



1st SouthStat Meeting

Encontro Sul Brasileiro de Estatística e Ciência de Dados

14 E 15 DE DEZEMBRO DE 2023. CURITIBA-PR.

Viés de amostragem nas análises filogeográficas

Vitória Silva Garcia¹

Universidade Federal do Rio Grande do Sul

RESUMO

A filogeografia tem como propósito caracterizar os processos de dispersão geográfica dos vírus e efetuar inferência sobre os eventos migratórios do passado. Para isso, são combinados dados de sequências genéticas, locais geográficos e as datas de amostragem. É possível fazer estimativas acerca da migração do vírus entre os locais geográficos e reconstruir o processo espacial de dispersão do vírus.

A filogeografia Bayesiana estuda as dispersões virais através de um modelo estocástico, mapeando as propagações dos vírus que estão sendo analisados. Desse modo é possível entender a origem da doença e a sua propagação nos locais geográficos.

Observa-se que essas análises podem ser afetadas por vieses de amostragem decorrentes de fatores que não estão relacionados ao processo de dispersão, como heterogeneidades espaciais e temporais devido à falta de recursos em diversas localidades para o sequenciamento de vírus. Esses vieses geralmente ocorrem durante a coleta da amostra. E assim, tem-se um recorte da amostra que não é representativo da população de estudo e por consequência ocorre um desequilíbrio na amostra.

Estudaremos as análises filogeográficas e o que ocorre com a estimação quando temos uma amostra aleatória com viés de amostragem, mensurando os efeitos que os vieses de amostragem possuem sobre a reconstrução filogeográfica e seus parâmetros.

Queremos investigar os diferentes graus de viés de amostragem, as diferentes estruturas de migração e como as estimativas comportam-se em face do viés. Utilizaremos um modelo que é uma cadeia de Markov na árvore filogeográfica, realizaremos inferência Bayesiana através do algoritmo de MCMC e estudaremos o comportamento do método por simulações de Monte Carlo.

Os resultados preliminares consistem em simulações de Monte Carlo que foram realizadas para estudar propriedades do processo gerador ligadas às distribuições dos estados nos nós externos da árvore filogenética (amostra). Constatou-se que conforme o número de nós externos aumenta juntamente com o comprimento total da árvore, a distribuição de equilíbrio da cadeia passa a ser atingida para taxas de migração menores e a variância dos nós internos é menor do que a variância dos nós externos da árvore. Nesses casos, temos que a diferença na variância entre os nós internos e externos é pequena.

Palavras-chave: Filogeografia; Viés de Amostragem; Cadeias de Markov; MCMC; Inferência Bayesiana.