"Evolution is not to be understood as a series of tournaments for the occupation of a fixed set of enviromental niches... Instead evolution brings about a proliferation of niches... Each new bird or mammal provides a niche for one or more new kind of flea."

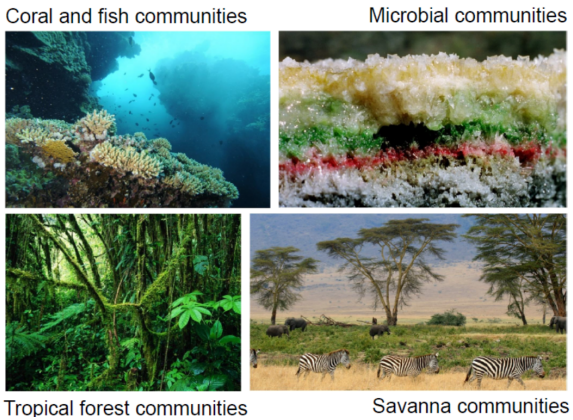
— Herbert Simon, The Sciences of the Artificial, 1969.

The aim of this project is to develop **realistic stochastic models** and **reliable statistical inference methods** for species diversification.

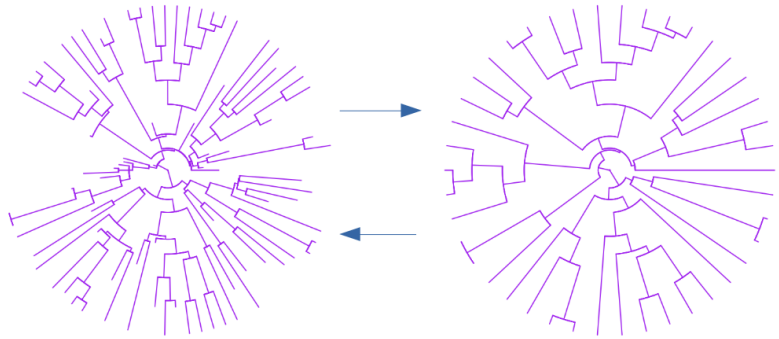
The aim of this project is to develop **realistic stochastic models** and **reliable statistical inference methods** for species diversification.

To do that we mainly face two big challenges:

- Complexity



- Incomplete information



Set up

The phylogenetic tree is mathematically determined by

- A set of branching times $\mathcal{T}$
- The topology Υ.

And its likelihood function is defined as

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

Example: Diversity-dependence model

For the simplest diversity-dependence model

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

The MLE can be finded partially analiticaly and partially numerically.

Example: Diversity-dependence model

For the simplest diversity-dependence model

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

The MLE can be finded partially analiticaly and partially numerically. First we consider σ_i and ρ_i

$$\sigma_i = \sum_{j=1}^{N} \lambda_0 - (\lambda_0 - \mu_0)\frac{n_i}{K} + \mu_0 = n_i(\lambda_0 + \mu_0) - n_i^2\beta_0$$

where $\beta_0 = \left(\frac{\lambda_0 - \mu_0}{K}\right)$, and

$$\rho_i = E_i(\lambda_0 - n_i\beta_0) + (1 - E_i)\mu_0$$

Firstly, after some algebra, we find a very nice analytical solution for the extinction rate parameter

$$\frac{\partial l(\lambda, \beta, \mu | Y)}{\partial \mu} = 0 \Leftrightarrow \hat{u}_0 = \frac{\sum_{i=1}^{N} (1 - E)}{\sum_{i=1}^{N} (n_i t_i)} \tag{2}$$

Firstly, after some algebra, we find a very nice analytical solution for the extinction rate parameter

$$\frac{\partial l(\lambda, \beta, \mu | Y)}{\partial \mu} = 0 \Leftrightarrow \hat{u}_0 = \frac{\sum_{i=1}^{N}(1 - E)}{\sum_{i=1}^{N}(n_i t_i)} \tag{2}$$

Moreover, with the other two equations, we have the following system

$$\begin{cases} \sum_{i=1}^{N} \frac{E_i}{\lambda - n_i \beta} = \sum_{i=1}^{N} n_i t_i \\ \sum_{i=1}^{N} \frac{E_i n_i}{\lambda - n_i \beta} = \sum_{i=1}^{N} n_i^2 t_i \end{cases}$$

Table: MLE estimation of 100 simulations. Simulations and estimations are from the algorithm described above.

