University of Groningen

Johann Bernoulli Institute for Mathematics and Computer Science

Groningen Institute for Evolutionary Life Sciences

Simultaneous estimation and selection of species diversification models

Author:
Francisco Richter

Supervisors:
Prof. Dr. Ernst C. Wit
Prof. Dr. Rampal S. Etienne

May 13, 2016

Contents

Chapter 1

Introduction

> *Disasters are called natural, as if nature were the executioner and not the victim.*
>
> Eduardo Galeano,

Biodiversity, the term used to describe the wide variety of species on Earth, is declining at enormous rates due to human-induced environmental changes. The last time that Earth experienced such high rates of biodiversity loss in a relatively short time was at the end of the Cretaceous period, 65 millon years ago, then most dinosaurs went extinct [Wake and Vredenburg, 2008]. Biodiversity loss, in turn, compromises ecosystem stability and productivity, which negatively impacts the ecosystem services on which human communities depend [Tilman et al., 2006].

To conserve biodiversity, we must understand the mechanisms how it comes about and how it is maintained, in assemblages of species, so-called ecological communities. Novel genomic tools make it possible only now to measure and model species in great detail. In the last decade sophisticated methods have been developed that primarily aimed at inferring the evolutionary relatedness of the species (the phylogenetic tree) from DNA sequence data on the basis of models of species diversification. On figure 1 we can see the phylogenetic tree of the most complete analysis of avian species [Jetz et al., 2012].

About 1.7 million species have been identified and given scientific names, but only about $100,000$ of these are popular enough for taxonomist to know them well. There are estimated to be anywhere from 6 to 15 million species on Earth. Many new species are found each year; for instance, 361 new species (mostly insects) were found in the remote rainforest of Borneo from 1999 to 2004 [Chivian and Bernstein, 2008].

Mathematical and statistical tools have been crucial to generate an understanding on diversification processes [Gillman, 2009]. However there is still a lot to improve. Most of current models have some major shortcomings: they are either too simplistic,

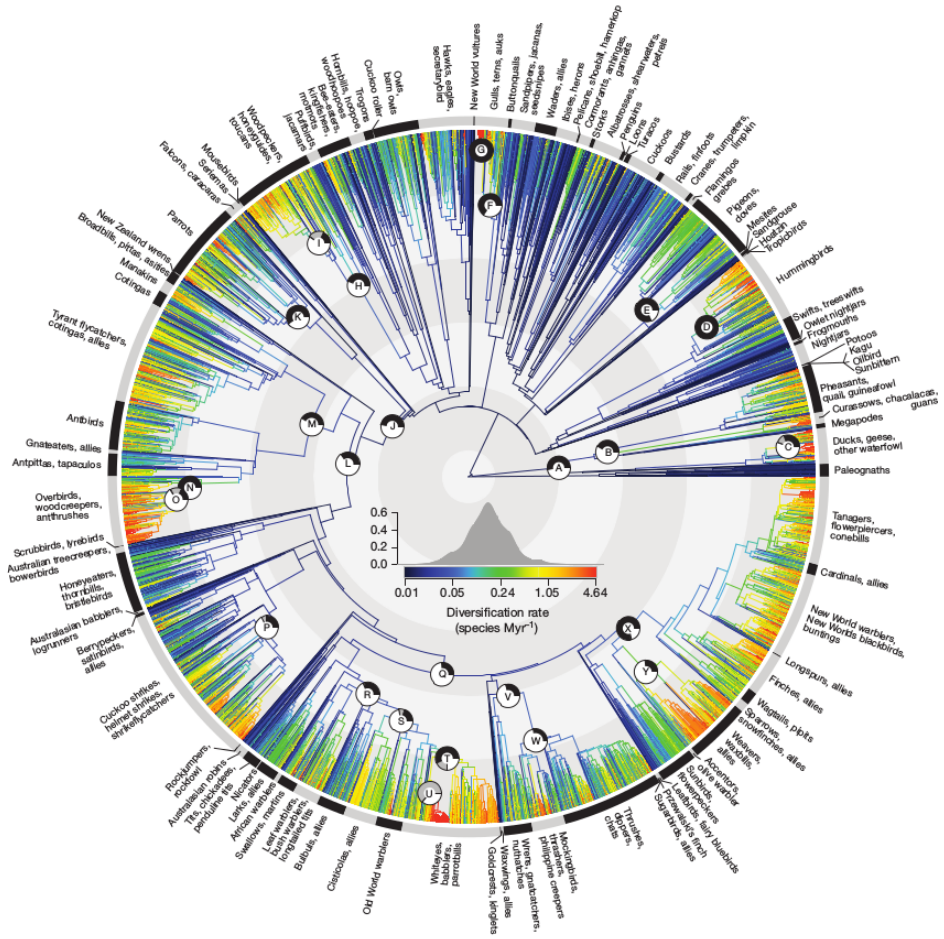

Figure 1.1: Diversification across the avian tree [Jetz et al., 2012]. Public phylogenetic information for avian species is available in this wonderful website http://birdtree.org/.

too specific or overly complex increasing the dimensionality of the system enormously.

Novel differential geometric approach to statistical methods have been successful implemented on biological data [Augugliaro et al., 2013][Abegaz and Wit, 2013], and give us a hope to reduce the inferential and computational complexity of the diversification scenario. In this project, we will consider models that account for local interactions and complex ecology information.

This manuscript has been written to be readable for both mathematicians and biologist, as well as a broader audience. Keeping that in mind, in chapter 2 we discuss an overview of the evolutionary diversification process and the ecological mechanism behind it. In chapter 3 we introduce the statistical modeling of the problem step by step. Finally, last tree chapters comment the knowledge utilization of this project, the timeline, and the bibliography respectively.

Chapter 2

Species Diversification

> *Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two facilities, which we may call intuition and ingenuity.*
>
> Alan Turing,

2.1 The evolutionary diversification process and the mechanism behind it. Review and Research Questions.

The mechanisms behind the variation in species diversity is one of the fundamental questions of evolutionary biology. Kendall [Kendall, 1948] considered a convenient birth-death process, which can be recast in terms of speciation and extinction events, and thus represents an extension of the Yule process [Yule, 1925]. Considering n_t as the number of species (of a related genus) at time t, the master equation for n_t in presence of extinction and speciation can be written as

$$\frac{dP(n_t)}{dt} = \lambda_{n_t-1}(n_t-1)P(n_t-1) + \mu_{n_t+1}(n_t+1)P(n_t+1) - (\mu_{n_t}+\lambda_{n_t})n_tP(n_t)$$

where λ_n and μ_n are the per capita speciation and extinction rates, respectively, and n_t the total number of species on moment t.

Based on this idea, the simplest and most widely applied of a variety of model is the random speciation-extinction process [Nee et al., 1994]. In a random speciation-extinction process, both speciation and extinction have instantaneous probability, or rates (λ and μ) which determine the probabilities that a clade either splits or terminates within a given time interval. In a simple branching process, the expected number of a clade increases exponentially, that is

$$E(n_t) = n_0 e^{(\lambda-\mu)t} \tag{2.1}$$

Despite this elegance, this approach is not realistic, evolutionary history of species is widely more complex:

- Each branch in the growing tree might not have the same potential fate to speciate or to get extinct. In fact, in many cases depends on several characteristics of the species.
- Ecological interactions are important.
- Geography, location and climate influence diversification of species.
- Evolution is likely to depend on diversity of species.
- Etcetera.

All this ecological factors needs to be considered if we want to build model consistent with real phylogenies; for example, typically we see a common slow down on diversification towards the present [Moen and Morlon, 2014] (but see [Jetz et al., 2012]) which is the opposite equation 1 predicts. Some biological explanations might underlie diversification slowdowns, like the protracted speciation model [Etienne and Rosindell, 2012], but those models do not have, for instance, a geographical component.

Spatial-temporal factors as well as traits and diversity dependence are some of the ecological information we should include to consider realistic models, however, this complex scenario increase the dimensionality of the system enormously, rendering them practically intractable.

