

Foundations of the Age-Area Hypothesis

Matthew J. Baker

Department of Economics

Hunter College and the Graduate Center, CUNY

February 28, 2018

Abstract

Historical mass migrations have played a seminal role in shaping the global geographical distribution of cultures and societies. A commonly used tool in theorizing about the nature and origins of mass migrations is the so-called Age-Area Hypothesis, originally advanced by Sapir (1915). The age-area hypothesis traces the origins of linguistic stocks, which are closely related to cultural and genetic similarities. The Hypothesis posits that the origin point of a linguistic/cultural stock is most likely where the languages comprising the stock are most divergent. While the hypothesis is compelling, and its predictions are often corroborated by other evidence, a cohesive theoretical structure for the Age-Area Hypothesis has never been described. In this paper I describe a micro-founded model of the Age-Area Hypothesis, and render this into a means of calculating the probability of points of cultural origination for a given linguistic stock. I also present a computational algorithm for computing the relative likelihood of historical locations and migratory paths. Applications and extensions ensue.

The Age-Area Hypothesis - evidently first developed and applied by Sapir (1916) in his study of Native American Languages and cultures- is often invoked when linguistic evidence is used in support of hypotheses about where different sorts of cultures originated, how they evolved over time, and how they came to be located where they are today.¹ Briefly stated, the hypothesis says that the geographic area where a language phylogeny originated is most likely the place where the component languages of the phylogeny are most divergent, or maximally differentiated.

While the question may have some of academic interest only, a wide variety of recent economic literature highlights the importance of understanding how we arrived at the current cultural mix around the globe. Coincident with this understanding is the realization that understanding where different cultures came from, and how they interacted in the past, is also of paramount importance.

Examples of the application of the linguistic age-area hypothesis abound. Sapir himself used these arguments to suggest that the so-called Na-Dene linguistic grouping originated in and around the North West Coast. As another example, Atkinson and Gray (2003) have used the results to suggest that the Indo-European languages originated in Anatolia, not, as is sometimes supposed, on the steppes of Siberia. Further applications of the idea - either implicit or explicit use of the idea - are almost too numerous to mention. Ruhlen (2001) uses the age-area hypothesis to lend insight into debates about the origins of the Na-Dene, as well as the origins of the Bantu linguistic group in Africa. More specifically, Darden (2001) uses the AAH extensively to buttress his picture of Ancient African History.

What is the existing theoretical basis for the age-area hypothesis? In spite of its wide range of applications and deep implications for human history, there is, to my knowledge, no theoretical basis for the hypothesis. In population genetics, it is well-known that the point of maximal genetic diversity of populations is closest to the geographical point of origin. For example, it is known that the domestic cat originated in Egypt because the genetic pool is deepest and most varied in Egypt. Migrant populations only carry off a subset of the genes of the parent population and hence are less diverse in a well-defined sense.

These principles certainly do not apply to language and linguistic drift. While there

¹One must be careful in invoking “the” Age-Area Hypothesis. As I will elaborate below, this paper concerns what might be called the Linguistic Age-Area Hypothesis - not the related Age-Area Hypothesis that posits the wider the distribution of a cultural trait, the older it likely is.

are many striking parallels between the evolution of language and population genetics, they do not apply in this case. A migratory group does not carry off only a fraction of the grammar, vocabulary, or language structure of its parent!

The lack of a theoretical basis for the age area hypothesis is also a problem for its use in technical situations. For example, suppose one wished to include an ancient migratory path in a statistical model of migration? Suppose one wanted to use a migratory path as part of a larger statistical model? Since the hypothesis has no theoretical underpinnings, it is difficult to see how to construct a likelihood for a certain path or migratory history.

In this paper, I develop a theoretical foundation for the age-area hypothesis, and make some of the terms supporting the hypothesis more rigorous. The model is micro-founded in the sense that it can be based on microeconomic principles and an associated back story.

The theory can then be rendered in such a fashion that it turns out with a probabilistic notion of parsimony and Occam's razor; that is, when adapted in a particular fashion, migratory paths that originate at deeper points in a phylogenetic tree can be thought of as simpler in a well-structured way. These ideas lead to a particular definition of divergence or dissimilarity for each constituent culture in the phylogeny.

Assembling the aforementioned methods, I then develop an Age-Area Theorem, which says that a culture that is more linguistically divergent from the others in the stock is also more likely to be the point of origin of the stock. I then present a computational algorithm which relies upon backwards traversal of the tree, along with some extensions of the model. I conclude by discussing some applications.

