of sins. Rigorous tests require a measure of exactly how well the data matches the proposed scenario. They

also require an explicit evaluation of alternative hypotheses. ...a framework for the rigorous evaluation of these hypotheses is clearly desirable. (Greenhill and Gray, 2005, p. 31)

This statement could easily have been written in describing the reason for the current paper, and as a justification for taking building a formal model around the AAH.

### 3 Problem Description and Motivation

Consider the hypothetical phylogenetic tree displayed in figure 1. For clarity, the figure depicts the phylogenetic relationship between cultures, which has perhaps been obtained through a careful analysis of the languages of the cultures so that one can deduce the times at which the cultures seem to have separated. *A* is evidently the most divergent culture in that the last time at which the cultures had a common ancestor is further in the past than it is for any of *B*, *C*, *D*, or *E*, all of whom are closer relatives to one another. *D* and *E* are the most closely related cultures. To avoid forming impressions based on evidence that is not part of the hypothesis, we will suppose that the locations of the groups have been concealed from us, and that we have no access to archaeological or historical information about the constituent cultures of the phylogenetic tree came to be in their current locations.

The AAH asserts that the geographical origins of this cultural group is *A*'s current location, as *A* is the most divergent culture (speaker of the most divergent language, that is) from the group. Recursive application of the AAH would lead one to a most likely migratory route: the stock originated at *A*'s location. There was then a migration from *A*'s location to *B*'s location, then from *B* to *C*, and then from *C* to either *D* or *E*.

Why would one believe this was the most likely explanation for how the cultures came to be in their current locations? There are a number of other possible routes, even when working within the constraints of the Phylogenetic Tree. It can never be the case that a migration from *D* to *E* preceded one from *D* to *C* as this would be inconsistent with the observed linguistic drift over time. But another possible sequence of migratory events would be for an initial migratory episode from *C* to *A*, followed by another from *C* to *B*, followed by yet another migration from *C* to *D* or *E*.

These two possible migratory routes are presented in figure 2, and the AAH asserts

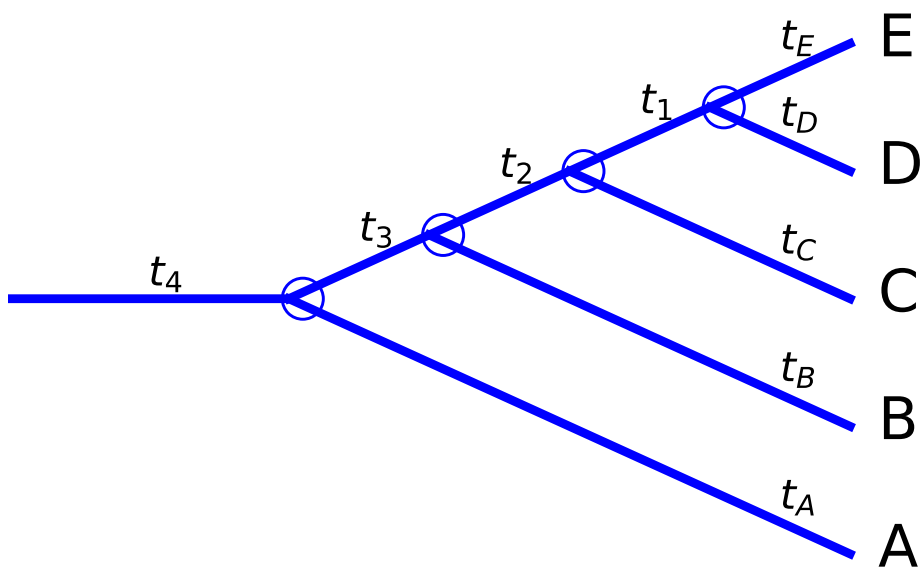


Figure 1: A Phylogenetic tree

## Two Hypothetical Migratory Routes

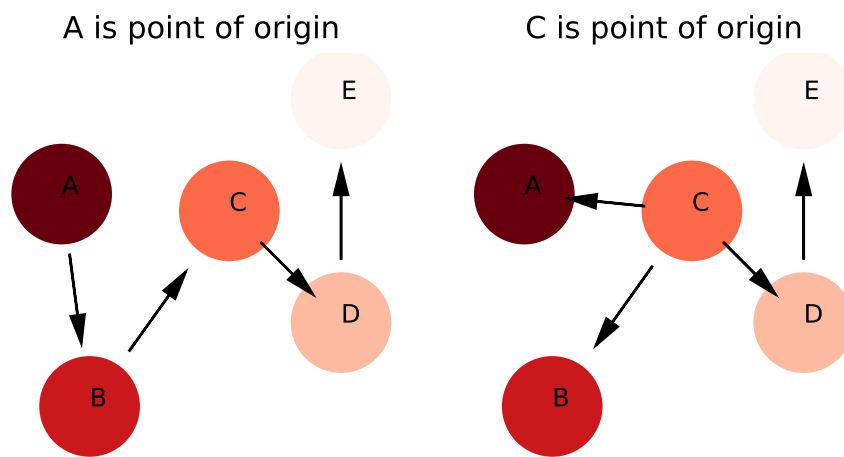


Figure 2: Potential migratory routes among a language stock consisting of five groups, following the Phylogenetic Tree in figure.



that the right-hand sequence is a less-likely migratory history than that on the left. Why? One might simply appeal to Occam's razor - the events on the left-hand side of figure 2 require only one migratory chain, while the events on the right-hand side require three separate expansions: an initial migration from A to C, followed by another from A to B. A third migratory event is sufficient to take care of the last two groups: it begins from C continues to D, and then to E. Also clear from this example is the limited usefulness of positing a minimal number of moves to explain migrations. The above paths both require only four distinct population movements.

I can now pose the question as follows: if one believes migratory events to be rare, and wishes to conserve them in explaining historical migrations, what sort of model accomplishes this? And one characterize parsimony in a meaningful mathematical and probabilistic way?

## 4 A model

### 4.1 Tree Basics

I assume throughout that the entire phylogenetic tree is known, so the universe of migratory events that needs to be explained is observed. The tree can be constructed by any number of methods.<sup>2</sup> I also assume the tree is a full, rooted, binary tree, meaning that there is a single origin node and branch, all other nodes have a single parent, all interior nodes have exactly two children, and all terminal nodes have zero children.

A binary tree with  $n$  taxa or leaves has  $n - 1$  internal nodes. The tree in figure 1, for example, has five terminal nodes/taxa and four internal nodes. This also applies to subtrees, which means that if a node has  $k$  successors (children and childrens' children, etc.), then it is followed by  $k - 1$  nodes.