On next chapter we will describe a novel methodological contribution to this project which aims to produce a computational feasible and consistent model selection procedure, able to take in consideration all ecological variables mentioned above. On the remaining part of this chapter we formulate and discuss the research questions this project is going to analyze. Figure 2 shows a Diagram of different factors we are interested to analyze as potential drivers of species diversification; on parenthesis we recognize the research questions related to these factors.

Q1: Do species interaction determine diversification?

Species richness varies widely over the surface of the Earth [Ricklefs, 2007] and patterns of species richness reflect the balance between speciation and extinction over the evolutionary history of life. Theories of how species evolve in changing environments mostly consider single species in isolation or pairs of interacting species [Barraclough, 2015].

We are interested on the impact on interactions between species which happen locally, and therefore are hard to tackle. Reducing the complexity of the system we aim

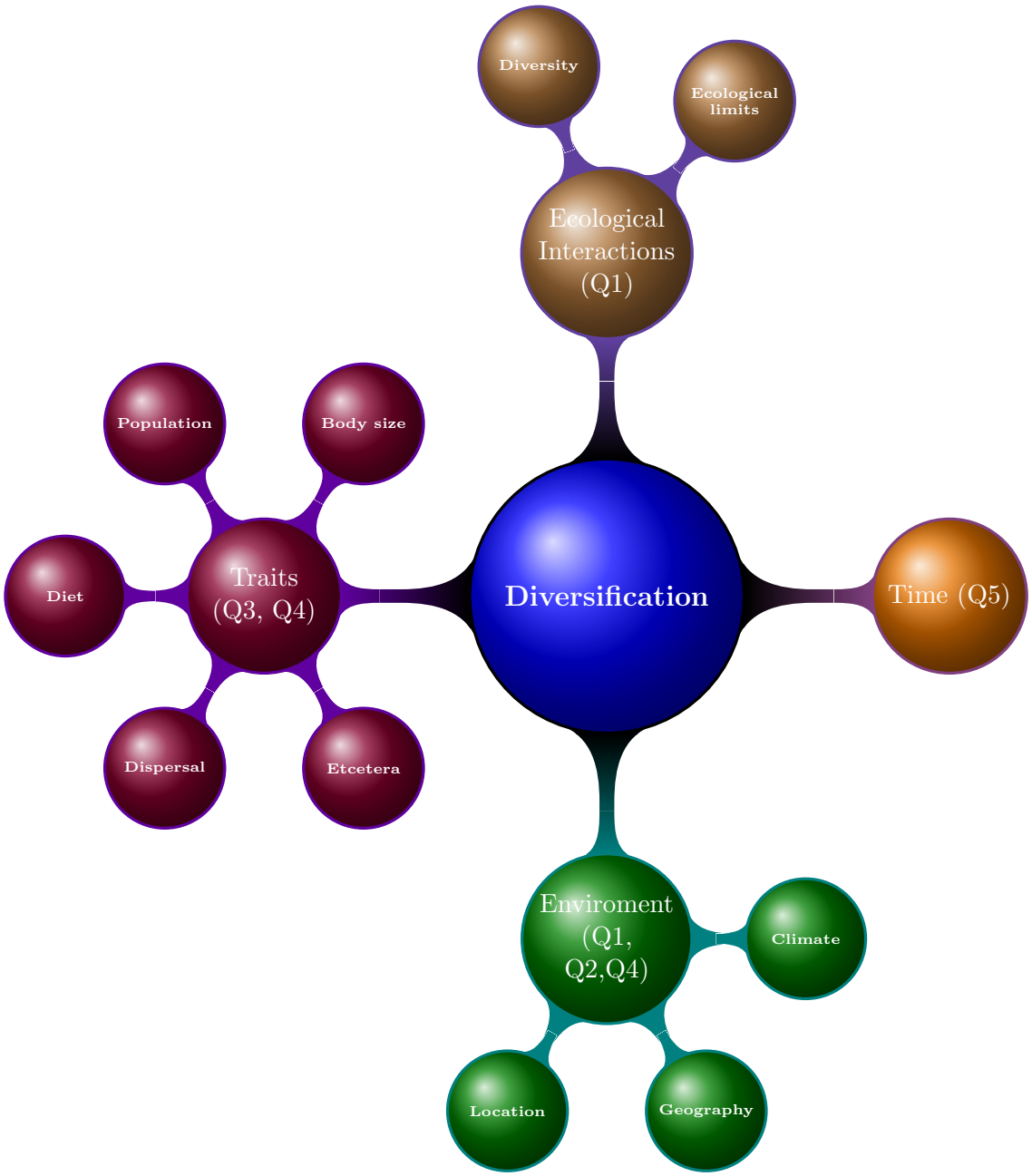

Figure 2.1: Representation of potential factors involved on species diversification we will include in the model. On parenthesis we see the associated research questions discussed on this chapter.

to incorporate a general approach to include species interaction on the diversification process.

An initial step would be to extend the approach of Etienne et al. [Etienne et al., 2011], which considers a caring capacity K based on competition and the idea that ecological constrains place upper limits on regional diversity and that diversity is usually close to its limit [Cornell, 2013]. Thus, they assume that the speciation rate is limited to the caring capacity K and therefore to the diversity of species n

$$\lambda_n = \max(0, \lambda_0 - (\lambda_0 - \mu_n)\frac{n}{K}$$

We will generalize this approach considering the caring capacity $K(x)$ as a function of location $x \in \mathbb{R}^2$. Based on this, an important underlying analysis of this research question, would be to understand what is the shape of $K(x)$ as a function of location, or more generally, the following question:

Q2: How diversification rates and ecological limits varies across location?.

The next question is related to the impact of species' traits on diversification:

Q3: Do different characteristics of species influence diversification? if yes, which ones and how they influence evolution?

The influence of species' traits on lineage diversification is an active area of macroevolutionary research, attributes of species, including population size, generation time, mechanism of pollination and seed dispersal, strength of sexual selection, body mass, longevity of species, diet in insects, latitude of birds and butterflies, sex allocation in flowering plants, and a large etcetera [Barraclough et al., 1998], has been studied and are still a matter of active research. Sister clade comparison [Magallon and Sanderson, 2001] has been a possible way to find relationship between traits and diversification rates, but approaches incorporating phylogenetic tree topology and the patterns of branching times had shown a greater statistical power because they incorporate more information about the patterns of diversification [Paradis, 2005]. Among these methods we find models for binary-states traits [Maddison et al., 2007], quantitative and continuous traits [FitzJohn, 2010], multiple traits [FitzJohn, 2012], etc. But not a general approach where we can include all different kind of traits in a flexible way.

Even though those methods have been successful finding important ecological insights in some way, they have some major drawbacks. For one side, most of the models require complete phylogenies, that is, extant specie must be present in a well-resolved phylogeny [FitzJohn et al., 2009], but in reality information on extinct species is absent or, at best, incomplete; and this will never get better, simple collecting more data is not possible. For other side, even models which deals with incomplete data have a big constrain: they lack of flexibility, in the sense that they are able to be applied to very special cases, but not in a general framework, for any kind of trait.

It is important to note that the question Q3 is very general, and actually triggers

many other research questions. For instance, dispersal is a life-history trait that influences the dynamics and persistence of populations, the distribution and abudance of species, and community structure [Silwood Park et al., 1999], we are also interested in their impact on evolution:

Q4: What is the role of dispersal on evolutionary processes?

Finally, we would like to incorporate the temporal component on evolution as well. Several studies (e.g [Rosindell et al., 2010]) claim that is necessary to incorporate protracted speciation for a proper understanding of biodiversity, this makes us formulate the following question:

Q5: Do we need to consider speciation as a protracted process?

In that way, our model should be flexible enough to incorporate protracted speciation, in the sense that it should also model speciation as a gradual process rather than an instantaneous one.

In this project we aim to build a general statistical framework able to incorporate any potential influential trait in a straightforward manner. Details of the methodology are described in next chapter.

2.2 Incomplete Phylogeny trees

A limitation of phylogenetic reconstruction based on extant species is that lineages are not represented (see Figure 3). This limitations makes estimating speciation and extinction rates problematic life [Ricklefs, 2007]. As was mentioned before information of extinct species is usually absent or, at best, incomplete.