1 Problem Description and Motivation

Consider the hypothetical phylogenetic tree displayed in figure 3. The figure depicts the genetic relationship between five related cultures. In the usual lingo, A is the 'most divergent,' with B, C, D, E all being more closely related. 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 language stock.

The Age-Area hypothesis (henceforth the AAH) concerns the location in which this

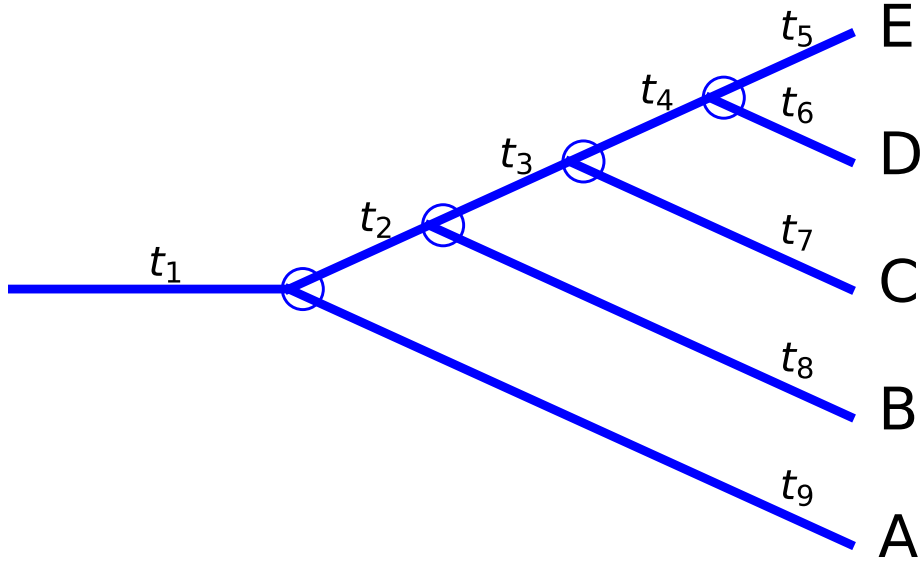


Figure 1: A phylogenetic tree

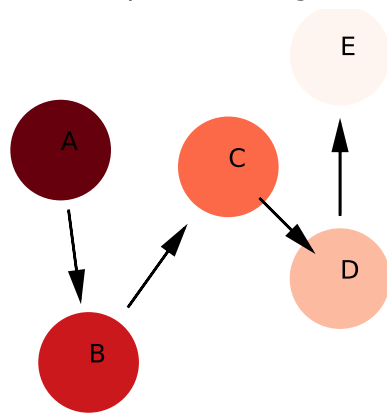
group of societies originated, and the times and sequences behind how they subsequently migrated to populate their current locations. The AAH posits that the point of origin was A's location, as A is the most divergent language from the group. If one wanted to push the hypothesis a bit further, or even apply it recursively, one would come up with a hypothetical migratory route: the hypothesis would lead one to believe that the point of origin of the stock was A, at which point there was a migration from A to B, then from B to C, and then from C to either D or E.

There are a large number of other possibilities as well, even though the structure of the tree constrains some sorts of events. For example, an additional plausible 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. The AAH suggests that the second sequence is a less likely migratory history than the sequence. Why? For a variety of reasons - one might simply appeal to Occam's razor - the events on

Two Hypothetical Migratory Routes

A is point of origin



C is point of origin

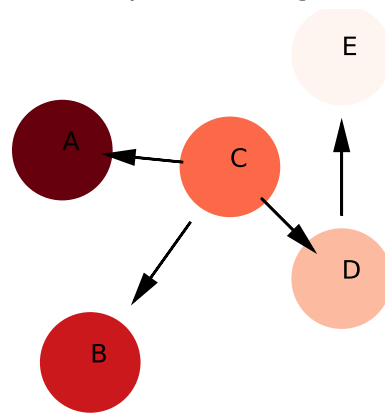


Figure 2: Potential migratory routes among a language stock consisting of five groups, following the Phylogenetic Tree in figure.

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. We can now pose our question as follows: if we believe these migratory events to be rare, and we wish to conserve them in explaining historical migrations, what sort of model would accomplish this? But can one make this operational in a more formal setting? Can one characterize parsimony in a meaningful mathematical and probabilistic way?

2 A model

In addition to providing a foundation for the age-area hypothesis, we might lean on some additional principles that we wish to encapsulate in the theory. We might like the model to conserve on estimation of parameters, and we might also like the model to have a consistent internal logic. A simple model with transparent machinery also creates the possibility of expansion and further development. While the assumptions listed below are restrictive, one can see how they might be relaxed in future work.