### 4.2 Migratory Chains

I build the model around *migratory chains*, which can be thought of as sequences of *migratory events*, similar to the way in which they are defined in Dyen (1956). A migratory chain is manifest on a phylogenetic tree as a directed path from the root of a tree (or subtree) to a terminal node. Geographically, a migratory event corresponds

---

<sup>2</sup>See, for example, Nichols (1997) for a survey of methods.

with a movement of a population to a new geographical location, with a migratory chain being composed of a sequence of such movements unfolding over time.<sup>3</sup>

I call a collection of migratory chains that span the entire phylogenetic tree a *migratory history*. Hypothesizing that a particular culture or society is the point of origin of the tree amounts to positing some sequence of migratory chains through the phylogenetic tree, of which the deepest chain starting from the root starts from the point of origin. I assume that a migratory chain has the following properties.

**Assumptions 1. Migratory Chains**

1. *Each migratory chain occupies a single location at any given point in time.*
2. *When a chain moves from its current location, a new chain starts in its place at the given location.*
3. *Migratory chains move at random times according to an exponential density.*
4. *Each migratory chain is distinct in that it has its own parameterized probability density.*

I will call a migratory history originating at location  $k$   $H_k$ , and will assume that candidate migratory histories require the minimal number of migratory events or moves to explain the current distribution of the member cultures of a phylogeny. In the time dimension, a migratory chain is then a directed path through a sequence of nodes to a leaf node, and in space, a migratory chain appears at a sequence of jumps at randomly distributed times. Collectively, chains comprise a history if the tree is covered by chains. I denote the component chains of a history as  $C_{jk}$ .

*Any* history employing the minimum number of moves will need exactly  $n - 1$  migratory events to span the tree. The events depicted in figure 2 following from the tree in figure 1 has 4 arrows between nodes, which corresponds with the number of migratory events required. This will be true regardless of where the migratory history is started.

Each location where a society or culture represented by a terminal node resides,  $k = 1, 2, 3, \dots K$ ; can be associated with a set of migratory histories  $\mathcal{H}_k$ . The previous statement accentuates the fact that there is typically more than one migratory history

---

<sup>3</sup>For the time being, I posit that a migratory event is an instantaneous jump to a new location. I will discuss this further below.

that is possible from origin point  $k$ . Each migratory history has some count of the total number of migratory chains required, which I refer to as  $N(H_k)$ . For the example on figure 2, for the first history  $N(H_A) = 1$ , while for the second,  $N(H_C) = 3$ . I also define a counting function for specific migratory chains as  $n(C)$ , which counts the number of migratory events or nodes spanned by the chain.

These ideas can be applied to develop a likelihood associated with any particular history, and a related measure of divergence which I refer to as ‘Dyen Divergence,’ after Dyen (1956), who anticipated some of the key ideas in this paper. The Dyen Divergence relates phylogenetic distance and probability precisely, so I can state and prove an ‘Age-Area Theorem’ that relates migratory histories, counts of migratory chains, counts of events within each chain, and likelihood. Parsimony or Occam’s razor suggests that simpler explanations - those with smaller values for  $C$ ’s - have higher probabilities and this is in indeed true under Assumptions 1, owing to the shape of the exponential distribution, as I now show in a preview of the logic used to state the theorem. I develop a comparison of the two scenarios described in figures 1 and 2 within the confines of the model. I begin by assuming that we do not know the exact times at which branching events occur, but instead only know the basic structure of the tree. The points can be treated as known but can be ‘integrated out’ by replacing the Exponential model with a Poisson model.

The first migratory history described on figure 2 requires an initial migratory chain to start at the root of the tree, which then proceeds from location  $A$  to  $B$ , then from  $B$  to  $C$  and then finally to  $D$  or  $E$ .<sup>4</sup> But each time the migratory chain proceeds to a new location, by Assumption 1, a new one starts in its place. Importantly, in the example in figures 1 and 2, these new chains never create any new migratory events and only lead to terminal nodes of the tree.

The probability of observing this sequence of events can be written by combining the densities of the components of  $H_A$ , which I can write as:<sup>5</sup>

$$L_A = \text{Prob}(H_A) = P(C_{1A})P(C_{AA})P(C_{BA})P(C_{CA})P(DA)$$

---

<sup>4</sup>In the next subsection, I will describe a more intricate example in which a decision must be made among multiple possibilities.

<sup>5</sup>Throughout I shall use numbers to denote interior branches and nodes along a tree, and the labels of endpoints to denote the terminal branches of the tree.

Using the form of the distribution for the chains, I have:

$$L_A = \frac{(\lambda_1 T)^4 e^{-\lambda_1 T}}{4!} \frac{(\lambda_2 t_6)^0 e^{-\lambda_2 t_6}}{0!} \frac{(\lambda_3 t_7)^0 e^{-\lambda_3 t_7}}{0!} \frac{(\lambda_4 t_8)^0 e^{-\lambda_4 t_8}}{0!} \quad (1)$$

Equation (1) is perhaps more expansive than necessary, as all the terms raised to the 0th power are degenerate. The probability in equation (1) could also be written as:

$$L_A = \frac{(\lambda_1 T)^4 e^{-\lambda_1 T}}{4!} e^{-\lambda_2 t_6} e^{-\lambda_3 t_7} e^{-\lambda_4 t_8} \quad (2)$$

The log-likelihood associated with equation (2) is:

$$\ln L_A = 4 \ln(\lambda_1 T) - 4\lambda_1 T - \ln(4!) - \lambda_2 t_6 - \lambda_3 t_7 - \lambda_4 t_8 \quad (3)$$

What values of  $\lambda_i$  maximize the likelihood in (3)? Evidently,  $\lambda_2 = \lambda_3 = \lambda_4 = 0$  at the optimum. Since these chains never go further than their current locations, the maximum likelihood estimate suggests a rate parameter equal to zero.

The derivative of (3) with respect to  $\lambda_1$ , however, generates a meaningful parameter estimate. Differentiating with respect to  $\lambda_1$  and solving the first-order condition gives  $\lambda_1^* = \frac{4}{T}$ . Substituting this and other optimal values back into the objective function gives the (concentrated) likelihood  $L_A$  as:

$$L_A = \frac{4^4 e^{-4}}{4!} \quad (4)$$

Equation (4) is simple in that only one non-degenerate migratory chain is needed to explain the whole tree, given that the migratory history starts at  $A$ . In this sense, this a parsimonious explanation for the current distribution of cultures.