The problem of missing data is widely considered to be the most significant obstacle in both, reconstructing phylogenetic relationships and modeling diversification process [Wiens, 2003]. In this project, following Friedman et al. [Friedman et al., 2002] we will use a structural EM approach to tackle this problem efficiently. The procedure is described on next chapter.

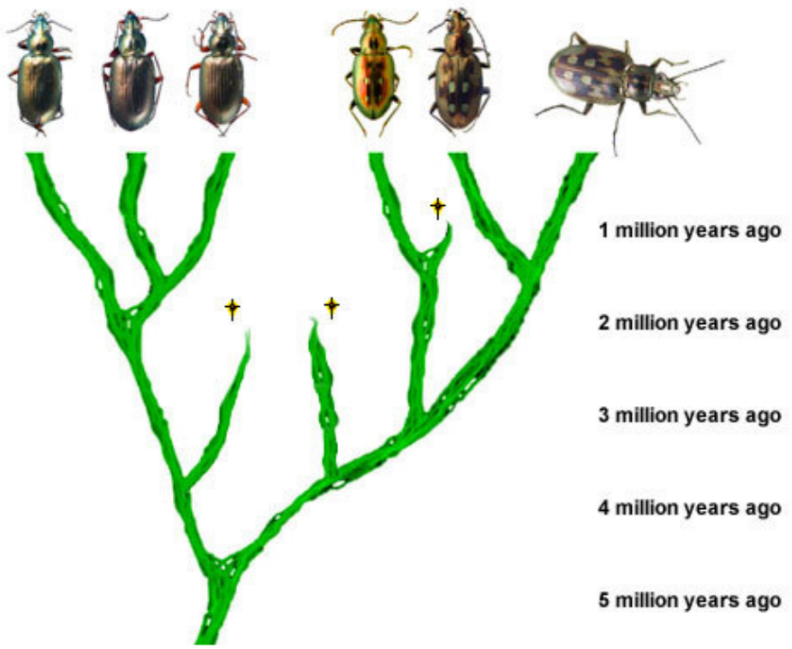

Figure 2.2: A phylogenetic tree of beach beetles of the genus *Bembidion*. Some branches have gone extinct in the past, while others represent species living today.

Chapter 3

Species diversification model

> *Nobody wants to read anyone else's formulas.*
>
> Finman's Law,

For this chapter we will use the following notation:

t_i	: Branching time i.
n_i	: Number of species at time t_i.
m	: Number of covariates
$S_i = \{s_{i,1}, ..., s_{i,n_i}\}$	: Set of extant species at time t_i.
$\lambda_{i,j}$	: Speciation rate for specie $s_{i,j}$ (at time t_i).
$\mu_{i,j}$	: Extinction rate for specie $s_{i,j}$ (at time t_i).
$\mathfrak{s}_i$	: Species which got speciation or extinction at time t_i (note that $\mathfrak{s}_i \in S_i$ as well).
$\mathfrak{x}_1$	: Binary value; $\mathfrak{x}_i$ = "extinction" (0) if an extinction occurs at time t_i and $\mathfrak{x}_i$ = "speciation" (1) if an speciation occurs at time t_i.
ρ_i	: $\mathfrak{x}_i$ rate for specie $\mathfrak{s}_i$.
$x_{i,j,k}$	: k-covariate of the specie j at time t_i.

3.1 Process

We start defining the following sets:

$$\mathcal{T} = \{t_1, ..., t_p\}$$
$$\mathfrak{S} = \{\mathfrak{s}_1, ..., \mathfrak{s}_p\}$$
$$\mathfrak{X} = \{\mathfrak{x}_1, ..., \mathfrak{x}_n\}$$

The phylogenetic tree is mathematically determined by

- A set of branching times $\mathcal{T}$.
- The topology $\Upsilon = \{\mathfrak{S}, \mathfrak{X}\}$.

It is natural to assume [Lemey, 2009] that the time needed for the species $s_{i,j}$ to get speciated follows an exponential distribution with parameter $\lambda_{i,j}$ and the time, of the same species, needed to get extinct follows an exponential distribution with parameter $\mu_{i,j}$. Thus, the waiting time T_i, defined as the minimum time, among all species, to have an speciation or extinction, also follows an exponential distribution

$$P(T_i = t) = \sigma_i e^{-\sigma_i t}$$

where σ_i is the sum of all speciation and extinction rates of extant species at time t_i[1]. That is $\sigma_i = \sum_{j=1}^{n_i} \lambda_{i,j} + \mu_{i,j}$.

Moreover, given the waiting time t_i, the probability of speciation of the species $s_{i,j}$ is $\frac{\lambda_{i,j}}{\sigma_i}$ and the probability of extinction of the same species is $\frac{\mu_{i,j}}{\sigma_i}$.

Then, the general version for the likelihood function of the phylogenetic tree is

$$L(Y|\Theta) = \prod_{i=1}^{p} \sigma_i e^{-\sigma_i t_i} \frac{\rho_i}{\sigma_i} \tag{3.1}$$

Note that

- λ and μ are non-constants but functions of explanatory variables, which includes traits values, environmental phenomena, location, diversity of species or time, which makes the model fairly flexible.
- The probability of Υ follows a Multinomial distribution $M(n|\lambda, \mu)$ with $n = 1$. However we could easily use $n > 1$ which would consider the case of multiple speciation [Scannell et al., 2006].

[1]Note that if $X_i \sim exp(p_i)$, then $X = \min\{X_i, ..., X_n\} \sim exp(\sum p_i)$. For details see [Casella and Berger, 2002]

By tacking partial derivatives equal to zero an analytical solution for 2 in the case of constant rates can be easily calculated; however, as mentioned on the previous chapter, our general (scientific) interest is to find insights of several ecological factors which could have a potential impact on the evolutionary processes of species, thus, the diversification rates are actually functions of explanatory variables

$$\lambda_{i,j} = f(\langle x_{i,j,\cdot}, \beta_1 \rangle)$$
$$\mu_{i,j} = f(\langle x_{i,j,\cdot}, \beta_2 \rangle)$$

where β_1, β_2 are m-dimensional vectors of parameters and $x_{i,j,\cdot}$ are m-dimensional vectors of ecological data (explanatory variables). $\langle \cdot, \cdot \rangle$ represents the inner product of two vectors.

For non-constant rates there is no analytical solution. We used standard optimization algorithms [Bertsekas, 1999] to minimize the likelihood for three different covariates models, however, those methods are not stable.

3.2 Estimation

We are seeking for an stable and flexible method for parameter estimation and model selection. With this purpose, and considering our particular likelihood function, which has parameters which are in turn function of explanatory variables, it worth to introduce a well known statistical kind of models, the Generalized Linear Models.

3.2.1 Generalized Linear Models

The unity of many statistical methods was demonstrated by Nelder and Wedderburn [Nelder and Baker, 1972] using the idea of the Generalized Linear Model (GLM). We present here the general set up for the Generalized Linear Model, specially applied to our research case. For a detailed study of GLM theory we refer the reader to [Dobson and Barnett, 2008] or [McCullagh and Nelder, 1989].

Let $Y_1, Y_2, ..., Y_n$ be independent random variables, from the exponential family [2], and X non-random matrix of explanatory variables or covariates [3]. In general, to specify a GLM we need

- The probability distribution of Y_i.
- Equation linking the expected value of Y_i with a linear combination of the explanatory variables.

[2] For details regarding to the required mathematical assumptions concerning the exponential family of distributions we refer the reader to [Casella and Berger, 2002]

[3] In our case, we are interested on potential effects of ecology factors on evolutionary processes. Then, different traits of species, climate, geographic characteristics, time and diversity of species are possible covariates of our random variable Y

On this way we split the model on a random component and a systematic component, where the random component specifies the distribution of Y_i and the systematic component specifies the way in which the explanatory variables comes into the model. Note that, unlike the current methods mentioned on previous chapter, we are not directly interested on estimate speciation and extinction rates itself, but we want to estimates the critical parameters which controls the underlining processes on ecological evolution. The diversification rates are functions of these parameters.