Simulated Parameters				Estimated parameters (25th, 50th, 75th percentiles)								
				λ_0			μ			K		
λ_0	K	crown age	μ	025th	50th	75th	025th	50th	75th	025th	50th	75th
0.8	40	5	0	0.71	0.87	1.04	0.00	0.00	0.00	31.20	39.09	440.16
			0.1	0.76	0.92	1.11	0.07	0.10	0.13	22.23	32.65	65.39
			0.2	0.80	0.96	1.28	0.13	0.19	0.26	12.74	31.76	83.12
			0.4	1.05	1.23	1.55	0.27	0.36	0.44	7.00	17.61	30.58
		10	0	0.68	0.79	0.87	0.00	0.00	0.00	39.63	40.98	43.92
			0.1	0.71	0.86	0.98	0.09	0.10	0.12	37.11	39.52	42.01
			0.2	0.78	0.91	1.05	0.18	0.20	0.23	34.08	38.13	43.32
			0.4	0.87	1.01	1.20	0.37	0.41	0.46	18.32	30.13	42.13
		15	0	0.69	0.77	0.87	0.00	0.00	0.00	39.58	40.00	41.03
			0.1	0.72	0.80	0.91	0.09	0.10	0.11	38.72	39.89	40.98
			0.2	0.78	0.84	0.96	0.18	0.20	0.22	38.29	40.25	41.93
			0.4	0.79	0.90	1.00	0.38	0.40	0.43	31.40	37.38	43.89

Table: MLE estimation of 100 simulations. Simulations are from the 'DDD' package and estimation from p1 algorithm.

Simulated Parameters				Estimated parameters (25th, 50th, 75th percentiles)								
λ_0	K	crown age	μ	λ_0			μ			K		
				025th	50th	75th	025th	50th	75th	025th	50th	75th
0.8	40	5	0	0.74	0.91	1.11	0.00	0.00	0.00	34.84	41.04	59.34
			0.1	0.94	1.15	1.28	0.08	0.11	0.14	26.57	32.55	39.98
			0.2	1.03	1.22	1.46	0.17	0.21	0.29	17.06	27.85	37.47
			0.4	1.13	1.43	1.67	0.33	0.42	0.52	9.40	17.52	27.29
		10	0	0.75	0.87	0.98	0.00	0.00	0.00	38.55	39.33	40.35
			0.1	0.78	0.90	1.05	0.09	0.10	0.12	37.09	38.55	40.33
			0.2	0.86	0.96	1.08	0.19	0.21	0.23	34.47	37.45	40.38
			0.4	0.96	1.11	1.22	0.38	0.42	0.45	28.11	33.88	40.41
		15	0	0.74	0.83	0.96	0.00	0.00	0.00	38.57	39.05	40.00
			0.1	0.80	0.88	0.99	0.09	0.11	0.12	37.57	38.67	39.57
			0.2	0.88	0.95	1.05	0.20	0.21	0.22	36.98	38.58	40.07
			0.4	0.92	1.03	1.14	0.38	0.41	0.44	34.02	37.45	41.96

