Approach

Given a phylogenetic tree, we define the following Random Variables,

T_i : branching time i of the phylogenetic tree
X_i : binary variable equal to 1 if the process i is an speciation, and 0 if is an extinction.

Then, we define the phylogenetic tree as a data frame $Y = \{X_i, T_i\}$.

We suppose that the random variable T follows an exponential distribution with parameter $\nu = \sum \lambda_i + \mu_i$ where λ_i and μ_i are the speciation and extinction rates, and X_i follows a Bernoulli distribution with parameter $p_i = \lambda_i/\nu$ corresponding to the probability of speciation (or extinction) of specie i.

Thus, we write the likelihood of the phylogenetic tree as

$$L(Y; \lambda, \mu) = \prod_{i=1}^{N} \nu e^{-\nu t_i} p_i$$

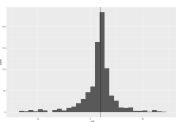where N es the number of branching times on the phylogenetc tree.

Some simulations

Moreover, we assumme the rates $\lambda = f(cov)$, $\mu = f(cov)$. In the simulation bellow we have a simple model $\lambda = exp\{\theta_0 + \theta_1 b\}$, $\mu = exp\{\varphi_0 + \varphi_1 b\}$

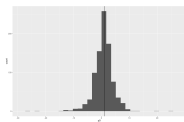| $n = 896$	real value	mean	median	min	max
θ_1	3.00	3.05	3.03	0.39	6.00
θ_2	4.00	0.61	3.85	-4136.88	949.77
ϕ_1	1.00	0.68	0.81	-26.92	24.46
ϕ_2	2.00	5.00	1.80	-2712.26	2164.94

(a) θ_0 (b) θ_1
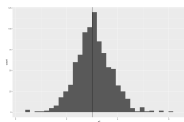
(c) φ_0 (d) φ_1
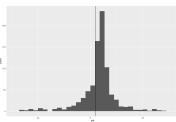
GLM & IWLS

Linear models of the form

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \beta_i; \qquad Y_i \sim N(\mu_i, \sigma^2),$$

are widely used for their well known estimation methods and computational convinience.

Advances in statistical theory and computers software allow us to use methods analogous to those developed for linear models in the following more general situation:

1. Response variables have distributions other than the Normal distribution (they may even be categorical rather than continuous).

2. Rlationship between the response and the explanatory variable need not be the simple linear form above.

Usually, for GLM, maximun likelihood estimators are obtained by an **iterative weighted least squares** procedure,

$$X^T W X b^{(m)} = X^T W z.$$