Al buio non si trova

Biostatistics in the 21st century[a]

Luiz Max Carvalho
lmax.fgv@gmail.com

FGV EMAp

[a] Available from https://github.com/maxbiostat/presentations/

Le dirò con due parole, chi son

Personal

- Born and raised in Petrópolis-RJ;
- Eldest of three kids;
- Married and father of a daughter;
- Mais Querido supporter.

Academic

- BSc in Microbiology & Immunology (UFRJ, 2012);
- PhD Evolutionary Biology (Edinburgh, 2018);
- Post doctoral researcher at ENSP/Fiocruz (2019);
- Lecturer (Assistant Professor) at EMAp since Jan/2020.

E che faccio

Applications of Statistics/Mathematics

Applications in Epidemiology, (Molecular) Biology, Ecology, Psychology, Linguistics, etc.

Applied Statistics

Markov Chain Monte Carlo, Model combination and selection, Statistical Phylogenetics.

That's like... Your opinion, man

Logarithmic Opinion pooling, aka log-linear mixtures[1]

$$\pi(\theta \mid \boldsymbol{\alpha}) = t(\boldsymbol{\alpha}) \prod_{i=1}^{K} f_i(\theta)^{\alpha_i}, \alpha_i \geq 0, \sum_{i=1}^{K} \alpha_i = 1.$$

Applications: combining forecast models, robust Bayesian inference.

- Hierachical approach: which prior $\pi_A(\boldsymbol{\alpha})$ will induce nice behaviour on both $p_A(\boldsymbol{\alpha} \mid \boldsymbol{y})$ and $p_T(\theta \mid \boldsymbol{y})$?
- Bayesian predictive synthesis: can we solve

$$\int_{\mathcal{X}} \alpha(y \mid \boldsymbol{x}) \prod_{j=1}^{J} h_j(x_j)\, d\mu(x_j) = \prod_{j=1}^{J} h_j(y)^{w_j}?$$

[1] https://arxiv.org/abs/1502.04206

Crystal ball

Next-generation epidemiological surveillance[2]

$$\tilde{f}(\boldsymbol{y}) = \mathbb{H}\left(\boldsymbol{M}(\boldsymbol{y}), \boldsymbol{\alpha}\right),$$

with $\boldsymbol{M}(\boldsymbol{y}) = \{m_1(\boldsymbol{y}), \ldots, m_K(\boldsymbol{y})\}$ a collection of forecasts.

- Can we build a unified framework to aggregate models from different groups, places, types?
- What form should the combining operator $\mathbb{H}$ take?
- Which proper scoring rules[3] to use?

[2] https://github.com/maxbiostat/EBDS_2021

[3] https://github.com/maxbiostat/proper_scoring_rules

I got the power

Normalised power prior[4]

$$\tilde{\pi}(\theta, a_0 \mid \boldsymbol{y}_0) = \frac{L(\boldsymbol{y}_0 \mid \theta)^{a_0} \pi(\theta \mid \eta) \pi_A(a_0 \mid \phi)}{c(a_0; \eta, \phi)}$$

Applications: clinical trials, quality control, policy-making.

- How pick π_A such that prediction error (say) is minimised?
- How to **efficiently** compute

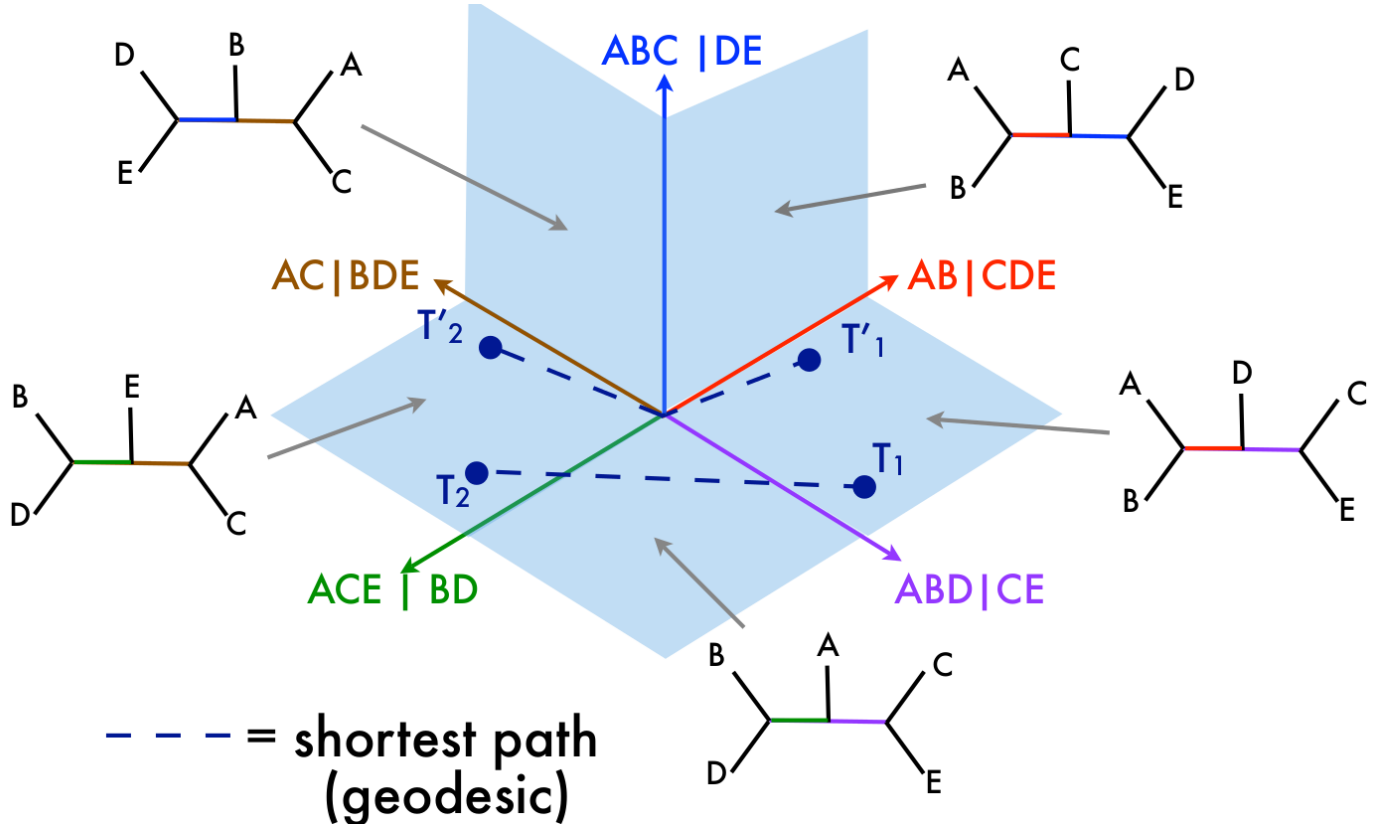
$$c(a_0; \eta, \phi) = \int_\Theta L(\boldsymbol{y}_0 \mid t)^{a_0} \pi(t \mid \eta) \, d\mu(t)$$

by leveraging its special properties as function of a_0?

[4] https://doi.org/10.1002/sim.9124

This place is weird

Traversing cubic complexes efficiently[5]



Applications: Molecular Epidemiology, Evolutionary Biology.

[5] https://youtu.be/h9bWRQ6aeKA

THE END