Contrast this with the case in which  $C$  is posited to be the origin point. To maintain consistency with the phylogeny, the requirements are: 1) a migratory chain starting at  $C$  leading to  $A$ , 2) another migratory chain starting at  $C$  going to  $B$ , and then 3) a migratory chain that starts at  $C$  and proceeds to  $D$  (or  $E$ ) and then finally to  $E$  (or  $D$ ). Degenerate chains start at location  $C$  and  $D$  (or  $E$ ). According to the model, each of these chains requires its own Poisson/Exponential parameter, which, omitting degenerate chains, gives:

$$L_C = \frac{(\lambda_1 t)^1 e^{-\lambda_1 t}}{1!} \frac{(\lambda_2 (t - t_0))^1 e^{-1}}{1!} \frac{(\lambda_3 (t - t_0 - t_1))^2 e^{-\lambda_3 (t - t_0 - t_1)}}{2!} \quad (5)$$

Maximizing  $L_C$  in (5) with respect to  $\lambda_1, \lambda_2$  and  $\lambda_3$ , and substituting the result back into the right-hand side of (5) gives:

$$L_C = \frac{1^1 e^{-1}}{1!} \frac{1^1 e^{-1}}{1!} \frac{2^2 e^{-2}}{2!} = \frac{2^2 e^{-4}}{2!} \quad (6)$$

Suppose that the origin point was known to be one of these two locations perhaps because of archaeological or historical evidence. According to equation (6), the likelihood that  $A$  is the point of origin relative to  $C$  depends upon  $L_A/(L_A + L_C)$ ; a race between the functions  $\frac{4^4}{4!}$  and  $\frac{2^2}{2!}$ , so the relative probability  $A$  is the point of origin would then be 84%.

A key feature of the previous analysis is the function:

$$h(n) = \frac{n^n}{n!}$$

which is convex in  $n$ . This convexity, which owes to the structure of the Poisson-Exponential distribution, implies that one is better off with explanations that require fewer non-degenerate parameters, or, equivalently, more degenerate parameters, which have the effect of lumping migratory events into longer chains. Breaking up a migratory chain into two smaller chains eschews this convexity, and results in a less likely explanation. That is, for any  $k \in (1, n - 1)$ , it is true that:

$$h(n) > h(n - k)h(k)$$

## 5 Divergence

The biggest detail that needs to be worked out in support of the above argument is how migratory chains that are more parsimonious and therefore more likely relate to a measure of phylogenetic divergence. For any migratory history  $H_k$ , there will be some migratory chain that has the largest value of  $n$ , and accordingly, define:

$$n_{H_k}^* = \max_{C_{ik} \in H_k} \{n(C_{1k}), n(C_{2k}), \dots, n(C_{N(H_k)k})\}$$

That is,  $n_{H_k}^*$  is the node count for the longest component chain of history  $H_k$ . The following couples this definition with the previously-introduced count of migratory chains to develop what I refer to as Dyen divergence, as Dyen (1956) went far in developing

the basic concepts.

**Definition 1. Dyen Divergence:** Define a function for a migratory history  $D_{H_k} = m(n_{H_k}^*, N(H_k))$ . that is increasing in the first argument and decreasing in the second. Define the **Dyen Divergence** of culture  $k$  in terms of the history  $H_k^*$  as:

$$D_k = \max\{D_{H_{1k}}, D_{H_{2k}}, \dots, D_{H_{I_k}}\}$$

The above definition defines not ‘a’ measure of divergence, but a family of such measures. The measure itself is a) increasing in the number of events associated with the largest chain in the history, and b) decreasing in the number of chains in the history. The largest such value is the Dyen Divergence of culture  $k$ . Using the probability functions and this measure of divergence, one can now show that larger divergences coincide with more likely chains.

## 6 Age-Area Theorem

**Theorem 1** (Age-Area Theorem). *Suppose that Assumptions 1 hold, and define a Dyen Divergence measure as in definition 1. Then*

$$D_k \geq D_j \implies L_k \geq L_j$$

and in particular

$$k = \arg \max [D_1, D_2, D_3, \dots, D_n] \implies k = \arg \max [L_1, L_2, L_3, \dots, L_n]$$

*Proof.* The probability that  $k$  is the point of origin is a product of the functions  $h(n)$ , which are convex in  $n$ , the number of migrations covered by a migratory chain. Thus, if  $C_i$  chains are required, the length of each are  $N_i$ , we may write:

$$L_k \propto \prod_{j=1}^{N(H_k^*)} h(n_j)$$

By the convexity of  $h(N)$ , and the fact that  $\sum n_j = I$ , where  $I$  is the number of internal nodes of the phylogenetic tree,  $L_k$  is maximized by the smallest number of chains, with the largest  $n$ . The second part of the theorem follows because any finite ordered set

has a maximum in the set.

[Needs to be expanded]

□

The theorem is vague about the exact definition of divergence, but suggests that any measure obeying the definitions 1, that is also decreasing in the number of chains required from a location, and increasing in the length of the longest chain will do. One candidate would be the ratio of the two:

$$D_i^1 = \frac{n_{H_k^*}^*}{N(H_k^*)}$$

another candidate would be the difference:

$$D_i^2 = n_{H_k^*}^* - N(H_k^*)$$

The above-mentioned functions are a measure that is easily constructed using any phylogenetic tree.

## 6.1 Additional Examples and Discussion

For a first example, in the phylogeny in figure 3, I have superimposed hypothetical migratory paths onto the tree with two possible points of origin in mind. The point of origin of the left-hand component figure is posited to be point  $E$ , which necessitates a migratory history with two migratory chains. The first history starts at  $E$  and goes to  $D$ 's location, and then from  $D$ 's location to  $A$ 's, from  $A$ 's to  $B$ 's, and finally from  $B$ 's to  $C$ . This first history is traced out in green on the figure, where 'dead' chains are marked with a dashed line. A second migratory chain (traced in red) must then be invoked to explain the sequence of events leading to the upper half of the tree. This chain begins at  $E$ 's location again, and then moves on to  $F$ 's, then  $I$ 's, then  $H$ 's, and finally to  $G$ 's location. Under the model, the likelihood of this sequence of events is:

$$L_E^* = \frac{4^4 e^{-4}}{4!} \frac{4^4 e^{-4}}{4!} = \left(\frac{4^4}{4!}\right)^2 e^{-8}$$

This reflects the two required, non-degenerate migratory chains to explain the sequence of events. If we were to use divergence measure  $D_1$ , two chains are required,

