

MÉTODOS FILODINÂMICOS PARA ESTUDO DE CORRELAÇÕES EVOLUTIVAS



Lauren Vieira (BIC-UFRGS), Gabriela Bettella Cybis (Orientadora-UFRGS)
laurendiasalves@gmail.com, gabriela.cybis@ufrgs.br

PROBLEMA

Para a genética evolutiva a estimação de correlações entre características genóticas e/ou fenotípicas, tem sido um problema desafiador, o modelo de Variável Latente propõe uma solução para esta questão, avaliando combinações de características contínuas e discretas (binárias, com múltiplos estados ordenados ou não), simultaneamente controlando as correlações quanto ao histórico evolutivo do conjunto de indivíduos estudados.

Enquanto métodos mais conhecidos, não consideram o histórico evolutivo, este modelo propõe uma abordagem Bayesiana filogenética para lidar com este problema, utilizando uma variável latente para análise dos dados que assumimos ter evoluído através de difusão browniana sob a filogenia (conhecida ou não) que associa os indivíduos em estudo. Por ser um modelo recente, poucas de suas propriedades foram avaliadas.

Neste trabalho desejamos avaliar a precisão do modelo ao encontrar associação entre uma variável contínua e uma variável discreta.

OBJETIVOS

1. Avaliar a capacidade do modelo, dados os diferentes possíveis ordenamentos, comparando com o Coeficiente de Correlação de Spearman em condições semelhantes
2. Avaliar o efeito de especificação incorreta do modelo de ordenamento, através da variabilidade das estimativas obtidas durante a análise

SIMULAÇÃO

Neste estudo optamos por gerar amostras de tamanho $n = 10$, para avaliar associação entre uma variável contínua e uma ordenada com múltiplos estados, estas amostras foram geradas através de difusão browniana em uma filogenia com correlação 0.9, utilizando para isto uma rotina do programa R. Geramos conjunto de 1000 amostras, para esta simulação e analisamos os dados gerados de acordo com os três ordenamentos (o original 012 e suas duas variantes 120 e 201). Para para cada um destes alinhamentos foram gerados conjuntos de 1000 elementos.

Os elementos gerados contavam com uma filogenia fixa sob qual os dados haviam sido gerados. Estes dados por sua vez foram avaliados através do BEAST, programa que correlação evolutiva através do Modelo de Variável Latente, entre outros modelos.

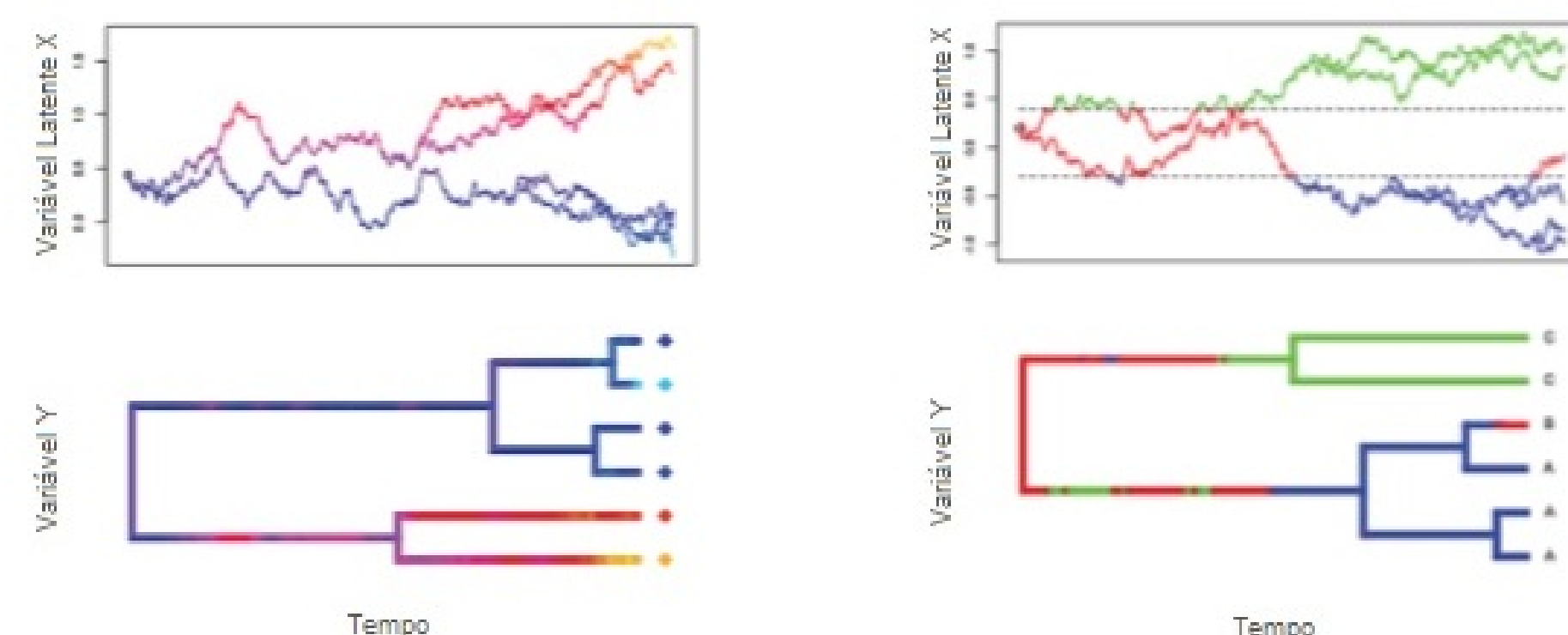
Para fins de comparação quanto a eficiência na estimação da associação entre as variáveis, tomamos a o Coeficiente de Correlação de Spearman, que avalia a associação entre variáveis, sem o pressuposto de continuidade e não condiciona os resultados ao histórico evolutivo compartilhado pelos indivíduos nas amostras.