Thus, we choose a function g such that

$$g(E(Y)) = \eta_i = \langle \mathbf{x_i}, \beta \rangle$$

where:

- g is a monotone, differentiable function called the *link function*.
- The vector $\mathbf{x_i}$ is a $m \times 1$ vector of explanatory variables.
- β is a $m \times 1$ vector of parameters.

The vector $\mathbf{x_i^T}$ is the ith row of the design matrix $\mathbf{X}$.

Note that the inference of β is crucial in the ecological sense, as those are the indicators of relationship of ecological factors and evolutionary processes. On this project we focus our attention on the development of an efficient framework able to capture this relationship; $g(\cdot)$ expresses the relationship of the covariates as a function of $E[Y]$.

Summarizing, a generalized linear model has three components:

1. Response variables $Y_1, ..., Y_n$ from the exponential family.
2. A set of parameters β and explanatory variables $\mathbf{X}$.
3. monotone link function g such that

$$g(E(Y_i)) = \langle \mathbf{x_i}, \beta \rangle$$

3.2.1.1 Iterative Re-Weighed Least Squares algorithm (IRWLS)

For the estimation of parameters we use the IRWLS method [Charnes et al. 1976], which is based on the the method of scoring which in turn is based on the Newton-Raphson formula,

Basically, the method has the following structure:

i. Obtain an initial estimate β,

ii. replace $f(y_i, \beta)$ with a Taylor series approximation,

iii. evaluate all expressions that involve β at the current estimate,

iv. solve the resulting system of equations

v. actualize the new β and repeat (ii)-(v) until $\{\beta^k\}$ converges.

.

Thus, the general formulation in the GLM context has the form,

$$\beta^{(\mathbf{m})} = \beta^{(\mathbf{m+1})} + [\mathcal{J}^{(\mathbf{m-1})}]^{-\mathbf{1}}\mathbf{U}^{(\mathbf{m-1})} \tag{3.2}$$

or

$$\mathcal{J}^{(m-1)}\beta^{(\mathbf{m})} = \mathcal{J}^{(\mathbf{m-1})}\beta^{(\mathbf{m+1})} + \mathbf{U}^{(\mathbf{m-1})} \tag{3.3}$$

where $\beta^{(\mathbf{m})}$ is the vector of estimates of the parameters $\beta_1, ..., \beta_l$ at the mth iteration, $[\mathbb{J}^{(m-1)}]^{-1}$ is the inverse of the information matrix and $U^{(m-1)}$ is the score function evaluated on $\beta^{(\mathbf{m-1})}$.

Thus, we can re-write the information matrix as

$$\mathcal{J} = X^T W X$$

where W is the diagonal matrix with elements

$$w_{ii} = \frac{1}{Var(Y_i)}\left(\frac{\partial E(Y_i)}{\partial \eta_i}\right)^2$$

and the expression on the right-hand side of 4 can be written as

$$X^T W \mathbf{z}$$

where $\mathbf{z}$ has elements

$$z_i = \langle \mathbf{x_i}, \beta \rangle + (\mathbf{y_i} - \mathbf{E}(\mathbf{Y_i}))\left(\frac{\partial \eta_\mathbf{i}}{\partial \mathbf{E}(\mathbf{Y_i})}\right)$$

with $E(Y_i)$ and $\frac{\partial \eta_i}{\partial E(Y_i)}$ are evaluated at $\beta^{(\mathbf{m-1})}$.
Hence the iterative equation 3 can be written as

$$X^T W X \beta^{(\mathbf{m})} = \mathbf{X^T W z}$$

This is the same form as the normal equations for a linear model obtained by weighed least squares, except that it has to be solved iteratively because, in general, the components of the equation depends on β.

3.2.2 Maximun GLM likelihood for phylogenetic trees

The implementation of the previous ideas on the phylogenetic tree has mainly two simultaneous GLM, the branching times (Br) GLM and the Topology (To) GLM. Thus, we re-write equation 2 considering both parts

$$L(Y|\Theta) = \prod_{i=1}^{p} \underbrace{\sigma_i e^{-\sigma_i t_i}}_{Bt} \underbrace{\rho_i/\sigma_i}_{To} \tag{3.4}$$

and we are looking for the log-likelihood function

$$l(\beta) = l^{Bt}(\beta) + l^{To}(\beta)$$

Our aim is to find the MLE of 2, that is

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}}\, l(\beta) = \underset{\beta}{\operatorname{argmax}}\, l^{Bt}(\beta) + l^{B}(\beta)$$

Then, for the whole phylogeny we consider

$$X^T W X = X^{Bt} W^{Bt} X^{Bt} + X^{To} W^{To} X^{To}$$

thus, our iterative procedure corresponds to

$$X^T W X \beta^{(m)} = X^T W \mathbf{z} \tag{3.5}$$

where

$$X = \begin{bmatrix} X^{Bt} \\ X^{To} \end{bmatrix},$$

$$W = \begin{bmatrix} W^{Bt} & 0 \\ 0 & W^{To} \end{bmatrix}$$

and

$$Z = \begin{bmatrix} Z^{Bt} \\ Z^{To} \end{bmatrix}$$

thus, to performs the MLE minimization we need the formulation of

$$X^{Br}, W^{Br}, Z^{Br}, X^{To}, W^{To}, Z^{To}.$$

The branching times (Bt)

As mentioned before, T_i follows an exponential distribution

$$P(T_i = t_i) = \sigma_i e^{-\sigma t_i}$$

thus

$$E(T_i) = \frac{1}{\sigma_i} \text{ and } Var(T_i) = \frac{1}{\sigma_i^2} \tag{3.6}$$

By other side, we have

$$\begin{aligned}\sigma_i &= \sum_{j=1}^{n_i} \lambda_{i,j} + \mu_{i,j} \\ &= \sum_{j=1}^{n_i} \langle x_{i,j,\cdot}, \beta_1 \rangle + \langle x_{i,j,\cdot}, \beta_2 \rangle \\ &= \langle \sum_j x_{i,j,\cdot}, \beta_1 + \beta_2 \rangle \\ &= \left\langle \mathbf{x_i^{Bt}}, \beta^{\mathbf{Bt}} \right\rangle\end{aligned}$$

then, by 7, the equation above, and following the GLM procedure, we get the link function $g(x) = 1/x$.

we use this to calculate the matrix W,

$$w_{ii}^{To} = \frac{1}{Var(T_i)} \left(\frac{\partial E(T_i)}{\partial \eta_i} \right)^2$$

where, $Var(T_i)$ is given by 7 and

$$\frac{\partial E(T_i)}{\partial \eta_i} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} = \frac{\partial \frac{1}{\eta_i}}{\partial \eta_i} = -\frac{1}{\eta_i^2} = -\frac{1}{\langle \mathbf{x_i}, \beta \rangle^2}$$

Then,

$$w_{ii} = 1/\langle \mathbf{x_i}, \beta \rangle^{\mathbf{2}}, \forall \mathbf{i}$$

Thus,

$$W^{Bt} = \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 1/\sigma_p \end{bmatrix}$$

Moreover,

$$z_i^{Bt} = \langle \mathbf{x_i}, \beta \rangle + (t_i - E(T_i)) \left(\frac{\partial \eta_i}{\partial E(T_i)} \right) = \sigma_i(2 - t_i\sigma_i)$$

Then, we could compute the IRWLS iterations for the exponential distribution case

$$(X^{Bt})^T W^{Bt} X^{Bt} \mathbf{b^{(m)}} = (\mathbf{X^{Bt}})^{\mathbf{T}} \mathbf{W^{Bt}} \mathbf{z^{Bt}}$$

where

$$X^B t = \begin{bmatrix} \sum_j x_{1,j,1} & \sum_j x_{1,j,2} & \cdots & \sum_j x_{1,j,m} \\ \sum_j x_{2,j,1} & \sum_j x_{2,j,2} & \ddots & \sum_j x_{2,j,m} \\ \vdots & \ddots & \ddots & \vdots \\ \sum_j x_{p,j,1} & \sum_j x_{p,j,2} & \cdots & \sum_j x_{p,j,m} \end{bmatrix}$$

with $x_{i,j,k}$ is the k-covariate for the specie j at time t_i.