2.1 Tree Basics

Along these lines, assume that we observe the entire phylogenetic tree, and that there are no unobserved migratory events. The tree is assumed to be a full, rooted, binary tree, meaning that there is a single origin node and branch, and all other nodes have a single parent. All the interior nodes have two children, and all terminal nodes (leaves) have zero children.

The most important fact about full binary trees that we employ is that if a binary tree has n leaves, then it has $n - 1$ internal nodes. The tree in 3 has five terminal nodes and four internal nodes. Importantly, this also applies to subtrees, which means that if a node has k ancestors, then it is followed by $k - 1$ nodes.

2.2 Migratory Chains

We flesh out a model by introducing a *migratory chain*. Roughly speaking, a migratory chain is defined as a directed path from the root of the tree to an end node, which, once traversed, divides the tree into subtrees. Positing 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. For example, consider figure

1. Migratory chains are unique in the sense that each is governed by its own parameters.
2. A migratory chain can begin with any location that is currently occupied.
3. The location where the chain appeared generates a mass-migration at time t according to an exponential distribution with parameter λ .
4. The group emitting the mass-migration loses its propensity to migrate. This propensity to migrate is carried off with the mass-migrating group to its new location.

Some additional terminology. We will call a mass-migration history originating at location i \mathcal{H}_i , and we will consider only mass-migration histories requiring the minimal number of mass-migrations. As a practical matter, this means that each history will have exactly $n - 1$ migrations. Each location for our terminal nodes $i = 1, 2, 3, \dots, I$ then will come along with a \mathcal{H}_i . We will refer to the number of migratory chains corresponding with \mathcal{H}_i as C_i .

The principal observation of forming a model in this fashion is now to use the ideas in definitions 1 to develop a probability corresponding with a history of migratory chains, and to relate this to some notion of distance between societies to prove an Age Area Theorem, which will come in the form of relationships between \mathcal{H} , \mathcal{C} , and \mathcal{P} , the probability that point i is the point of origin for the societies in I .

To preview the basic logic, I develop a comparison of the two scenarios described in figure 2 within the confines of the model. We will 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. But we will start as treating these points as known and then “integrate them out.” The first scenario requires an initial migratory event to start after time t_0 from point A . At this point, according to the model assumptions, A is dormant, and the migratory bug is transitioned to location B . After time t_1 , B emits the migratory bug to C and is then dormant, followed by C ’s emission and subsequent dormancy, and then D ’s. Finally, E retains the bug but never emits. The result of this is that over the entire time span of the tree, there have been exactly 4 migratory events

through the one necessary migratory chain. These probability of this happening can then be written as

$$P_a = \frac{(\lambda T)^4 e^{-4T}}{4!} \quad (1)$$

Which is labelled as P_a because it is effectively the probability of this one migratory chain occurring.² The log-likelihood associated with the above is:

$$\ln P_a = 4 \ln(\lambda T) - 4T - \ln(4!) \quad (2)$$

Differentiating 2 with respect to λ , and solving gives $\lambda^* = \frac{4}{T}$, and substituting this back into the objective function gives the concentrated or profile likelihood:

$$P_a = \frac{4^4 e^{-4}}{4!} \quad (3)$$

There are a couple of points of interest about equation (3). First, since only one migratory chain is needed to explain the sequence of events, only one parameter is needed for the whole process. In this sense, this a parsimonious - it requires only one distribution and one parameter.

Contrast this with the case in which we posit that C is the initial location of the culture area. To maintain consistency with the linguistic or genetic analysis, we require 1) a migratory chain that starts at C and goes to A , 2) a migratory chain that starts at C and goes 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). According to the model, each of these chains requires its own Poisson/exponential distribution and parameter, which results in a likelihood of:

$$P_c = \frac{(\lambda_1 t)^1 e^{-\lambda_1 t}}{1!} \frac{(\lambda_2 (t - t_0))^1 e^{-1}}{1!} \frac{(\lambda_3 (t - t_0 - t_1))^2 e^{-\lambda_3 (t - t_0 - t_1)}}{2!} \quad (4)$$

Maximizing P_c in (5) with respect to λ_1, λ_2 and λ_3 , and substituting the result back into the right-hand side of (5) gives:

$$P_c = \frac{1^1 e^{-1}}{1!} \frac{1^1 e^{-1}}{1!} \frac{2^2 e^{-2}}{2!} = \frac{2^2 e^{-4}}{2!} \quad (5)$$

According to this model, we now see that the relative likelihood that A is the

²Note that there is a slight abuse of notation here. One might argue that we should include the probability of migratory events in general occurring, or write the above as a conditional probability where we are really talking about $P(a|N_m = 1)$.

point of origin relative to C depends upon $P_A/(P_a + P_c)$, which, roughly put is a race between the functions $\frac{4^4}{4!}$ and $\frac{2^2}{2!}$. If for some reason we had restricted things to these two possibilities, the relative probability that A was the point of origin would be 84%.