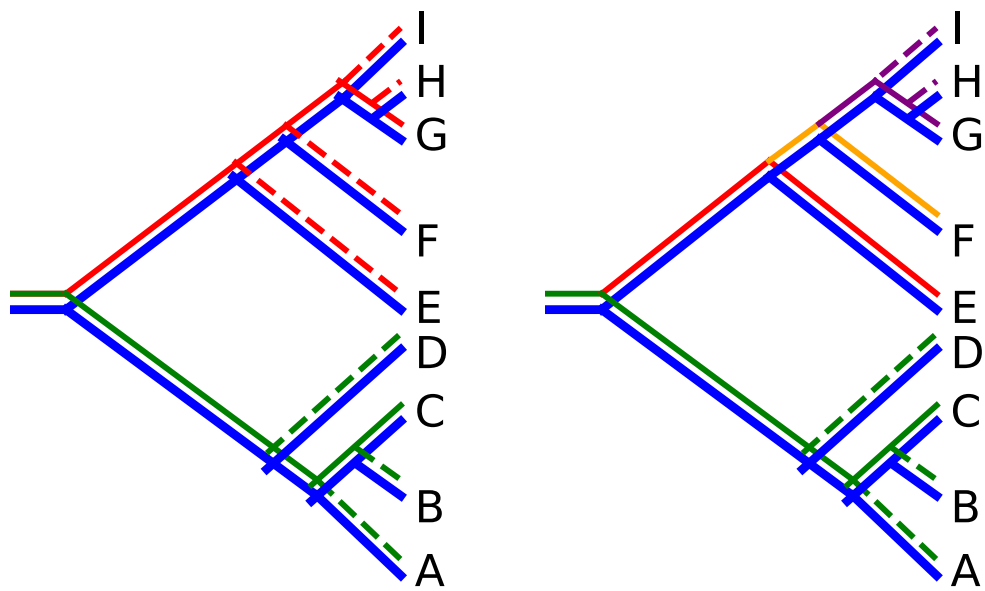


Figure 3: Two hypothetical migratory histories through a given Phylogenetic Tree. The picture on the left takes location  $E$  as the point of origin, while that on the right takes point  $I$  as the origin.



each of which has a maximal length of 4. Hence,  $D_E^1 = 4 - 2 = 2$  and  $D_E^2 = 4/2 = 2$ .

If  $I$ 's location is to be the point of origin, more migratory chains need to be invoked. The first chain is again traced out in green, mimicking the case in which  $E$  is the point of origin. But to explain how a group got to point  $E$  given  $I$  was the starting point, a single migratory chain consisting of a sole migratory event is needed. Also needed is another migratory chain to explain how a group got to point  $F$  from  $E$ . Finally, a fourth migratory chain going from  $I$ , and then to  $H$ , and finally on to  $G$  is needed. The likelihood of these events is then:

$$L_I^* = \frac{4^4 e^{-4}}{4!} \frac{1^1 e^{-1}}{1!} \frac{1^{e-1}}{1!} \frac{2^2 e^{-2}}{2!} = \frac{4^4 2^2}{4! 2!} e^{-8}$$

One finds that in this case, the measures of divergence are  $D_I^1 = 1$ , and  $D_I^2 = 0$ . One can see in these examples the roles that convexity and degeneracy plays in driving differences in likelihood. One can also observe that it is not sufficient to pick the location from which the longest chain can be assembled, as both  $E$  and  $I$  admit chains of length 4. If one wished to compare hypotheses about the two points, one might use the relative likelihood; in this case, the probability that  $E$  is the point of origin is  $L_E^*/(L_I^* + L_E^*) = .84$ . Of course, one would like to have a more rigorous method for computing likelihoods in every settings, and this is described in a subsequent section which presents a computational algorithm.

## 6.2 Known branch lengths

If branch lengths are known, no difficulties are posed for the basic ideas of the model. All that happens is that exponential densities replace the Poisson densities. The chief technical assumption that is required in this instance is that none of the branch lengths are too short. To illustrate how the model changes, return to figure 1. Now, with branch lengths treated as known, the likelihood is governed by a collection of exponential distributions:

$$P_A = \lambda_1^4 e^{-\lambda_1 T} \lambda_2^0 e^{-\lambda_2 t_6} \lambda_3^0 e^{-\lambda_3 t_7} \lambda_4^0 e^{-4t_8} \quad (7)$$

While there are still degenerate branches, in that the probability in equation 7 is maximized where  $\lambda_2 = \lambda_3 = \lambda_4 = 0$ , now  $\lambda_1^* = \frac{4}{T}$  maximizes the likelihood, so the profile likelihood becomes:

$$L_A = \left(\frac{4}{T}\right)^4 e^{-4} \quad (8)$$

The likelihood for  $C$  is:

$$L_C = \lambda_1^1 e^{-\lambda_1 T} \lambda_2^1 e^{-\lambda_2 (T-t_1)} \lambda_3^2 e^{-\lambda_3 (T-t_0-t_1)} \quad (9)$$

The profile likelihood associated with equation (9) is

$$\frac{1}{T} \frac{1}{T-t_1} \left( \frac{2}{T-t_1-T_2} \right)^2 \quad (10)$$

So, as before, we still have a basic race between a rapidly increasing denominator and a slower-decreasing denominator; in expression (10), the decreasing length of the branches spanned in the denominator fulfills the role of the factorial term in the denominator of the Poisson model. The danger for the age-area theorem is that the branches will be very small, which will lead to the denominator being very large. However, if we impose a minimum branch length on the branches, we get the same convexity of the concentrated function as before, so we have:

**Theorem 2** (Age-Area Theorem - known branch lengths). *Suppose that Assumptions 1 hold, and define Divergence as in definition 1. Also, suppose that branches are not too short. Then*

$$D_k \geq D_j \implies L_k \geq L_j$$

and in particular

$$i = \arg \max [D_1, D_2, D_3, \dots, D_K] \implies i = \arg \max [L_1, L_2, L_3, \dots, L_K]$$

*Proof.* Under the assumptions, the concentrated likelihood remains convex. Proceed as in the previous statement of the theorem.  $\square$

## 7 Microeconomic foundations

The previous sections have leaned on properties of the Exponential/Poisson distribution. What justifies the choice of this distribution? One would like a model in which migrations move forward probabilistically, but at the same time, the impetus for the

migration resets or takes on a different character once the migration moves on.

First, suppose that there is a discrete set of habitable locations, and suppose that a shock randomly occurs that applies to all currently unoccupied locations. This creates an abundance of resources at all unoccupied places throughout the system. The shock is such that a natural carrying-capacity parameter increases, but this extra carrying capacity can be exhausted. If the population ever reaches some barrier level  $\bar{b}$  at a particular location, the carrying capacity is exhausted and falls to a level immediately reverts to a level of  $\underline{K}$ . Once the population has adjusted, a new value of  $\bar{b}$  is drawn, a new shock emerges, and time continues on.

The key thing is that there are now  $\bar{b}$  people at a place where carrying capacity is  $\underline{K}$ . At this point in time, there are now a superabundance of people in the given area, and some mass of these people may migrate to a new location.<sup>6</sup>

According to an arbitrage condition, a segment of the population realizes that it could migrate to a new location where the birds are plentiful even if migration is costly, and finds it in its interests to do so given the localized collapse of the bird population. The population splits into a migratory group based on arbitrage, the migratory group randomly selects a new location, and emigrates.<sup>7</sup> After this emigration, the remaining population embarks on the discovery of a new resource.