REFERENCIAS

Cybis, G.B., Sinsheimer, J.S., Bedford, T., Mather, A.E., Lemey, P. and Suchard, M.A., Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. The Annals of Applied Statistics, 9(2): 969-991. 2015.

MODELO DE VARIÁVEL LATENTE

O Modelo de Variável Latente modela variáveis fenotípicas Y , cujos valores são determinados por uma variável latente X , que evolui através de difusão browniana com matriz de precisão Σ^{-1} , sob a filogenia τ que conecta os indivíduos da amostra.



A densidade da variável latente X , pode ser obtida através da convolução de normais multivariadas decorrentes do processo de difusão browniana sob τ .

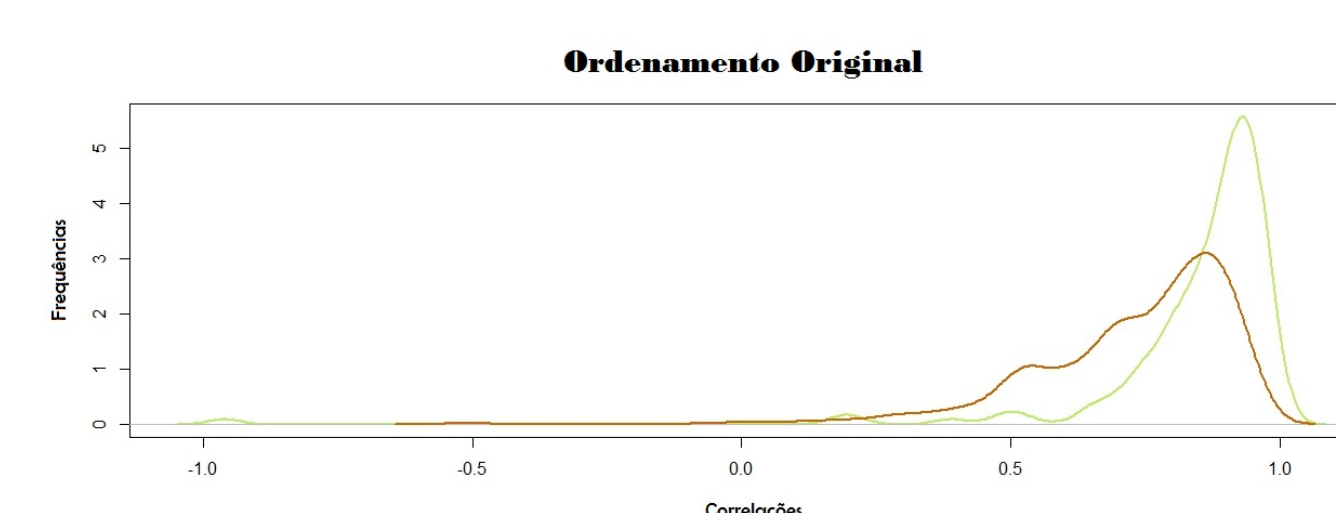
$$P(X, Y|\tau, \Sigma^{-1}, \theta) = P(X|\tau, \Sigma^{-1}, \theta) * P(Y|X)$$

Onde θ são os parâmetros do modelo filogenético.