The topology

As described previously the topology Υ_i follows a multinomial distribution $MN(1, \lambda_1, \mu_1, \lambda_2, \mu_2, ..., \lambda_{n_i}, \mu_{n_i})$,

for convenience, we define $\pi_i = \frac{\rho_i}{\sigma_i}$, where ρ_i is the $\mathfrak{x}_i$ rate of the specie $\mathfrak{s}_i$ [4]. In other words, π_i is the probability that the specie $\mathfrak{s}_i$ evolves at time t_i as had happened.

Then, we define the Bernoulli distributed random variable $Z_i \sim B(\pi_i)$,

note that

$$l^{To} = L(Z_i|\pi) = \prod_{i=1}^{n} \pi_i$$

and $E[Z_i] = \pi_i$, $Var(Z_i) = \pi_1(1 - \pi_i)$
then, we can work with Z_i instead of Υ_i as they have same Likelihood function.
we can π_i this as

$$\pi_i = \frac{\eta_i}{\eta_i + c_i}$$

where $\eta_i = \langle x_{i,\mathfrak{s}_i}, \beta \rangle$ and $c_i = \langle \sum_{j \neq \mathfrak{s}_i}^{n_i} x_{i,j}, \beta \rangle$. Thus,

$$\eta_i^{To} \frac{1}{c_i} = \frac{\pi_i}{1 - \pi_i}$$

and

$$g(\pi_i) = \frac{\pi_i}{1 - \pi_i}$$

then,

$$X^{To} = \begin{bmatrix} x_{1,\mathfrak{s}_1,1} & x_{1,\mathfrak{s}_1,2} & \cdots & x_{1,\mathfrak{s}_2,m} \\ x_{2,\mathfrak{s}_2,1} & x_{2,\mathfrak{s}_2,2} & \cdots & x_{2,\mathfrak{s}_2,m} \\ \vdots & \ddots & \cdots & \vdots \\ x_{p,\mathfrak{s}_p,1} & x_{p,\mathfrak{s}_p,2} & \cdots & x_{p,\mathfrak{s}_p,m} \end{bmatrix}$$

$$w_{ii}^{To} = \frac{1}{\pi_i(1 - \pi_i)} \frac{c_i^2}{(\eta_i + c_i)^4} = c_i \pi_i^2$$

and

$$W^{To} = \begin{bmatrix} c_1\pi_1^2 & 0 & \cdots & 0 \\ 0 & c_2\pi_2^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & c_p\pi_p^2 \end{bmatrix}$$

[4] Note that $\mathfrak{x}_i$ is a string corresponding to "speciation" or "extinction", please see the notation section for details.

moreover,

$$z_i^{To} = \sum_{k=1}^{m} x_{i,\mathfrak{s}_i,k}\beta_k + \frac{\pi_i c_i^2}{(1-\pi_i)^3}$$

then, $Z^{To} = [z_1^{To}, ..., z_p^{To}]^T$.

The phylogenetic tree (branching times + topology)

Finally, we have

$$X^{Br}, W^{Br}, Z^{Br}, X^{To}, W^{To}, Z^{To}.$$

With this we can perform the equation 6 iterativelly.

3.2.2.1 Example 1: Diversity-dependence model

As a first example we formulate a diversity-dependence model [Etienne et al., 2011], where the speciation rate depends linearly on the number of species n_i and the extinction rate is constant,

$$\lambda_{i,j} = \lambda_0 - (\lambda_0 - \mu_0)\frac{n_i}{K}, \qquad \mu_n = \mu_0$$

where λ_0, μ_0 and K are the parameters corresponding to initial speciation rate, extinction rate and carrying capacity respectively.

Note that this model assumes equal rates for different species, or in other words, all topologies are equally probable.The log-likelihood function has the form

$$l = \sum_{i=1}^{p} ln(\sigma_i) - \sigma_i t_i + ln\left(\frac{1}{n_i}\right)$$

where

$$\begin{aligned}\sigma_i &= \sum_{j=1}^{n_i} \lambda_0 - (\lambda_0 - \mu_0)\frac{n_i}{K} + \mu_0 \\ &= n_i(\lambda_0 + \mu_0) - n_i^2\left(\frac{\lambda_0 - \mu_0}{K}\right) \\ &= \langle x_i^{Bt}, \beta^{Bt}\rangle = \eta_i\end{aligned} \tag{3.7}$$

where

$$x_i^{Bt} = \begin{bmatrix} n_i \\ n_i^2 \end{bmatrix}, \qquad \beta^{Bt} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_0 + \mu_0 \\ \frac{\mu_0 - \lambda_0}{K} \end{bmatrix}$$

then,

$$\frac{\partial E[T_i]}{\partial \eta_i} = \frac{\partial (1/\sigma_i)}{\partial \sigma_i} = -\frac{1}{\sigma_i^2}$$

and

$$\frac{\partial \eta_i}{\partial E[T_i]} = \frac{\partial (1/E[T_i])}{\partial E[T_i]} = -\frac{1}{E^2[T_i]} = -\sigma_i^2$$

thus,

$$w_{ii}^{Bt} = \sigma_i^2 \frac{1}{\sigma_i^4} = \frac{1}{\sigma_i^2}$$

and

$$W^{Bt} = \begin{pmatrix} \frac{1}{\sigma_1^2} & & & \\ & \ddots & 0 & \\ & & \ddots & \\ & 0 & & \ddots \\ & & & \frac{1}{\sigma_p^2} \end{pmatrix}$$

moreover,

$$z_i^{Bt} = n_i(\lambda_0 + \mu_0) + n_i^2 \frac{\mu_0 - \lambda_0}{K} - (t_i - \frac{1}{\sigma})\sigma_i^2$$

then

$$\mathbf{z^{Bt}} = \begin{bmatrix} n_1(\lambda_0 + \mu_0) + n_1^2 \frac{\mu_0 - \lambda_0}{K} - (t_1 - \frac{1}{\sigma_1})\sigma_1^2 \\ \vdots \\ n_p(\lambda_0 + \mu_0) + n_p^2 \frac{\mu_0 - \lambda_0}{K} - (t_p - \frac{1}{\sigma_p})\sigma_1^2 \end{bmatrix}$$

and

$$X^{Bt} = \begin{bmatrix} n_1 & n_1^2 \\ \vdots & \vdots \\ n_p & n_p^2 \end{bmatrix}$$

Finally, with all these values we are ready to run the following iterative procedure which will find the MLE parameter estimator

$$\beta^{(m)} = [(X^{Bt})^T W^{Bt} X^{Bt}]^{-1} (X^{Bt})^T W^{Bt} \mathbf{z^{Bt}}$$

Note that the right side of the equation above depends on β, so we choose β_0 and then we replace β with the previous calculated beta on each iteration.

This is the simplest model for diversity dependence, however this model does not account for geographic variables, and does not even use topology information. We aim to generalize it considering location information and incorporating traits dependence, under the GLM approach this extension is quite straightforward.

