

PROGRAMA DO WORKSHOP

Descrição geral

O objetivo principal do workshop é entender como uma análise usando Markov Chain Monte Carlo (MCMC) funciona e então explorar aplicações em modelos de evolução de caracteres em árvores filogenéticas. Nas primeiras partes do workshop vamos usar modelos simples (com poucos parâmetros) como, por exemplo, regressão linear e média de uma distribuição. Vamos interpretar resultados de análises, fazer experimentos para entender conceitos e utilizar analogias e simplificações sempre que necessário. O objetivo central do curso é entender os passos do MCMC e as ideias em que a análise se baseia. Sendo assim, durante o curso não vamos explorar em detalhe a matemática e estatística formal por trás do método. Por fim vamos aplicar os conceitos explorados para estimar parâmetros de modelos mais complexos como os métodos comparativos filogenéticos.

Na primeira parte do curso vamos explorar conceitos básicos da estimativa de modelos, entender a diferença entre um “point estimate” como a Máxima Verossimilhança e uma distribuição posterior gerada por um MCMC. Vamos também analisar as características de superfícies de verossimilhança (“likelihood surface”).

Na segunda parte vamos mergulhar um pouco mais fundo. Vamos entender a ideia geral de um MCMC e comparar com estratégias usadas para estimar a máxima verossimilhança. Por fim, vamos caminhar passo-a-passo por cada um dos componentes de uma análise de busca usando MCMC. Vamos explorar os resultados de análises assumindo diferentes distribuições a priori (priors) para entender como os priors podem influenciar as estimativas.

Na terceira parte vamos explorar uma série de modelos comparativos filogenéticos com ênfase nos modelos que descrevem padrões de evolução de características ao longo de uma árvore filogenética. Vamos utilizar estes modelos como estudo de caso para explorar convergência de cadeias, seleção de modelos e outros tópicos importantes do uso de estatística Bayesiana.

Resultados esperados do curso

Espera-se que o aluno que não possui experiência prévia com análises usando MCMC seja capaz de entender a lógica e as aplicações deste método. O aluno será apresentado à terminologia utilizada com frequência em publicações científicas e terá a oportunidade de associar estes termos com os tópicos abordados no curso.

O aluno que já tem alguma prática em análises utilizando MCMC poderá entender melhor o papel de cada um dos componentes e parâmetros destas análises. Também espera-se que o aluno desenvolva sua capacidade de detectar potenciais problemas em cadeias de MCMC e conduzir análises com senso crítico.

O aluno com conhecimentos de programação poderá aplicar os algoritmos apresentados no workshop para implementar programas com sua própria análise de MCMC.

Tópicos abordados no curso

Introdução

1. O que é um modelo?
 1. Parâmetros de um modelo.
 2. Estimativa dos parâmetros de um modelo.
 3. Solução exata por derivação.
 1. Exemplo simples de Máxima Verossimilhança.
 4. Estimativa usando minimização.
 1. Trabalhar o conceito de busca ('Quente ou frio?')
2. Função de verossimilhança.
 1. O que é uma função de verossimilhança?
 2. Usando a função de verossimilhança para buscar a Máxima Verossimilhança.
 3. Exemplos de busca de verossimilhança.
3. Superfície de verossimilhança.
 1. Exemplos de superfícies em trabalhos publicados.
 2. O conceito de pico.
 3. O conceito de vale.
 4. O conceito de região plana.

Comportamento da estimativa usando máxima verossimilhança.

1. Pontos de partida das buscas.
 1. Greedy searches e ótimos locais.
 2. Múltiplas estimativas para verificar convergência em máxima verossimilhança.

A ideia por trás do MCMC.

1. Analogia: Explorando uma paisagem desconhecida.
 1. Robôs em Marte.
 2. Distribuição posterior em função da superfície de Verossimilhança.
 3. Analogia: Escolhendo chapéus.
 4. Porque escolher somente um chapéu se podemos levar mais de um? (Point estimate *versus* posterior distributions)

MCMC passo a passo

1. Visão geral do algoritmo.
 1. Conceito de Markov Chain.
 2. Conceito de convergência.
 3. Papel da função de verossimilhança e prior.
2. Algoritmo do MCMC.
 1. Estado inicial.
 2. Distribuições e esquemas de movimentos ("proposal schemes").
 3. Metropolis e Metropolis-Hastings.
 4. Mistura e número de gerações.
3. Explorando a posterior.
 1. E aí!? Convergiu mesmo?
 2. Melhorando a mistura da cadeia.
 3. Como evitar desperdiçar seu trabalho?

Modelos comparativos filogenéticos

1. Estimando parâmetros de evolução de caracteres em filogenias.
 1. Apresentação dos modelos.
 2. Interpretação das funções de verossimilhança.
 3. Implementação do MCMC.
 4. Explorando a posterior dos parâmetros.
2. Comparando modelos e testando hipóteses.
 1. Usando Bayes Factors.
 1. 'Marginal likelihood' e 'stepping-stone'
 2. Comparando distribuições posteriores.
 3. Simulações e aproximações.
 1. Máxima Verossimilhança e Bayesiana como complementos.
 2. Posterior predictive checks.