A inferência do modelo é feita através de MCMC (Monte Carlo Markov Chain), onde a estimação da correlação entre as variáveis é feita através de Σ^{-1} . (Cybis et al.2015).

ORDENAMENTO ORIGINAL

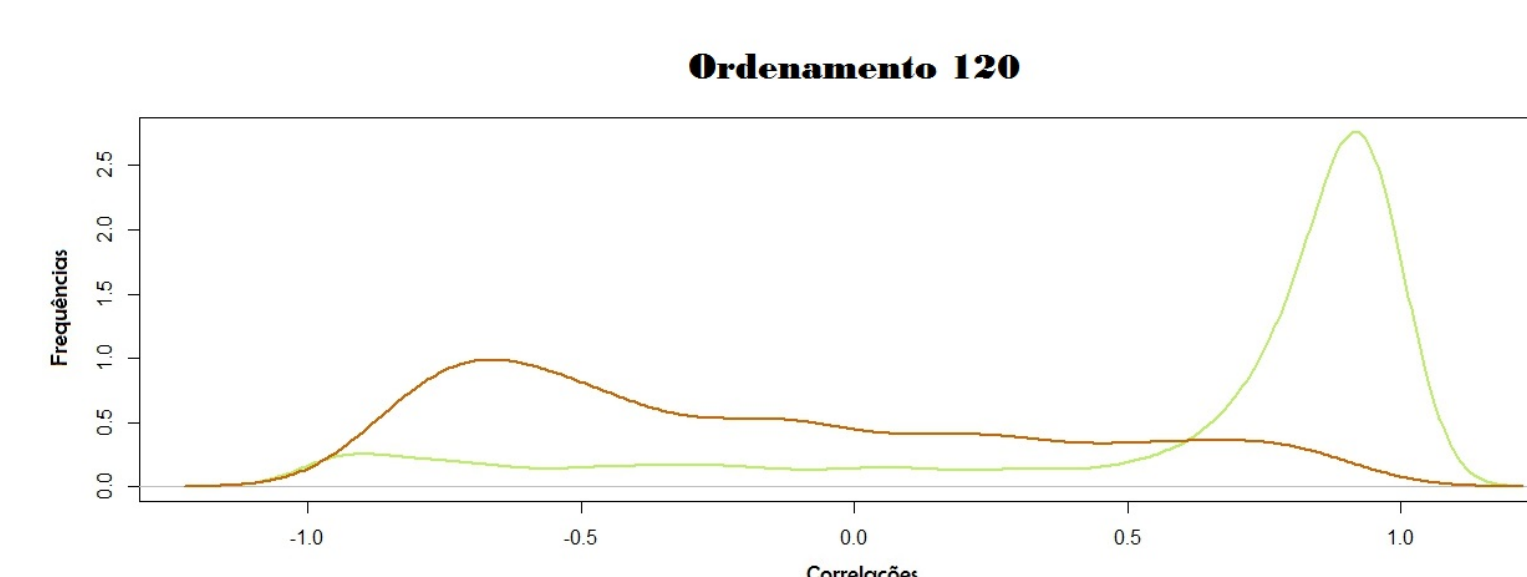
Como era esperado, dado o ordenamento original, a eficiência do Modelo de Variável Latente (em verde) comparado ao Coeficiente de Correlação de Spearman (em marrom), encontra associação mais fraca entre as variáveis. Porém a variabilidade das estimativas do modelo é bem alta (0.607) comparada a das estimativas por Spearman(0.1932).