3.2.2.2 Example 2: Linear model for traits dependence

As a second example, we generalize the linear model for traits dependence [Paradis, 2005] including a non-constant extinction rate. The relation of body mass and diversification rates have been studied with different approaches [Gittleman and Purvis, 1998], based on that we formulate diversification rates as

$$\lambda_{i,j} = \beta_0 + \beta_1 ln(bodymass_{i,j}) = \beta_0 + \beta_1 v_{i,j}$$
$$\mu_{i,j} = \beta_3 + \beta_4 ln(bodymass_{i,j}) = \beta_2 + \beta_3 v_{i,j}$$

Here we define $v_{i,j}$ as the log of the body mass of specie j at time t_i, then for the branching times formulation we calculate

$$\begin{aligned} \sigma_i &= \sum_{j=1}^{n_i} \lambda_{i,j} + \mu_{i,j} \\ & n_i(\beta_0 + \beta_2) + \left[\sum_{j=1}^{n_i} v_{i,j}\right](\beta_1 + \beta_2) \\ & = \langle x_i^{Bt}, \beta^{Bt} \rangle = \eta_i^{Bt} \end{aligned} \tag{3.8}$$

where

$$x_i^{Bt} = \begin{bmatrix} n_i \\ \sum_{j=1}^{n_i} v_{i,j} \end{bmatrix}, \qquad \beta^{Bt} = \begin{bmatrix} \beta_0 + \beta_2 \\ \beta_1 + \beta_3 \end{bmatrix}$$

moreover, we have

$$W^{Bt} = \begin{pmatrix} \frac{1}{\sigma_1^2} & & & \\ & \ddots & 0 & \\ & & \ddots & \\ & 0 & & \frac{1}{\sigma_p^2} \end{pmatrix},$$

$$\mathbf{z^{Bt}} = \begin{bmatrix} n_1(\beta_0 + \beta_2) + \left(\sum_{j=1}^{n_1} v_{1,j}\right)(\beta_1 + \beta_2) + (t_1 - 1/\sigma_1)\sigma_1^2 \\ \vdots \\ n_p(\beta_0 + \beta_2) + \left(\sum_{j=1}^{n_p} v_{p,j}\right)(\beta_1 + \beta_2) + (t_p - 1/\sigma_p)\sigma_p^2 \end{bmatrix}$$

and

$$X^{Bt} = \begin{bmatrix} n_1 & \sum_{j=1}^{n_1} v_{1,j} \\ \vdots & \vdots \\ n_p & \sum_{j=1}^{n_p} v_{p,j} \end{bmatrix}$$

For the topology part we consider

$$\pi_i = \frac{\rho_i}{\sigma_i} = \frac{(\beta_0 + \beta_1 v_{i,\mathfrak{s}_i})\mathbb{1}(\mathfrak{x}_\mathfrak{i}) + (\beta_2 + \beta_3 v_{i,\mathfrak{s}_i})(1 - \mathbb{1}(\mathfrak{x}_\mathfrak{i}))}{n_i(\beta_0 + \beta_2) + (\sum_j^{n_i} v_{i,j})(\beta_1 + \beta_3)}$$

where

$$\mathbb{1}(\mathfrak{x}_\mathfrak{i}) = \begin{cases} 1 & if \quad \mathfrak{x}_\mathfrak{i} = "speciation" \\ 0 & if \quad \mathfrak{x}_\mathfrak{i} = "extinction" \end{cases}$$

then, following the procedure described on previous chapter, we re-write π_i on the form

$$\pi_i = \frac{\rho_i}{\rho_i + c_i}$$

where $c_i = \sigma_i - \rho_i$. moreover,

$$\begin{aligned} \rho_i &= \beta_0 \mathbb{1}(\mathfrak{x}_\mathfrak{i}) + \beta_1 v_{i,\mathfrak{s}_i} \mathbb{1}(\mathfrak{x}_\mathfrak{i}) + \beta_2 + \beta_3 v_{i,\mathfrak{s}_i} - \beta_2 \mathbb{1}(\mathfrak{x}_\mathfrak{i}) - \beta_3 v_{i,\mathfrak{s}_i} \mathbb{1}(\mathfrak{x}_\mathfrak{i}) \\ &= \langle x_i^{To}, \beta^{To} \rangle = \eta_i^{To} \end{aligned}$$

where,

$$x_i^{To} = \begin{bmatrix} \mathbb{1}(\mathfrak{x}_\mathfrak{i}) \\ v_{i,v_{i,\mathfrak{s}_i}} \mathbb{1}(\mathfrak{x}_\mathfrak{i}) \\ 1 \\ v_{i,v_{i,\mathfrak{s}_i}} \end{bmatrix}, \qquad \beta^{To} = \begin{bmatrix} \beta_0 - \beta_2 \\ \beta_1 - \beta_3 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

and

$$W^{To} = \begin{bmatrix} c_1\pi_1^2 & 0 & \cdots & 0 \\ 0 & c_2\pi_2^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & c_p\pi_p^2 \end{bmatrix}$$

moreover,

$$z_i^{To} = \sum_{k=1}^{m} x_{i,\mathfrak{s}_i,k} \beta_k + \frac{\pi_i c_i^2}{(1 - \pi_i)^3}$$

then, $Z^{To} = [z_1^{To}, ..., z_p^{To}]^T$.
With all these values we perform iteratively the following equation

$$\beta^{(m)} = [(X^{Bt})^T W^{Bt} X^{Bt}]^{-1} (X^{Bt})^T W^{Bt} \mathbf{Z^{Bt}}$$

Discussion

As we can see, this method allows easy generalization to important extensions on ecology:

- We can include any kind of multiple traits, quantitative and continuous, binary traits (using identity functions), geographical variables like range of species, etcetera.
- We can also study migration influences including on the link function of the GLM.
- As mentioned before we can extend to the case of multiple speciation allowing the multivariate distribution to have parameter $n > 1$.
- We can include protracted speciation adding an additional parameter to the exponential distribution.

Most of the data is indeed incomplete. To overcome this statistical problem we introduce the well known Expectation Minimization algorithm in next section.

3.2.3 Incomplete phylogenetic trees. The EM algorithm

As discussed on the introductory chapter, in evolutionary ecology, or more specifically, on phylogenic trees, we are actually full of missing observations since we can only observe extant species with molecular data and fossil record is very poor and most of the time absent. Then, the described model is not really enough to afford real data.

In statistics, an expectation-maximization (EM) [Dempster et al., 1977] algorithm is an iterative method for finding MLE of parameters in statistical models, where the model depends on unobserved latent variables.

More formally, we consider $\mathcal{Y} = \{\tau_{obs}, \Upsilon_{obs}\}$ as the tree containing observed branching times and topology of the tree. In a similar way we define $y_{miss} = \{\tau_{mis}, \Upsilon_{mis}\}$ as the "missing part" of the real phylogenetic tree, that is information of extinct species not included in current data. Finally we define the real phylogenetic tree of a clade as $\mathcal{Z} = \mathcal{Y} \cup y_{miss}$.

We then define the Q function as the expectation of the log-likelihood function

$$Q(\beta, \beta_0) = \int_{Y_{miss}} log(p(\beta|y_{miss}, \mathcal{Y}))p(y_{miss}|\beta_0, \mathcal{Y})dy_{miss}$$

Where Y_{miss} is the (infinite) set of all possible latent trees, or missing part of the real tree. Then $y_{miss} \in Y_{miss}$ is a variable, and Q is an integral along all possible values of y_{miss}. In this way, given the current approximation to the maximizer of the observed posterior ($\beta^{(i)}$), the E step of the EM algorithm is defined by computing

$$Q(\beta, \beta^{(i)}) = \int_{Y_{miss}} log(p(\beta|y_{miss}, \mathcal{Y}))p(y_{miss}|\beta^{(i)}, \mathcal{Y})dy_{miss} \tag{3.9}$$

The M step then corresponds to maximize the Q function with respect to β to obtain the update $\beta^{(i+1)}$.

In the case of phylogenetic trees, the calculation of 10 is not straightforward since p has the form of equation 5. To perform this integration we use the Monte Carlo EM(MCEM) algorithm [Wei and Tanner, 1990]. The general structure is as follows

i. Obtain an (initial) estimate $\beta^{(i)}$,

ii. generate a sample $y_{miss}^{(1)}, ..., y_{miss}^{(h)}$ from the current approximation to the conditional predictive distribution $p(y_{miss}|\beta^{(i)}, \mathcal{Y})$,

iii. update the current approximation to $Q_{i+1}(\beta, \beta^{(i)})$, to be the mixture of augmented log-posteriors of β, mix over the latent data patterns from (ii)

$$Q_{i+1}(\beta, \beta^{(i)}) = \frac{1}{N} \sum_{j=1}^{N} log(p(\beta|y_{miss}^{(j)}, \mathcal{Y}))$$

iv. the M step corresponds to the maximization of the right side of the equation above.

v. update the conditional predictive distribution, actualize the new β and repeat (ii)-(v) until convergence.