Generically speaking, the function

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

is convex in n , which implies breaking up a migratory chain into two smaller chains results in a less believable or likely explanation. That is, for any $k \in (1, n - 1)$, it is true that:

$$a(n) > a(n - k)a(k)$$

In context, this means that breaking up a migratory chain into migratory chains results in an explanation of events that is always less likely. There are a few details that need to be worked out, but the above example illustrates most of the inner workings of the model. The biggest detail that needs to be worked out is how the above theory that longer migratory chains are more parsimonious and therefore more likely relates to a measure of linguistic divergence. This task and others is taken up in the next section.

3 Age-Area Theorem

There are many measures of divergence, and the exact approach one takes to measuring divergence of language are irrelevant here, although to fix ideas it is useful to present a brief list of popular approaches. Some are more complex than others, and some involve glottochronology, lexicostatistics, etc. Ultimately, we take as given a full, rooted, binary tree, \mathcal{T} , which has N terminal nodes which represent our $i = 1, 2, 3, \dots, N$ cultural groups. As was established in the previous section, \mathcal{T} has $N - 1$ internal nodes and branches.

Along with this tree, we imagine that there is some dissimilarity matrix surrounding the constituents \mathcal{T} , in the form of pairwise distance. That is:

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \dots & d_{1N} \\ d_{21} & 0 & d_{23} & \dots & d_{2N} \\ \vdots & & \ddots & & \\ d_{N1} & & & & 0 \end{bmatrix}$$

To be clear, this is a linguistic distance matrix, not a physical distance. Also, define average to be the distance of a given society from all others; this is equivalent to computing the row sum in the above matrix and dividing by $n - 1$. So, we write:

$$D_i = \frac{1}{N-1} \sum_{j \neq i} d_{ij}$$

Theorem 1 (Age-Area Theorem). *Suppose that the above model of chain migration as rare events holds and that phylogenetic distance is proportional to separation time. Then*

$$D_i \geq D_j \iff p_i \geq p_k$$

Proof. The basic of the proof stems from the fact that there is a relationship between the length of migratory chains, likelihood, and the average distance from a bunch of ancestors. In fact, this is obvious so let's move on. \square

Corollary 1.

$$p_i = \max[p_1, p_2, p_3, \dots] \iff D_i = \max[D_1, D_2, D_3, \dots, D_n]$$

Proof. Suppose not. Then p_i could be a maximal probability but not have maximal average distance. This would imply that there is a $D_j > D_i$, which would contradict the Age-Area Theorem. \square

4 Micro foundations

Pursuant to the theorem, one might ask: what sort of process might produce a model like this of migration? One can actually develop a model that creates random scarcity, which incites mass migration. The trick is developing a theory in such a way so that it preserves migrations. That is, one requires that migrations move forward probabilistically and stops. Here is one way of doing this.

Imagine that there are a large, nonbinding set of viable locations in the migratory universe. At some an initial point, a new event arrives. This creates a system-wide increase in resources.

At a particular point in the universe of places, imagine we imagine that income per capita depends upon a resource and also the current number of inhabitants. That is,

flow per capita income at a particular location is

$$y_t = f(n_t, \theta_t) \quad (6)$$

Suppose that when income is above some subsistence level s population increases proportionally. (This can be derived from a standard model). Normalize the subsistence level to unity. Then, in a given instant of time, current population creates additional population so long as there is additional income. That is, we have:

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

Let's further suppose that $f(n_t, \epsilon_t, \epsilon_{t+\Delta})$ can be captured by a first-order Taylor expansion so we get:

$$n_{t+\Delta} = n_t (\bar{f} + f_1 n_t - s + f_2(\epsilon_{t+\Delta} - \epsilon_t)) - 1)\Delta + n_t$$

If we let Δ go to zero, and suppose that ϵ is governed by a standard brownian motion, we can write the above as a stochastic differential equation of the form:

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

We can reparameterize this by supposing that $\bar{f} = r$, $f_1 = \frac{r}{K}$, and $f_2 = \sigma$. The result is then a stochastic logistic population growth model:

$$dn = rn \left(1 - \frac{n}{K}\right) + \sigma^2 n^2 dz \quad (7)$$

The exact parameterization of the model is not necessary; what is important is that the mean-reverting nature of the SDE in equation (4) has a time-independent, stationary distribution, as Ricciardi et al. (1999) note. This distribution is given by:

$$W(n) = \Gamma \left[\frac{2a}{\sigma^2} - 1 \right]^{-1} \left(\frac{2b}{\sigma^2} \right)^{\frac{2a}{\sigma^2} - 1} x^{\frac{2a}{\sigma^2} - 2} \exp \left[\frac{2bx}{\sigma^2} \right]$$

Note to self. σ has the same meaning, and $a = r$, while $b = r/K$. If we plug this in, the above gets quite simple.

