

Utilizando prioris de complexidade penalizada para estimação de tamanho de uma população a partir de filogenias

Cristiana Aparecida Nogueira Couto

Orientador: Luiz Max Carvalho

EMAp-FGV

25 de junho 2021



Introdução

Justificativa

Métodos

Cronograma

Considerações finais

Referências bibliográficas

1.

Um apanhado geral

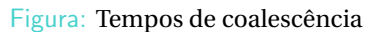


Tabela: Revisão de literatura

Conjunto de dados	Descrição do objetivo
63 sequências de HCV, vírus transmissor da hepatite C, amostrados em 1993 no Egito.	Investigar a dinâmica populacional do HCV no Egito (MININ; BLOOMQUIST; SUCHARD, 2008)
Três conjuntos de dados correspondendo a três épocas gripais, período recorrente anual caracterizado pelos surtos de influenza	Investigar a dinâmica populacional intra-sazonal da influenza (MININ; BLOOMQUIST; SUCHARD, 2008)
399 sequências de HIV-1, amostradas no Senegal entre 1990 e 2014 . Todas as 179 sequências são do subtipo CRF02_AG.	Inferir a função demográfica (DIDELOT; VOLZ, 2021)

2.

A motivação matemática

Qual é o problema e quais as motivações para investigá-lo?

Modelo Demográfico por Partes:

Seja uma árvore filogenética T com n folhas. Seja $s = (s_2, s_3, \dots, s_n)$ os intervalos de inter-coalescência de T , ou seja, intervalos entre os eventos de coalescência.

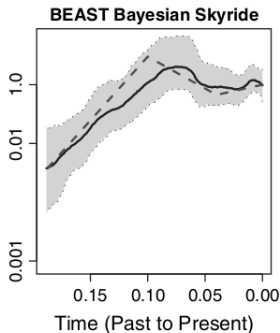
Seja θ o tamanho da população, onde $\theta = (\theta_2, \dots, \theta_n)$

Modelo Skyride

- ▶ Supõe que o tamanho da população muda suavemente ao longo do tempo;
- ▶ Define uma função priori no tamanho da população como um Processo Gaussiano suave.

Nesse modelo, precisamos inferir o parâmetro de precisão τ que é desconhecido.

Figura: Estimação do tamanho efetivo de uma população com efeito gargalo



Fonte: (MININ; BLOOMQUIST; SUCHARD, 2008)

Modelo Skyride

$$Pr(s|\theta) = \prod_{i=2}^n Pr(s_k|\theta_k) \quad (1)$$

Transformação $\gamma_k = \log(\theta_k)$, $k = 2, \dots, n$.

$$Pr(\gamma|\tau) \propto \tau^{\frac{n-2}{2}} \exp\left(\frac{-\tau}{2} \sum_{k=2}^{n-1} \frac{(\gamma_{k+1} - \gamma_k)^2}{2\theta}\right) \quad (2)$$

Fonte: (MININ; BLOOMQUIST; SUCHARD, 2008)

Objetivo

Entender a influência da escolha da distribuição a priori nas inferências sobre a dinâmica populacional obtidas a partir de filogenias. Em particular, das chamadas prioris de complexidade penalizada (Penalised Complexity), PC.

3.

Como fazer?

Experimentos e colocando a mão na massa.

Prioris de Complexidade Penalizada
(SIMPSON et al., 2017) propõem um conceito para a construção de distribuições a priori que são robustas, invariantes à reparametrização e baseada em alguns princípios.

Experimentos com dados simulados:

- ▶ População constante;
- ▶ População exponencial;
- ▶ População que teve um efeito gargalo, ou seja, um evento que reduz drasticamente o tamanho de uma população.

OS conjuntos de dados podem ser simulados com o pacote
phylodyn do **R**.

Experimentos com dados reais:

- ▶ Dengue do tipo 4;
- ▶ Influenza H3N2;
- ▶ Outros conjuntos de dados comumente usadas na literatura, como as sequências de HIV amostradas no Egito, por exemplo.

Figura: Cronograma de atividades mensais

Tarefas\ Mês	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Fundamentação teórica	x	x	x	x	x				
Estudar e compreender os principais conceitos de estatística bayesiana	x	x							
Compreender a teoria e utilização de priors de complexidade penalizada		x							
Aprender os conceitos de filogenética e filodinâmica		x	x						
Escrita do relatório de entrega referente a A2		x	x						
Apresentação 1			x						
Experimentos com os dados				x	x				
Expandir experimentos					x	x			
Escrita da monografia			x	x	x	x	x		
Revisão de português e ABNT do texto final							x	x	
Preparar a apresentação (slides)								x	x
Entrega do relatório									x
Apresentação final									x

Fonte: Elaborado pela autora

Obrigada pela atenção!

Contato:

cristiana.couto@fgv.edu.br



CARVALHO, L. M. F. de. A better prior for the precision parameter in skyride/grid/track. 2021.



DIDELLOT, X.; VOLZ, E. M. Maximum likelihood inference of pathogen population size history from a phylogeny. **bioRxiv**, Cold Spring Harbor Laboratory, 2021. Disponível em: <<https://www.biorxiv.org/content/early/2021/01/19/2021.01.18.427056>>.



KÜHNERT, D.; WU, C.-H.; DRUMMOND, A. J. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. **Infection, Genetics and Evolution**, Elsevier BV, v. 11, n. 8, p. 1825–1841, dez. 2011. Disponível em: <<https://doi.org/10.1016/j.meegid.2011.08.005>>.



MININ, V. N.; BLOOMQUIST, E. W.; SUCHARD, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. **Molecular Biology and Evolution**, Oxford University Press (OUP), v. 25, n. 7, p. 1459–1471, abr. 2008. Disponível em: <<https://doi.org/10.1093/molbev/msn090>>.



NORDBORG, M. Coalescent theory. **Handbook of statistical genetics**, Wiley Online Library, 2004.



SIMPSON, D. et al. Penalising model component complexity: A principled, practical approach to constructing priors. **Statistical Science**, Institute of Mathematical Statistics, v. 32, n. 1, fev. 2017. Disponível em: <<https://doi.org/10.1214/16-sts576>>.



VOLZ, E. M.; KOELLE, K.; BEDFORD, T. Viral phylodynamics. **PLoS Computational Biology**, Public Library of Science (PLOS), v. 9, n. 3, p. e1002947, mar. 2013. Disponível em: <<https://doi.org/10.1371/journal.pcbi.1002947>>.