



SALÃO DE INICIAÇÃO CIENTÍFICA XXVIII SIC

paz no plural



| | |
|-------------------|--|
| Evento | Salão UFRGS 2016: SIC - XXVIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS |
| Ano | 2016 |
| Local | Campus do Vale - UFRGS |
| Título | Análise estatística de correlações evolutivas estimadas pelo modelo filogenético de Variável Latente |
| Autor | VITÓRIA MARIA MARTINI WENDT |
| Orientador | GABRIELA BETTELLA CYBIS |

**Análise estatística de correlações evolutivas estimadas pelo
modelo filogenético de Variável Latente**

Aluna: Vitória Martini Wendt

Orientadora: Gabriela Bettella Cybis

Universidade Federal do Rio Grande do Sul

O modelo filogenético de variável latente pode ser usado na biologia evolutiva para estimar correlações entre traços fenotípicos enquanto controla para a história evolutiva dos indivíduos em estudo. Sua aplicação abrange análises em que são utilizados diferentes tipos de dados: contínuos, discretos binários e discretos múltiplos (ordenados ou não). Este trabalho tem como objetivo avaliar as propriedades estatísticas da estimação de correlações evolutivas calculadas através deste modelo. Para isto, foi desenvolvida uma rotina para simulação do modelo utilizando ferramentas estatísticas e computacionais apropriadas.

Seja $Y = (y_0, \dots, y_N)$ um vetor cujas entradas são os valores observados para a variável de interesse nos N indivíduos da amostra, $X = (x_0, \dots, x_N)$ um vetor cujas componentes representam as variáveis latentes do processo evolutivo e F uma filogenia (representação da história evolutiva dos N indivíduos). A variável X evolui através de F por movimento Browniano simulado pelo cálculo de sucessivas normais multivariadas. Ao final do movimento o valor da variável latente determinará o valor de Y , de modo que $Y = X$ quando Y contínuo. No caso em que Y é discreto, define-se o valor de Y de acordo com a posição de X em relação a um limiar. Uma rotina de simulação do processo descrito acima é construída na plataforma *R*.

A matriz de correlações estimada pelo modelo de variável latente é então obtida a partir da análise de Y por MCMC no software de inferência bayesiana para filogenias, BEAST. Com esta matriz é possível comparar as estimativas das correlações das variáveis geradas pelo modelo de variável latente com aquelas geradas pelo coeficiente de correlação de Pearson.

Até o momento foram confrontadas estimativas de correlações entre variáveis contínuas, binárias e, contínuas e binárias. Os resultados obtidos mostram que os estimadores de correlações calculados através do modelo de variável latente apresentam pouco ou nenhum vício. Além disso, a taxa de falso positivo, definida como a fração de simulações no cenário de independência em que a análise indicou uma correlação significativamente não nula (com intervalo de credibilidade de 95%), mostra um índice de 0.03 de falso positivo para correlações entre variáveis discretas enquanto, para variáveis de ambos os tipos o índice é de 0.12, chegando a 0.18 quando tratadas duas variáveis contínuas.