Ainda assim quando observados apenas os valores significantes ($\alpha = 0.05$ bicaudal), os valores médios das correlações bem como a variabilidade das estimativas é semelhante (Modelo de Variável Latente $\mu = 0.89$ e $\sigma = 0.0869$, Spearman $\mu = 0.8275$ e $\sigma = 0.0743$)

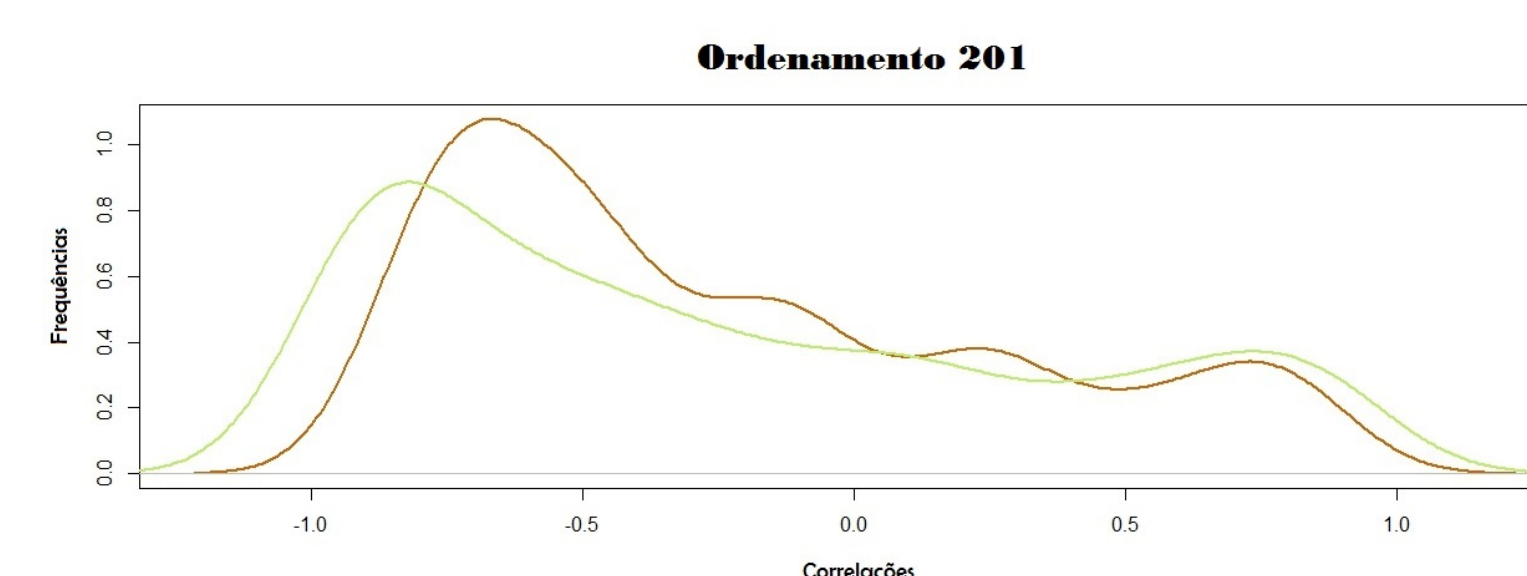
OUTROS ORDENAMENTOS

A despeito do que se esperava o modelo de variável latente (em verde), conseguiu estimar a associação entre as variáveis em questão, quando o ordenamento avaliado foi 120.



Diferente do Coeficiente de Spearman (em marrom), o modelo estimou associação significativa para as simulações apenas 79%, enquanto por Spearman o mesmo ocorreu para 29.3%

Quando observado ordenamento201, os resultados de ambos os métodos são semelhantes quanto a estimação das associação, ambos encontram associação fraca entre as variáveis.



CONCLUSÕES

Verificou se que o Modelo de Variável Latente consegue estimar correlação entre uma variável contínua e uma discreta com múltiplos estados ordenados. Se pode perceber que dado ordenamento 120, diferente do que se esperava o modelo identificou correlação forte entre as variáveis. Resultado que se manteve, quando observamos as estimativas da média e da moda a posteriori deste modelo.

Foi observado ainda, que o Modelo de Variável Latente, de um modo geral retorna estimativas maior variabilidade que o Coeficiente de Spearman, ainda que este seja mais afetado dado o ordenamento 120, os resultados obtidos por ambos os métodos são semelhantes.