As Ricciardi et al. (1999) and ? show, the existence of a time-independent steady state density implies that the time to hitting a remote barrier is approximately exponential: that is:

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

Armed with these facts, we can now create a version of the model that produces migratory chains that are approximately exponentially distribution. So, consider the following back story for the model.

1. The landscape is uniform in that all locations have the same initial carrying capacity K_0 .
2. A new migratory chains raises the carrying capacity across all locations to $K_1 > K_0$.
3. There is a catch; if the population should ever reach the high boundary S , the carrying capacity immediately reverts to K_0 .

The above sequence of events lead to a temporal population surplus. Suppose that a mass-migration requires a certain level of population \bar{p} without which it will not succeed. The

s is often the case when dealing with estimation over trees, it behooves one to consider an algorithm that works backwards from the branches to the base. The algorithm presented here is essentially a pruning algorithm like Felsenstein's with a hint of dynamic programming. The dynamic programming aspect introduces a sort of continuation maximum likelihood into the tree.

The algorithm works as follows. At termination, the result is a concentrated log-likelihood of each location being the point of origin. The following is an illustration showing how the algorithm works on the example given above.

0. Initialization: a "active" vector of zeros $A = [0, 0, 0, 0, 0]$, a vector of exponential parameters that start with zero, $E = [0, 0, 0, 0, 0]$, and a continuation likelihood $L = [-t_4, -t_5, -t_6, -t_7, -t_8]$
1. Select an interior node and pick out its branches.
2. Poo I can add stuff again!

Gildner (1991) Dowd et al. (2014) Ricciardi et al. (1999)

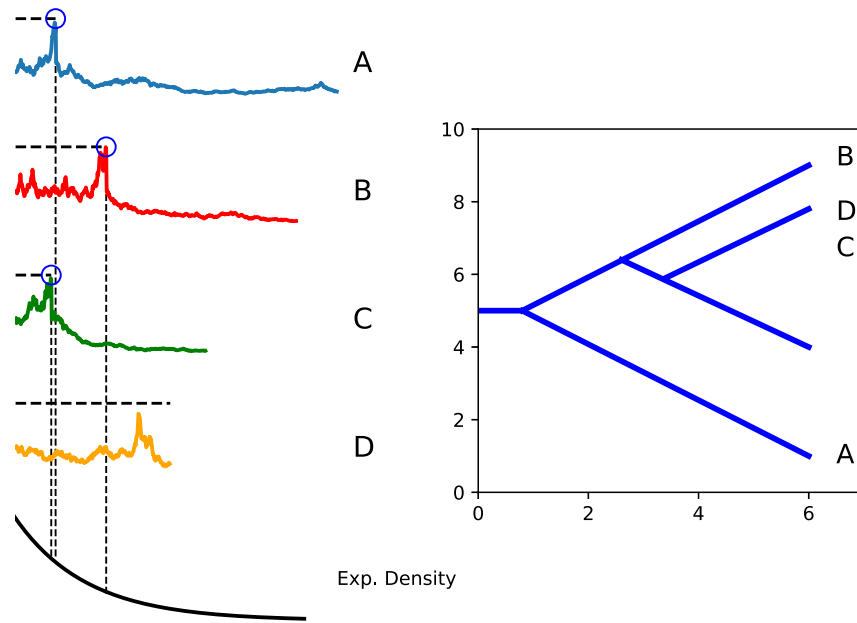


Figure 3: An illustration of the formation of a phylogenetic tree. The left-hand side of the figure denotes times at which the barrier is hit, triggering mass migrations. Note that the length of the tree corresponds with the length of the path starting at *A*, which is also the same as the sums of the other branches.

References

- Dowd, B. E., Greene, W. H., and Norton, E. C. (2014). Computation of standard errors. *Health Services Research*, 49(2):731–50.
- Gildner, G. M. (1991). *The Chickasaw Bayou Campaign (Master's Thesis)*. U. S. Army and General Staff College, Leavenworth, Kansas.
- 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.