

Estimando matrizes de covariância com modelos hierárquicos

Estratégia geral

Utilizar métodos hierárquicos para estimar matrizes de covariância e médias fenotípicas de forma filogeneticamente estruturada.

Resumo do Modelo

Dada um filogenia e um conjunto de medidas quantitativas, podemos escrever um modelo hierárquico para estimar médias e covariâncias entre as medidas utilizando aproximações gaussianas.

Passos do modelo:

1. Notação:

- z_{ij} matriz de medidas, i populações e j caracteres.
- θ_i médias das populações
- σ_i matriz de covariância das populações
- Θ_k hiper parâmetros, médias nos nós da filogenia
- C_{mn} hiper parâmetro, tensor de covariância das médias, não entendi direito essa parte ainda (pode incluir distância filogenética)
- Σ_k hiper parâmetros, matriz de covariâncias nos nós da filogenia

2. Verossimilhança das populações

- gaussiana básica: $p(z|\theta, \sigma) = N(z|\theta, \sigma) = \sum_{ij} \exp\left(-\frac{1}{2} * (z_{ij} - \theta_i)\sigma_i^{-1}(z_{ij} - \theta_i)\right)$

3. Priors hierárquicos

- Cada nó k da filogenia equivale a um conjunto de prior para os ramos acima dele
- Para um terminal i, os priors de θ_i e σ_i seriam da forma $N(\theta_i|\Theta_k, C_{mn})$ e $Wis(\sigma_i|\Sigma_k)$
- Para os nós internos, o processo se repete até a raiz, onde um prior não informativo (mas integrável) deve ser definido. Esse é o ponto mais subjetivo do processo e seria bom fazer análise de sensibilidade aqui.

4. Integração

- Depois de montado o modelo, como produto de priors e verossimilhança, ele é integrado numericamente via monte carlo, tomando amostras de todos os parâmetros $(\theta, \sigma, \Theta, \Sigma)$ seguindo a distribuição de probabilidade a posteriori

5. A partir das amostras os parâmetros são estimados.

6. PROFIT!

Referências

A base do modelo é tratar as estimativas como uma sequencia de modelos hierárquicos aninhados na filogenia. As estimativas são todas gaussianas. Mesmo assim tem partes difíceis de estimar, relacionados com os priors das matrizes de covariância e os passos de monte carlo associados a elas. Cada passo deve conter uma matriz simétrica, e isso é difícil. É o velho problema de amostrar direito o espaço de matrizes positivas definidas.

Básico:

1. Modelos hierárquicos: Capítulos 5, 15, 19 do Bayesian Data Analysis
2. MCMC e estimativas: Capítulos 10 a 14, especialmente do 14.6 em diante, onde tem a parte de covariâncias não uniformes e correlação. Com sorte a gente não vai precisar implementar essas técnicas na mão.

Pacotes

- [PyMC](#)