Statistical Methods Assignment V

Università della Svizzera italiana

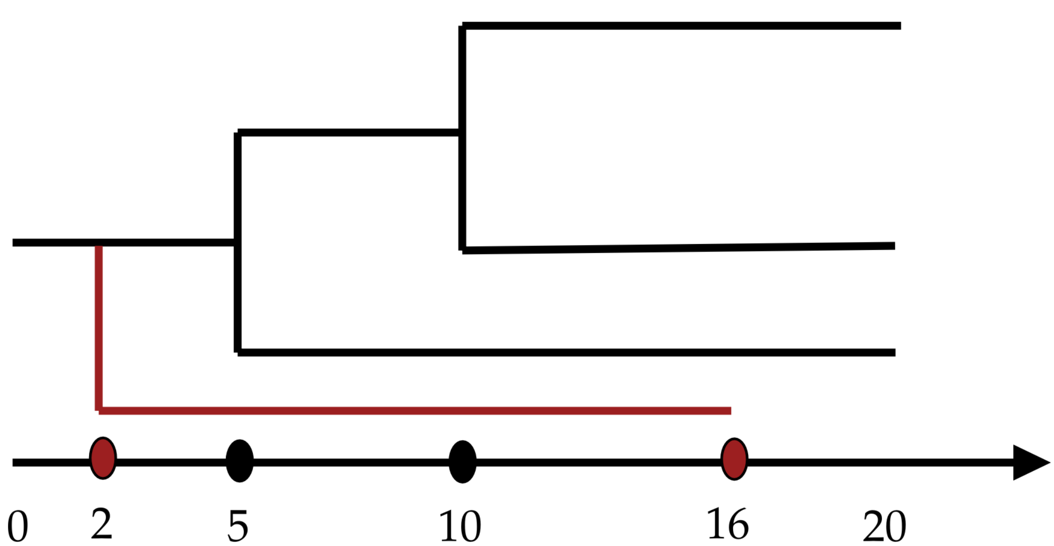
Due Date: 05/18/2024

[This assignment is to be completed using a Jupyter Notebook. Your code, figures, and analyses should be documented clearly within the notebook.]

Computational evolutionary processes

Evolutionary dynamics explore the underlying processes that shape the development and diversity of systems over time across a variety of fields. These dynamics are commonly modeled using stochastic processes, such as birth-death processes, where ’births’ and ’deaths’ metaphorically denote the emergence and disappearance of elements like languages, species, innovations, or financial products. Such models help illustrate the evolution of these diverse systems through time.

Consider the following simple evolutionary tree, where the x-axis represents time in millions of years:



Modeling

Consider a model that addresses the following rates:

- **Innovation rate** (λ): This rate is diversity-dependent and formulated as:

$$\lambda_t = \max\{(\lambda_0 - \beta_N N_t), 0\}$$

  where λ_0 is the baseline innovation rate under ideal conditions (no competition) and N_t is the number of branches at time t. The function max ensures that the innovation rate does not drop below zero.

- **Extinction rate** (μ): The extinction rate is assumed to be a constant value, μ_0, reflecting a baseline risk of extinction that does not vary with time:

$$\mu = \max\{\mu_0, 0\}$$

By adjusting the innovation rate based on diversity, the model accounts for competition and niche filling, which are pivotal in understanding evolutionary patterns. The constant extinction rate simplifies the complexity of the model while focusing on the dynamics of innovation.

1. Assuming that this process was generated under rates $\lambda_0 = 0.8$, $\beta_N = -0.075$ and $\mu_0 = 0.1$, calculate the probability of the evolutionary process described in the tree of the picture.

2. Now consider the case where the fossil record is missing. Assume that we know there is only one extinction and formulate the probability of the extant tree.

3. Use a Monte Carlo method, together with importance sampling, to estimate the integral above. To do that:

   (a) Use the uniform sampler. That is, sample the innovation time, the extinction time, and the allocation of the missing branches from uniform distributions.

   (b) We can obtain a numerical approximation by using the Monte-Carlo approach considering

$$\begin{aligned} f(x_{obs}|\theta) &= \int_{x \in \mathcal{X}(x_{obs})} f(x|\theta) dx \\ &= \int_{x \in \mathcal{X}(x_{obs})} \frac{f(x|\theta)}{f_m(x|\theta, x_{obs})} f_m(x|\theta, x_{obs}) dx \\ &\approx \frac{1}{M} \sum_{x_i \sim f_m(x|\theta, x_{obs})} \frac{f(x_i|\theta)}{f_m(x_i|\theta, x_{obs})} \end{aligned} \tag{1}$$

   for M the Monte-Carlo sample size and f_m is your sampler of the missing part of the full tree given an extant tree x_{obs}.

4. Now assume we do not know how many extincted branches are missing. Design an implement a stochastic method to calculate the probability of the number of missing branches.

5. Propose another sampler, designed by you, and compare the results.
6. Extend the model to non-constant extinction rate, perform simulations with both models and compare the results. Calculate the probabilities of points 3. and 4. for the non-constant extinction rate model.