To make these ideas concrete, and see how they explicitly relate to the Exponential distribution, imagine that income per capita depends upon a resource level and also the current number of inhabitants. That is, flow per capita income at a particular location is

$$y_t = f(n_t, \theta_t) \tag{11}$$

I imagine that  $f(n, \theta)$  is bounded above at some reasonable level. Individuals gain utility from consumption  $x_t$  and children  $k_t$  and children according to a Bernoulli utility function  $u = (x_t k_t)^{\frac{1}{2}}$ . Raising a child costs  $\kappa$  units of the consumption good, so individuals face the budget constraint  $y_t = x_t + \kappa k_t$ . This leads to  $k_t^* = \frac{y_t}{2\kappa}$ , and  $x_t^* = \frac{y_t}{2}$ .

---

<sup>6</sup>A parable: a people currently occupy an island with a population of friendly birds, who also populate some unsettled islands. While plump, these birds taste terrible. By accident, one day it is discovered that a spice on the island makes the birds palatable, leading to an abundance of food. Human population adjusts to the newfound resource, and while the birds are plentiful, there are no problems. But one year, an unusually good agricultural crop pushes the human population higher than it usually is. The additional population strains the bird population leading to a collapse. Half of the population leaves for a new island, where birds are still plentiful.

<sup>7</sup>The location does not actually have to be random. The minimum cost reachable location is in fact the best, as is discussed in the section on model extensions.

Equilibrium population growth is  $n_t k_t^*$  and equilibrium indirect utility is  $v(y, \kappa) = \frac{y_t}{2\kappa}$

A ramification of this is that when income is above some subsistence level, population increases proportionally, according to a Malthusian model of population growth in which people have the optimal number of children. Normalizing the subsistence level to unity. Then, in a given instant of time, current population creates future population according to the following relationship:

$$n_{t+\Delta} = n_t f(n_t, \epsilon_{t+\Delta} - \epsilon_t, \Delta) + n_t(1 - \delta\Delta)$$

Suppose that  $f(n_t, \epsilon_t, \epsilon_{t+\Delta})$  can be captured by a first-order Taylor expansion so we get:

$$n_{t+\Delta} = n_t (\bar{f} - f_1 n_t - \delta)\Delta + f_2(\epsilon_{t+\Delta} - \epsilon_t)n_t + n_t$$

This can be rewritten as:

$$n_{t+\Delta} - n_t = n_t(\bar{f} - \delta - f_1 n_t) + f_2(\epsilon_{t+\Delta} - \epsilon_t)n_t$$

Letting  $\Delta$  go to zero, and assuming that  $\epsilon$  is governed by a standard Brownian motion, the above can be rewritten as a stochastic differential equation:

$$dn = n(\bar{f} - \delta + f_1 n) + f_2^2 n^2 dz$$

Re-parameterizing so that  $\bar{f} - \delta = r$ ,  $f_1 = \frac{r}{K}$ , and  $f_2 = \sigma$ . The result is then a stochastic logistic population growth model:

$$dn = rn \left(1 - \frac{n}{K}\right) + \sigma^2 n^2 dz \tag{12}$$

Here, our carrying-capacity parameter  $K$  is where the shock can be built into the model. That is, once a shock occurs,  $K = \bar{K}$  across available locations. But once population approaches some value close to the carrying capacity,  $\rho K$ , the carrying capacity crashes to  $\underline{K}$ .<sup>8</sup> Once the stressing population has moved on, a new value for the carrying capacity is drawn and the process starts anew.

The exact parameterization in terms of a stochastic logistic growth model is not

---

<sup>8</sup>Details need to be worked out, but this concept works. A fixed number of people to have a viable migration, along with assumptions on how large  $K$  is, should be sufficient.

the critical fact of the matter. What is crucial is the mean-reverting nature of the stochastic differential equation (7). Ricciardi et al. (1999) show that the above process has a time-independent, initial-condition-independent stationary distribution given by:

$$W(n) = \left( \Gamma \left[ \frac{2r}{\sigma^2} - 1 \right] \right)^{-1} \left( \frac{2r}{K\sigma^2} \right)^{\frac{2r}{\sigma^2} - 1} x^{\frac{2r}{\sigma^2} - 2} \exp \left[ \frac{2rx}{K\sigma^2} \right]$$

As Ricciardi et al. (1999) and Nobile et al. (1985) show, the existence of a time-independent steady state-density implies that the first passage time to a ‘large’ barrier is approximately exponential. That is, let  $g(\bar{b}, t|y)$  denote the distribution of first passage times to a barrier  $\bar{b}$ . Then:

$$g(\bar{b}, t|y) \sim \frac{1}{t_1(\bar{b}|y)} \exp \left( -\frac{t}{t_1(\bar{b}|y)} \right)$$

Where  $t_1(\bar{b}|y)$  is the mean first passage time corresponding with the distribution. That is, the distribution of times to which the barrier is attained is approximately exponential. To review, the Exponential/Poisson distribution foundations for the model can now be justified by the following back story:

1. Any migratory chain has implies that unoccupied locations have a carrying capacity  $K$  and an upper barrier  $\bar{b}$ .
2. When a population arrives at the new location, it evolves according to the given model, so that the time at which it hits the barrier is approximately exponential according to the above density.
3. Once the barrier population is attained, the carrying capacity crashes until a segment of the population leaves for a new location.

The above sequence of events leads to a local superabundance of population, which finds it in its interests to divide and find a new place to live. The idea and its inner workings are shown in figure 4.

<sup>9</sup> The result of these ideas is something like figure 5. The figure shows mean-reverting processes eventually hitting an upper limit, which resets the process and produces a migratory event. This sequence of events continues at the next location. The idea behind the process is illustrated in figure 5.

---

<sup>9</sup>This has yet to be shown and this part of the paper is weak.

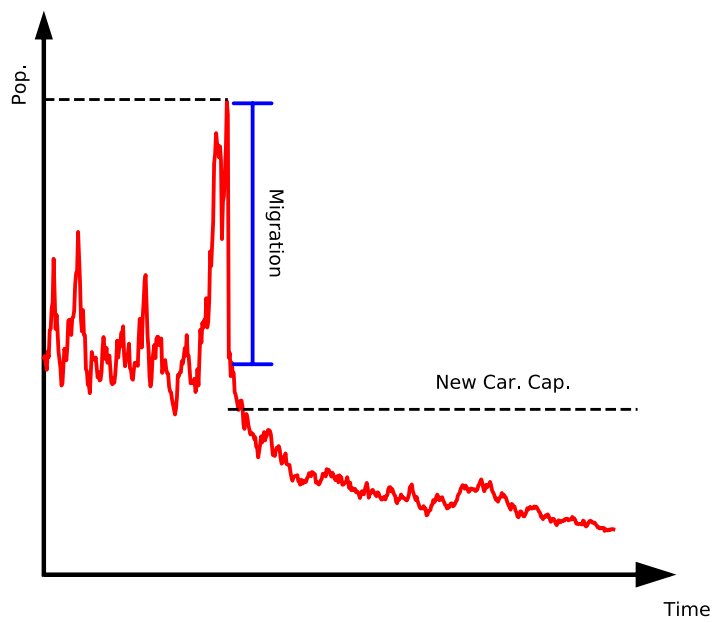


Figure 4: An illustration of the hitting process, in which a new carrying capacity is drawn after the migration.