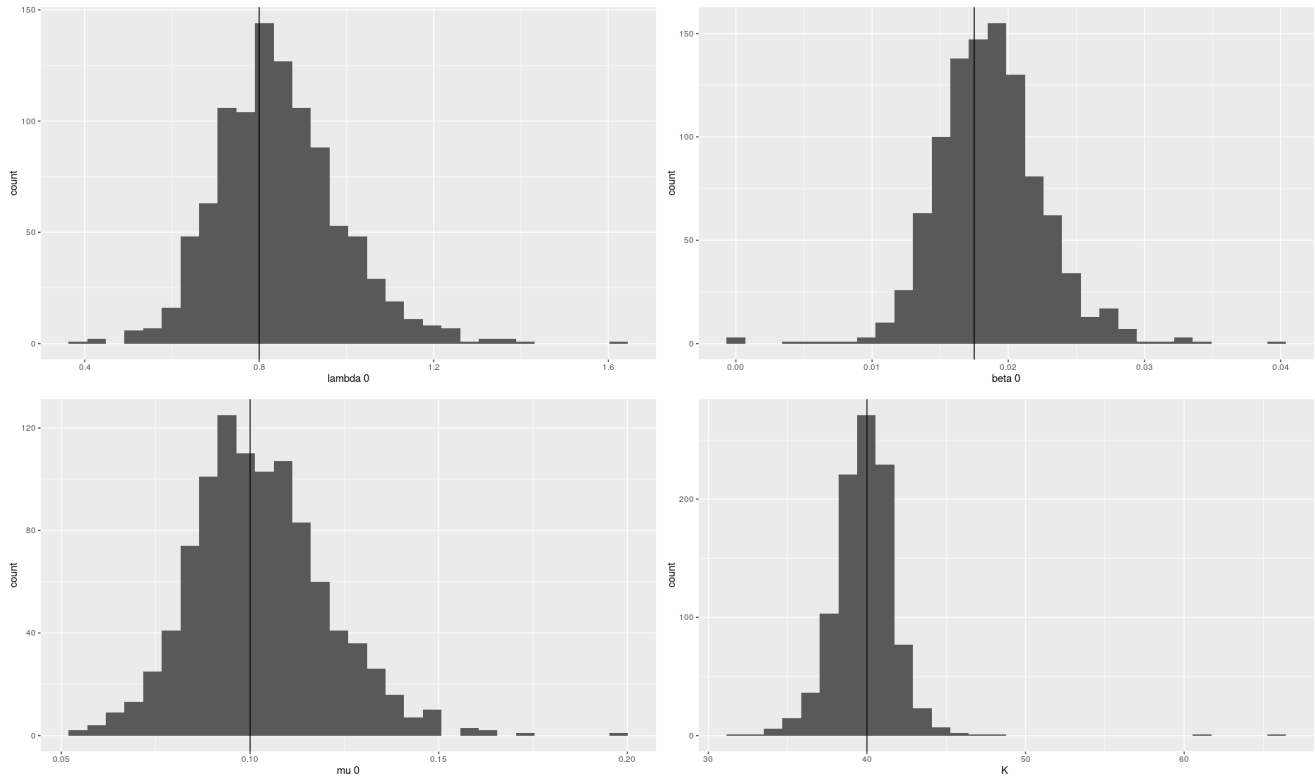


Figure: Estimations over 1000 simulations of the diversity-dependence process with true values $\lambda_=0.8, \beta_0 = 0.0175, \mu_0 = 0.1, K = 40$ and crown time $= 15$. The black vertical lines shows the real values.

Developing statistical SDE inference procedures

The likelihood is only **explicit** given the **full** phylogenetic history S:

$$S = (D, M),$$

where M is missing data (extinct species, incomplete fossil record, ...).

EM algorithm

For unimodal likelihoods, its unique maximum can be found by maximizing iteratively

$$Q(\theta|\theta^{(i-1)}) = E[l_{D,M}(\theta)|D, \theta^{(i-1)}]$$

MCEM

$$Q(\theta|\theta_{(i)}) = \int_{\{\text{🌳},\text{🌳},\text{🌳},\dots\}} \log L(\theta|\text{🌳})\mathrm{d}\text{🌳} \longrightarrow \theta_{(i+1)} = \arg\max_{\theta} Q\left(\theta|\theta_{(i)}\right)$$