3.2.4 Automatic selection of variables: differential geometric extension of the least angle regression method.

So far, we have described the mathematical methodology for inference of parameters based on incomplete phylogenies. However, as mentioned on previous chapter, the complexity of the problem is just huge, we need to deal with an enormous amount of ecological and complex information which will be almost impossible to handle with the structural EM algorithm itself. To make this framework feasible we embed a differential geometric path finding method [Augugliaro et al., 2013] inside the M-step if the EM algorithm. This will produce a sparse, computationally feasible and consistent model selection procedure.

As defined previously $\mathcal{Y}$ as the observed phylogenetic tree, and $\mathcal{Z}$ indicates the complete phylogenetic tree (see figure 3). Then, the the q-th score function, measuring the likelihood increase in the q-th direction is given by

$$\partial_q l(\beta; y) = \frac{\partial log(p(y; \beta))}{\partial \beta_q}$$

To measure the scale of this change, relative to the size of the q-th predictor, we scale the score function by the square root of the conditional Fisher information, i.e $I_q(\beta) = \partial^2 l(\beta; y)/\partial\beta_q^2$ to obtain the conditional Rao score statistic,

$$r_q^u(\beta) = \frac{\partial_q l(\beta; y)}{\sqrt{I_q(\beta)}}$$

We can note that, under regularity conditions, at the maximum likelihood estimate the Rao score statistic is equal to zero, since the derivatives of the likelihood are zero.

On the other hand, for an initial estimate $\hat{\beta}_0$, for instance $\hat{\beta}_0 = (b_0, 0, ..., 0)$, we define γ_{max} to be the largest absolute value of the Rao score statistic at $\hat{\beta}_0$

$$\gamma_{max} = \max_{q=1,...,m} |r_q^u(\hat{\beta}_0)|$$

This value points out the best instantaneous normalized contribution to the likelihood of a single variable. This particular variable would make an excellent candidate for being included in the model. With these definitions in hand, we can now define the model extension estimator.

Let $\gamma \in [0, \gamma_{max}]$ be a fixed value. The model extension estimator of a given spatial birth-death model, denoted by $\hat{\beta}(\gamma) \in \mathbb{R}^m$, is such that the following conditions are satisfied:

$$|r_q^u(\hat{\beta}(\gamma))| = \gamma, \qquad \forall q \in \mathcal{A}(\gamma)$$
$$|r_q^u(\hat{\beta}(\gamma))| < \gamma, \qquad \forall q \notin \mathcal{A}(\gamma)$$

where $\mathcal{A}(\gamma) = \{m : \hat{\beta}_q(\gamma) \neq 0\}$ is called active model set.

Here we have defined the model extension estimator in terms of the Rao score statistic. Augugliaro et al. [Augugliaro et al., 2013] shows that the conditions above follow naturally from a differential geometric interpretation of a model, which allows us to generalize the least angle regression method (LARS) introduced in Efron et al [Efron et al., 2004]. The new method generalized the geometric description of LARS and is based on the following simple differential geometric identity

$$r_q^u(\beta) = cos\rho_q(\beta) \cdot \|r_\beta(\mathcal{Y})\|_{p(E(\beta))}$$

where in this case $\rho_q(\beta)$ is the angle between $\partial_q l(\beta; \mathcal{Y})$ and the tangent residual vector $r_\beta(\mathcal{Y})$ while $\|r_\beta(\mathcal{Y})\|_{p(E(\beta)}$ is the length of this vector – which, crucially, does not depend on variable q. Using figure 4 the dgLARS method can be described in the following way. First the method selects the predictor, say $\mathcal{X}_{a_1}$, whose basis vector $\partial_{a_1} l(\hat{\beta}(\gamma_{max}); \mathcal{Y})$ has the smallest angle with the tangent residual vector, and includes it in the active set $\mathcal{A}(\gamma^{(1)} = \{a_0, a_1\}$, where a_0 stands for the intercept and $\gamma^{(1)} = \gamma_{max}$. The solution curve $\hat{\beta}(\gamma) = (\hat{\beta}_{a_0}(\gamma), \hat{\beta}_{a_1}(\gamma), 0, ..., 0)^T$ is chosen in such a way that the tangent residual vector is always orthogonal to the basis $\partial_{a_0} l(\hat{\beta}(\gamma); \mathcal{Y})$, while the direction of the curve $\hat{\beta}(\gamma)$ is defined by the projection of the tangent residual

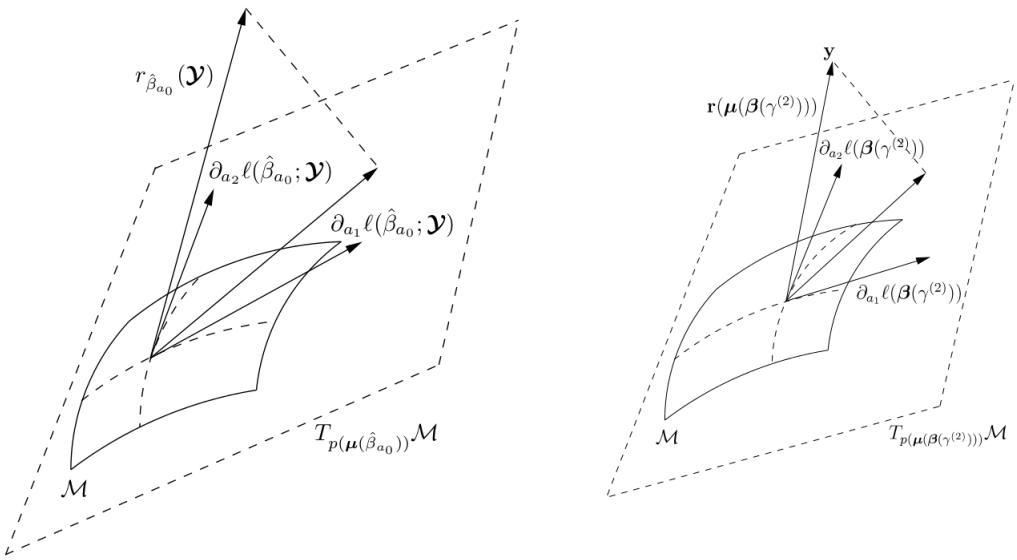

Figure 3.1: Differential geometrical description of the model extension method for birth-death speciation model with two covariates; in the left side the first predictor $\mathcal{X}_{a_1}$ is found and included in the active set; in the right side the generalized equiangularity condition 11 is satisfied for $\mathcal{X}_{a_1}$ and $\mathcal{X}_{a_2}$ [Augugliaro et al., 2013]

vector onto the basis vector $\partial_{a_1} l(\hat{\beta}(\gamma); \mathcal{Y})$. The curve $\hat{\beta}(\gamma)$ continues as defined above until $\gamma^{(2)} < \gamma^{(1)}$, for which there exists a new predictor, say $\mathcal{X}_{a_2}$, that satisfies the equiangularity condition, namely

$$\rho_{a_1}(\hat{\beta}(\gamma^{(2)})) = \rho_{a_2}(\hat{\beta}(\gamma^{(2)})) \tag{3.10}$$

At this point $\mathcal{X}_{a_2}$ is included in $\mathcal{A}(\gamma^{(2)})$ and the curve

$$\hat{\beta}(\gamma) = (\hat{\beta}_{a_1}(\gamma), \hat{\beta}_{a_1}(\gamma), \hat{\beta}_{a_2}(\gamma), 0, ..., 0)^T$$

continues, such that the tangent residual vector is always orthogonal to the basis vector $\partial_{a_1} l(\hat{\beta}; \mathcal{Y})$ and $\partial_{a_2} l(\hat{\beta}(\gamma); \mathcal{Y})$, as shown on the right side of figure 4.

At the end, model reduction requires selecting a tuning parameter $\hat{\gamma}$ that controls the sparsity or shrinkage of the model. Especially in complex models, like this one, it is important to tune this inference to the available data [Wit et al., 2012]. For that purposes we will adjust the fast generalized cross-validation and generalized information criterion [Abbruzzo et al., 2014] [Vujacic et al., 2013] adjusted to phylogenetic trees.