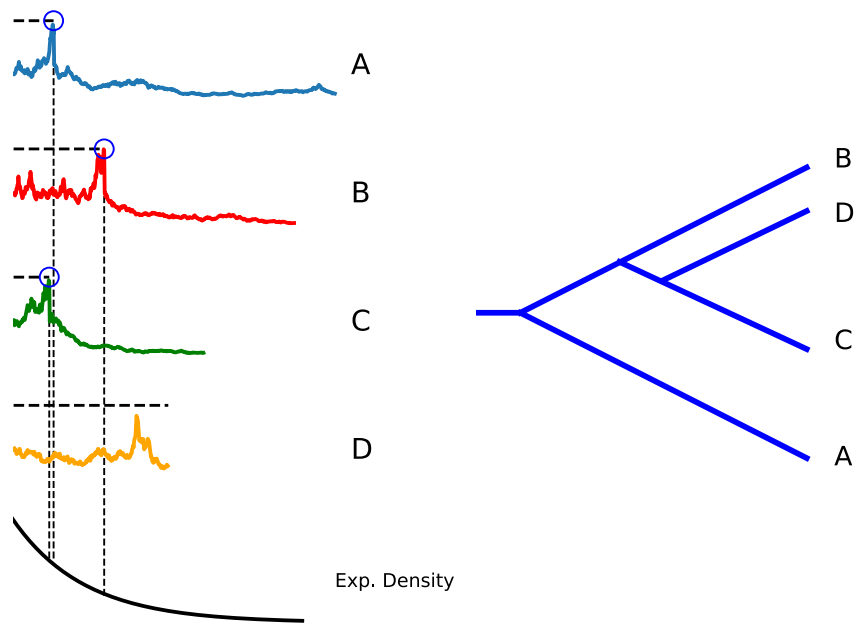


Figure 5: An illustration of the formation of a phylogenetic tree. The left-hand side of the figure denotes times at which the barrier is hit, at which point a segment of the population leaves for a new area. Note that the length of the tree corresponds with the length of the path starting at *A*, which is also the same as the sums of the other branches. Also note the exponential density of migration times.

## 8 Computation

In this section I show how one can couple choice of the optimal path through a tree by working recursively back through the tree, employing a dynamic programming algorithm to obtain the probability that all points in the tree were the geographic point of origin of the tree. This exercise also shows how one could expand the model to include different sorts of things in the likelihood, such as physical distance between locations. First, imagine that all interior nodes have been enumerated in depth-first order, so that nodes nearer to leaves or taxa carry lower indices. This allows a backwards traversal of the tree via the index. Both the trees in 1 and 3 have followed this convention. In particular, label interior nodes along the tree  $v = 1, 2, 3, \dots, V$ , with terminal branches carrying their ‘names’  $k = 1, 2, 3, \dots, K$ . ?

I can now recursively develop an expression for the probability that location  $k$  is the origin for the family of cultures after having traversed the  $v$ th node, and once the root node is reached, there will be a full complement of probabilities associated with the hypothesis that  $k$  was the point of origin. I begin by treating the case in which branch lengths are known, as the case in which they are not known can be handled by appropriate normalization of branch lengths. Let  $L_{v,k}^*$  denote the likelihood that  $k$  is the point of origin after considering the subtree spanned by node  $v$ . This is the most likely migratory history emanating from group  $k$ ’s location, given the subtree starting at  $v$ . Let  $t_{kv}$  denote the length of a chain starting at node  $v$  given it emanated from  $k$ ’s location.

A useful idea is the *descendants* of node  $v$ . These are all the terminal nodes that can be reached from a directed path starting at node  $v$  proceeding to the roots. A related useful function is the function  $A(i, v)$  - which I define to be all of the descendant nodes that were not reachable from  $k$  until node  $v$  is considered.

At  $v = 0$ , the algorithm starts with  $L_{k,0} = 1$ , and the two state variables  $t_{k,0} = t_i$ , and  $n_{ik} = 0$ , which are, respectively, the length of a branch and the number of jumps given a migratory chain starting at  $k$ . Now, define:

$$t_{i,k} = t_k + t_j^*, \quad n_{i,k} = n_j, \quad [t_j^*, n_j] = \arg \max_{j \in A(i,k)} \{t_j : j, f(t_j + t_k, n_j + 1) + LL_{j,k-1}\} \quad (13)$$



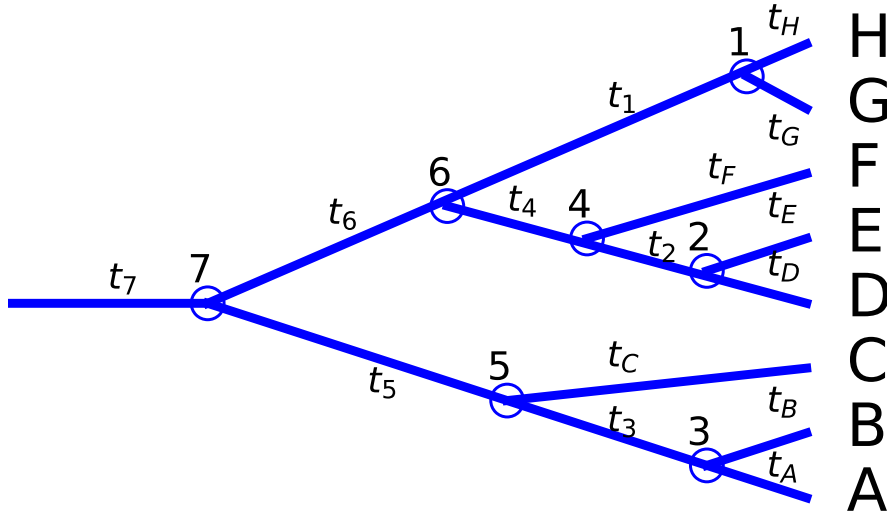


Figure 6: An illustration of the formation of a phylogenetic tree. The left-hand side of the figure denotes times at which the barrier is hit, triggering mass migrations. Note that the length of the tree corresponds with the length of the path starting at  $A$ , which is also the same as the sums of the other branches.

Further, define the conditional log-likelihood as:

$$\ln L_{ik}^* = \ln L_{ik-1}^* + \ln(f(t_{ik}, n_{ik}))$$

### 8.1 An example

The algorithm begins by keeping a running tally of the log-likelihood for each location, the length of a current migratory chain for each location, and the number of migratory events associated with the current migratory chain. I refer to these three quantities as  $LL_v$ ,  $t_v^*$ , and  $n_v$ .

Initially at  $v = 0$ , all values in the vectors  $LL_v$  and  $n_v$  are zero, with the vector  $t_v$  holding each culture's terminal branch, so  $t_{kv}^* = t_k$  to start. Consider first the three

interior nodes 1, 2, and 3. The first node involves groups  $H$  and  $G$  only. The longest chain that can be constructed from location  $H$  at node 1 has one migratory event, and spans  $t_1 + t_G$ . Similarly, for  $G$ , the longest chain has 1 event and spans  $t_1 + t_H$ . This logic can be repeated for nodes 2 and 3, so after the first three nodes, the is as depicted in table 1.

Culture	$t_3^*$	$n_3$	$LL_3$
H	$t_1 + t_G$	1	0
G	$t_1 + t_H$	1	0
F	$t_F$	0	0
E	$t_2 + t_D$	1	0
D	$t_2 + t_E$	1	0
C	$t_c$	0	0
B	$t_3 + t_A$	1	0
A	$t_3 + t_B$	1	0

Table 1: Algorithm at  $v = 3$

Next, consider nodes  $v = 4$  and  $v = 5$ . Node 4 has descendants  $D, E$  and  $F$ . While  $D$  and  $E$  require a migratory chain with a single event to explain how  $F$  was reached, There are two possibilities for explaining how  $D$  and  $E$  arrived in their locations given they started at  $F$ ; the chain must run from  $F$  to  $D$  to  $E$ , or from  $F$  to  $E$  to  $D$ . Suppose that for some reason  $D$  seems more likely to follow  $F$  than  $E$ . Following the algorithm, the existing information on  $t^*$  and  $n$  for  $D$  and  $E$  must be pushed into the likelihood vector, which start a new migratory chain for these two nodes that produces a jump to location  $F$ . Similar ideas apply for node  $v = 5$ . The situation after  $v = 5$  is depicted in table 2.

Culture	$t_3^*$	$n_3$	$LL_3$
H	$t_1 + t_G$	1	0
G	$t_1 + t_H$	1	0
F	$t_4 + t_2 + t_D$	2	0
E	$t_4 + t_F$	1	$-\ln(t_2 + t_E)$
D	$t_4 + t_F$	1	$-\ln(t_2 + t_E)$
C	$t_5 + t_3 + t_B$	2	0
B	$t_5 + t_C$	1	$-\ln(t_3 + t_A)$
A	$t_5 + t_C$	1	$-\ln(t_3 + t_B)$

Table 2: Algorithm at  $v = 5$

Groups  $D$  through  $H$  are all descendants of node 6, and hence an optimal decision must be taken by each of  $H$  and  $G$  as how migration proceeds through  $D, E$ , and  $F$ , while another decision must be made about how migration goes from each of  $D, E$ , and  $F$  through  $G$  and  $H$ . Table 3 suggests that the former proceeds from  $F$ , to  $E$ , to  $D$ , while the latter goes to  $G$  then to  $H$ .

Culture	$t_6^*$	$n_6$	$LL_6$
H	$t_6 + t_4 + t_2 + t_D$	1	$-\ln(t_1 + t_G)$
G	$t_6 + t_4 + t_2 + t_D$	1	$-\ln(t_1 + t_H)$
F	$t_6 + t_1 + t_H$	2	$2(\ln 2 - \ln(t_4 + t_2 + t_D))$
E	$t_6 + t_1 + t_H$	1	$-\ln(t_4 + t_F) - \ln(t_2 + t_D)$
D	$t_6 + t_1 + t_H$	1	$-\ln(t_4 + t_F) - \ln(t_2 + t_E)$
C	$t_5 + t_3 + t_B$	2	0
B	$t_5 + t_C$	1	$-\ln(t_3 + t_A)$
A	$t_5 + t_C$	1	$-\ln(t_3 + t_B)$

Table 3: Algorithm at  $k = 6$

The final node,  $v = 7$  requires further decisions to be made about how migrations proceed. A configuration of the quantities used in the algorithm is show in table 4:

Culture	$t_7^*$	$n_7$	$LL_7$
H	$t_7 + t_5 + t_3 + t_B$	3	$3(\ln 3 - \ln(t_6 + t_4 + t_2 + t_D)) - \ln(t_1 + t_G)$
G	$t_7 + t_5 + t_3 + t_B$	3	$3(\ln 3 - \ln(t_6 + t_4 + t_2 + t_D)) - \ln(t_1 + t_H)$
F	$t_7 + t_5 + t_3 + t_B$	3	$2(\ln 2 - \ln(t_6 + t_1 + t_H)) + 2(\ln 2 - \ln(t_4 + t_2 + t_D))$
E	$t_7 + t_5 + t_3 + t_B$	3	$2(\ln 2 - \ln(t_6 + t_1 + t_H)) - \ln(t_4 + t_F) - \ln(t_2 + t_D)$
D	$t_7 + t_5 + t_3 + t_B$	3	$2(\ln 2 - \ln(t_6 + t_1 + t_H)) - \ln(t_4 + t_F) - \ln(t_2 + t_E)$
C	$t_7 + t_6 + t_4 + t_2 + t_E$	4	$2(\ln 2 - \ln(t_5 + t_3 + t_B))$
B	$t_7 + t_6 + t_4 + t_2 + t_E$	4	$-\ln(t_5 + t_C) - \ln(t_3 + t_A)$
A	$t_7 + t_6 + t_4 + t_2 + t_E$	4	$-\ln(t_5 + t_C) - \ln(t_3 + t_B)$

Table 4: Algorithm at  $k = 7$

At this point, all nodes of the tree have been traversed, so, the remaining task is to collect the state variables into the likelihood. The results of performing this operation are shown on table 5.