Chapter 4

Knowledge utilization

4.1 Implementation of results

4.1.1 Software development

We will develop free software packages for the often mathematically and computationally demanding inference techniques, allowing non-specialist scientist to apply these newly developed methods easily to their own systems.

4.1.2 Application to real data

We will test the methods on published phylogenies, but also on a new phylogeny of microlandsnails on limestone outcrops on Malaysian Borneo that will be constructed in research funded by a VICI grant awarded to R.S. Etienne. This system is unique in that contains well-defined local communities (the snails are strongly restricted to the limestone) with limited dispersal and high endemism. Besides contituting a miniature world ideally suited for trying out our new methods, the microsnail ecological evolutionary dynamica is a real life system that can reveal the effects of global warming on biodiversity.

4.2 Contribution of research to specific fields beyond mathematics, ecology and evolution: Language evolution

Phylogenetic methods are increasingly used to study language evolution. This project can help identify the phylogeographic context and the influence of community effects (e.g number of languages present) on language diversification in different human societies. We will collaborate with Quentin Atkinson (University of Auckland) and localized evolution models are also there an important inferential hurdle that has not been taken yet.

4.3 Contribution of research to society: Global conservation

Understanding the processes driving biodiversity is of crucial importance for assessing the effects of global change on the diversity of life on this planet. This project will help addressing the following key issues: (1) if local diversity limits further diversification, will a reduction in diversity then speed up diversification and restore the balance, and if so, over what time scale will equilibrium be recovered; or (2) will lost diversity never be regained, because the loss happens too quickly for natural processes to compensate?

4.4 Open Access policy

Scientific publications arising from this project will be made publicly accessible to the research community, by depositing submitted/accepted manuscripts at arXiv.org and choosing Open Access journals options. This will allow the unconditional and immediate availability of the scientific results for use by other scientist and general public, particularly in developing countries.

Chapter 5

Time Planning

2016 May - 2017 Apr	**Project 1:** Spatial birth-death model of diversification. **Deliverable:** Paper.
2017 May - 2018 Apr	**Project 2:** Robust inference of spatial diversification. **Deliverable:** Paper, R package (version 1). Organize statistical workshop on geographical dynamics of speciation..
2018 May - 2019 Apr	**Project 3:** Geographic dynamics of speciation events in real applications. **Deliverable:** Paper, R package (version 2).
2019 May - 2019 Oct	Complete PhD thesis. **Deliverable:** Thesis..

Bibliography

[Abbruzzo et al., 2014] Abbruzzo, A., Vujačić, I., Wit, E., and Mineo, A. M. (2014). Generalized information criterion for model selection in penalized graphical models. *arXiv preprint arXiv:1403.1249*.

[Abegaz and Wit, 2013] Abegaz, F. and Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, page kxt005.

[Augugliaro et al., 2013] Augugliaro, L., Mineo, A. M., and Wit, E. C. (2013). Differential geometric least angle regression: a differential geometric approach to sparse generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3):471–498.

[Barraclough, 2015] Barraclough, T. G. (2015). How do species interactions affect evolutionary dynamics across whole communities? *Annual Review of Ecology, Evolution, and Systematics*, 46:25–48.

[Barraclough et al., 1998] Barraclough, T. G., Vogler, A. P., and Harvey, P. H. (1998). Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1366):241–249.

[Bertsekas, 1999] Bertsekas, D. P. (1999). Nonlinear programming.

[Casella and Berger, 2002] Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

[Chivian and Bernstein, 2008] Chivian, E. and Bernstein, A. (2008). *Sustaining life: how human health depends on biodiversity*. Oxford University Press.

[Cornell, 2013] Cornell, H. V. (2013). Is regional species diversity bounded or unbounded? *Biological Reviews*, 88(1):140–165.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

[Dobson and Barnett, 2008] Dobson, A. J. and Barnett, A. (2008). *An introduction to generalized linear models*. CRC press.

[Efron et al., 2004] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

[Etienne et al., 2011] Etienne, R. S., Haegeman, B., Stadler, T., Aze, T., Pearson, P. N., Purvis, A., and Phillimore, A. B. (2011). Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20111439.

[Etienne and Rosindell, 2012] Etienne, R. S. and Rosindell, J. (2012). Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Systematic Biology*, 61(2):204–213.

[FitzJohn, 2010] FitzJohn, R. G. (2010). Quantitative traits and diversification. *Systematic biology*, page syq053.

[FitzJohn, 2012] FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in r. *Methods in Ecology and Evolution*, 3(6):1084–1092.

[FitzJohn et al., 2009] FitzJohn, R. G., Maddison, W. P., and Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic biology*, 58(6):595–611.

[Friedman et al., 2002] Friedman, N., Ninio, M., Pe'er, I., and Pupko, T. (2002). A structural em algorithm for phylogenetic inference. *Journal of Computational Biology*, 9(2):331–353.

[Gillman, 2009] Gillman, M. (2009). *An introduction to mathematical models in ecology and evolution: time and space*, volume 4. John Wiley & Sons.

[Gittleman and Purvis, 1998] Gittleman, J. L. and Purvis, A. (1998). Body size and species–richness in carnivores and primates. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1391):113–119.

[Jetz et al., 2012] Jetz, W., Thomas, G., Joy, J., Hartmann, K., and Mooers, A. (2012). The global diversity of birds in space and time. *Nature*, 491(7424):444–448.

[Kendall, 1948] Kendall, D. G. (1948). On some modes of population growth leading to ra fisher's logarithmic series distribution. *Biometrika*, 35(1/2):6–15.

[Lemey, 2009] Lemey, P. (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press.

[Maddison et al., 2007] Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic biology*, 56(5):701–710.

[Magallon and Sanderson, 2001] Magallon, S. and Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution*, 55(9):1762–1780.

[McCullagh and Nelder, 1989] McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press.

[Moen and Morlon, 2014] Moen, D. and Morlon, H. (2014). Why does diversification slow down? *Trends in ecology & evolution*, 29(4):190–197.

[Nee et al., 1994] Nee, S., May, R. M., and Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 344(1309):305–311.

[Nelder and Baker, 1972] Nelder, J. A. and Baker, R. J. (1972). Generalized linear models. *Encyclopedia of Statistical Sciences*.

[Paradis, 2005] Paradis, E. (2005). Statistical analysis of diversification with species traits. *Evolution*, 59(1):1–12.

[Ricklefs, 2007] Ricklefs, R. E. (2007). Estimating diversification rates from phylogenetic information. *Trends in Ecology & Evolution*, 22(11):601–610.

[Rosindell et al., 2010] Rosindell, J., Cornell, S. J., Hubbell, S. P., and Etienne, R. S. (2010). Protracted speciation revitalizes the neutral theory of biodiversity. *Ecology Letters*, 13(6):716–727.

[Scannell et al., 2006] Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., and Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*, 440(7082):341–345.

[Silwood Park et al., 1999] Silwood Park, U., Dytham, C., and Perrin, N. (1999). The evolutionary ecology of dispersal.

[Tilman et al., 2006] Tilman, D., Reich, P. B., and Knops, J. M. (2006). Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature*, 441(7093):629–632.

[Vujacic et al., 2013] Vujacic, I., Abbruzzo, A., and Wit, E. (2013). A computationally fast alternative to cross-validation in penalized Gaussian graphical models. *ArXiv e-prints*.

[Wake and Vredenburg, 2008] Wake, D. B. and Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? a view from the world of amphibians. *Proceedings of the National Academy of Sciences*, 105(Supplement 1):11466–11473.

[Wei and Tanner, 1990] Wei, G. C. and Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, 85(411):699–704.

[Wiens, 2003] Wiens, J. J. (2003). Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology*, 52(4):528–538.

[Wit et al., 2012] Wit, E., Heuvel, E. v. d., and Romeijn, J.-W. (2012). all models are wrong...: an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236.

[Yule, 1925] Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of dr. jc willis, frs. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213:21–87.