One can see by inspecting the table which group seems to be the likeliest to be the point of origin. It is group  $C$ . While the example has branch lengths, if one were to replace these with factorials, one would see that for group  $C$ , the divergence measure

Culture	Log-likelihood
H	$3(\ln(3) - \ln(t_7 + t_5 + t_3 + t_B)) + 3(\ln 3 - \ln(t_6 + t_4 + t_2 + t_D)) - \ln(t_1 + t_G)$
G	$3(\ln(3) - \ln(t_7 + t_5 + t_3 + t_B)) + 3(\ln 3 - \ln(t_6 + t_4 + t_2 + t_D)) - \ln(t_1 + t_H)$
F	$3(\ln(3) - \ln(t_7 + t_5 + t_3 + t_B)) + 2(\ln 2 - \ln(t_6 + t_1 + t_H)) + 2(\ln 2 - \ln(t_4 + t_2 + t_D))$
E	$3(\ln(3) - \ln(t_7 + t_5 + t_3 + t_B)) + 2(\ln 2 - \ln(t_6 + t_1 + t_H)) - \ln(t_4 + t_F) - \ln(t_2 + t_D)$
D	$3(\ln(3) - \ln(t_7 + t_5 + t_3 + t_B)) + 2(\ln 2 - \ln(t_6 + t_1 + t_H)) - \ln(t_4 + t_F) - \ln(t_2 + t_E)$
C	$4(\ln(4) - \ln(t_7 + t_6 + t_4 + t_2 + t_E)) + 2(\ln 2 - \ln(t_5 + t_3 + t_B))$
B	$4(\ln(4) - \ln(t_7 + t_6 + t_4 + t_2 + t_E)) - \ln(t_5 + t_C) - \ln(t_3 + t_A)$
A	$4(\ln(4) - \ln(t_7 + t_6 + t_4 + t_2 + t_E)) - \ln(t_5 + t_C) - \ln(t_3 + t_B)$

Table 5: Algorithm at conclusion, known branch lengths. Poisson model can be recovered by replacing time terms with factorials.

$D^1$  is  $D_C^1 = 4 - 2 = 3$ . For other candidates like  $B$ , for example, one finds  $D_C^1 = 4 - 3 = 1$ . If we were to replace the branch lengths with corresponding factorials (this can be done by counting terms in parentheses and then subtracting one), we would find that  $L_D = \frac{4^4}{4!} \frac{2^2}{2!} e^{-6}$ . A competitor that might be considered is  $H$ , which would have likelihood  $L_H = \frac{3^3}{3!} \frac{3^3}{3!} e^{-6}$ .  $L_D$  computes out to be 0.05288, while  $L_h$  computes out to be .05019, so, while the hypotheses are close,  $C$  is still the more likely point of origin.

## 9 Extensions

I outline some potential extensions to the basic methods, including making the algorithm take more information into account, approximating probabilities, and expanding the likelihood to include supplementary information.

### 9.1 Additional information

One can also modify the model to include explicit information about distance and space. One possibility is to do this in such a way so as to reflect that moving larger distances is less likely, and many analyses commonly include such spatial features.

[To be written]

### 9.2 Expanded likelihoods

One extension is to interpret the likelihood presented in the previous few pages as a conditional likelihood, and incorporate it into a joint estimation procedure. For exam-

ple, if one has a means of computing the likelihood of a particular tree, one can then think about the joint distribution of a likelihood and a tree at the same time using the simple probabilistic relationship:

$$P(\mathcal{T}, \mathcal{H}) = P(\mathcal{H}|\mathcal{T})P(\mathcal{T})$$

This is important because there are so many methods for computing the likelihood of a given linguistic tree. This allows one to consider a migratory route as part of the estimation process, rather than just something loosely implied by the structure of the tree. One can also form a picture of the distribution of trees and roots once one is equipped with a likelihood function.

Another important facet is the ability to include information for other sources, as in a Bayesian analysis.

[To be written]

## 10 Applications

### 10.1 Afroasiatic

[To be written]

### 10.2 Na Dene

[To be written]

Additional information and applications, including a Python implementation of the ideas in this paper, can be found on the project site: <http://github.com/mbaker21231/instevo>.

## 11 Conclusions

[To be written]

## References

- Ahlerup, P. and Olsson, O. (2012). The roots of ethnic diversity. *Journal of Economic Growth*, 17(2):71–102.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kerlat, S., and Wacziarg, R. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.
- Ashraf, Q. and Galor, O. (2013). The “out of africa” hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.
- Atkinson, Q. D. and Gray, R. D. (2003). Language-tree divergence times support the anatolian theory of indo-european origin. *Nature*, (426):435–439.
- Cavalli-Sforza, L. L. and Cavalli-Sforza, F. (1995). *The Great Human Diasporas: The History of Diversity and Evolution*. Perseus Books, Cambridge, MA.
- Dimmendaal, G. J. (2011). *Historical Linguistics and the comparative study of african Languages*. John Benjamins Publishing Company, Amsterdam and Philadelphia.
- Dolgopolsky, A. (1988). The indo-european homeland and lexical contacts of proto-indo-european with other languages. *Mediterranean Language Review*, 3:7–31.
- Dyen, I. (1956). Language distribution and migration theory. *Language*, 32(4):611–626.
- Ehret, C. (2002). *The Civilizations of Africa: A history to 1800*. University of Virginia Press, Charlottesville.
- Greenhill, S. and Gray, R. D. (2005). Testing population dispersal hypotheses: Pacific settlement, phylogenetic trees, and austronesian languages. In Mace, R., Holden, C. J., and Shennan, S., editors, *The evolution of cultural diversity: a phylogenetic approach*, chapter 3, pages 31–65. Left Coast Press.
- Latham, R. G. (1851). *The Ethnology of the British Colonies and Dependencies*. J. Van Voorst, London.
- Mace, R., Holden, C. J., and Shennan, S., editors (2005). *The Evolution of Cultural Diversity: A Phylogenetic Approach*. Left Coast Press, Walnut Creek, CA.

- Mallory, J. P. (1997). The homelands of the indo-europeans. In Blench, R. and Spriggs, M., editors, *Archaeology and Language 1: Theoretical and methodological orientations*, pages 93–121. Routledge, London.
- Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–1539.
- Nichols, J. (1997). Modelling ancient population structures and movements in linguistics. *annual review of anthropology*, 359-84:26.
- Nobile, A. G., Ricciardi, L. M., and Sacerdote, L. (1985). Exponential trends of first-passage-time densities for a class of diffusion processes with steady-state distribution. *Journal of Applied Probability*, 22(3):611–618.
- Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. Cape, London.
- Ricciardi, L. M., Crescenzo, A. D., Giorno, V., and Nobile, A. G. (1999). An outline of theoretical and algorithmic approaches to first passage time problems with applications to biological modeling. *Mathematica Japonica*, 50(2):247–322.
- Ruhlen, M. (1994). *The Origin of Language: Tracing the Evolution of the Mother Tongue*. John Wiley and Sons, Inc., New York, New York.
- Sapir, E. (1916). *Time Perspective in aboriginal American culture: A study in method*. Number 19 in Geological Survey, Memoir 90, Anthropological Serie. Government Printing Bureau, Ottawa.
- Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *The Quarterly Journal of Economics*, 124(2):469–529.
- Spolaore, E. and Wacziarg, R. (2013). How deep are the roots of economic development. *Journal of Economic Literature*, 51(2):325–369.
- Trask, R. L. (2000). *The Dictionary of Historical and Comparative Linguistics*. Fitzroy Dearborn Publishers, Chicago and London.