Análise de verossimilhança bayesiana

- Agora vamos ver o modo bayesiano de ajuste de modelos. O teorema de Bayes diz que:

$$prob(\vec{\alpha}|X_i) \propto \mathcal{L}(X_i|\vec{\alpha})prob(\vec{\alpha})$$

  de modo que, claramente, a verossimilhança continua sendo muito importante.
- A principal diferença do método da máxima verossimilhança, além do fator *a priori* é que o resultado, o posterior, é uma função que tem os parâmetros $\vec{\alpha}$ como variáveis, não apenas o valor mais provável e sua incerteza.
- As duas maiores vantagens do método bayesiano são a possibilidade de lidar com parâmetros indesejáveis via marginalização e o uso da *evidência* ou *fator da Bayes* para escolher entre modelos.

Análise de verossimilhança bayesiana

- Uma outro produto útil do modo bayesiano é a distribuição assintótica da função de verossimilhança. Esta é uma gaussiana multivariada centrada nas MLE $\hat{\vec{\alpha}}$. Sua matriz de covariância é a inversa da matriz de informação de Fisher:

$$F = -\begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_1^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_1 \partial \alpha_2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_1 \partial \alpha_3} & \cdots \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_2 \partial \alpha_1} & \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_2^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_2 \partial \alpha_3} & \cdots \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_3 \partial \alpha_1} & \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_3 \partial \alpha_2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha_3^2} & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}_{\vec{\alpha}=\hat{\vec{\alpha}}}$$

- A verossimilhança então é:

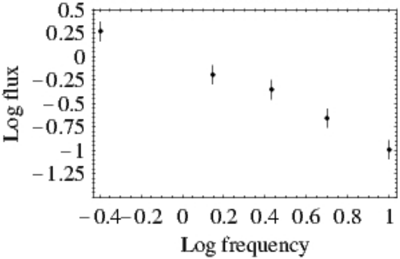
$$\mathcal{L}(\vec{\alpha}|X_i) = \mathcal{L}(\hat{\vec{\alpha}}|X_i) exp\left[-\frac{1}{2}(\vec{\alpha}-\hat{\vec{\alpha}})^T F(\vec{\alpha}-\hat{\vec{\alpha}})\right],$$

onde T denota matriz transposta. Essa é a chamada aproximação de Laplace.

Análise de verossimilhança bayesiana

-

Exemplo W&J P. 143

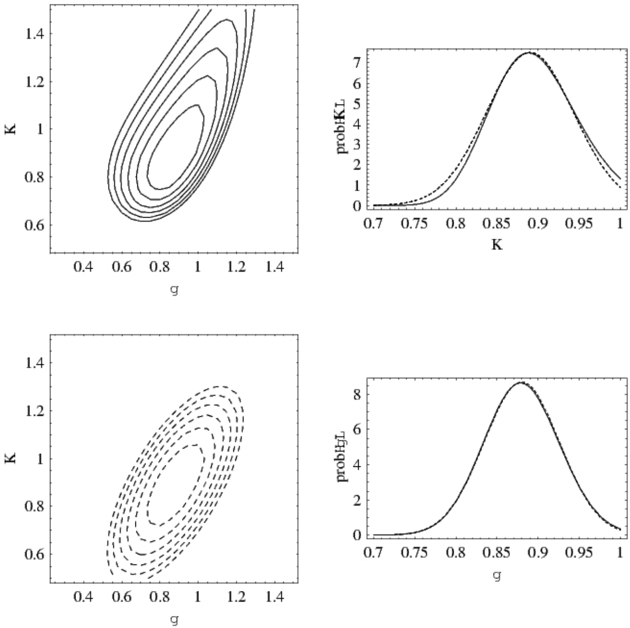


-
-

$$S = Kf^{-\gamma}$$

-

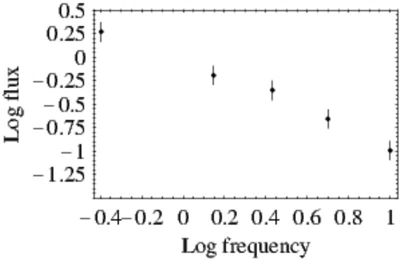
$$\mathcal{L} = \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\epsilon K f^{-\gamma}} exp\left[-\frac{(S_i - Kf_i^{-\gamma})^2}{2(\epsilon K f_i^{-\gamma})^2}\right]$$

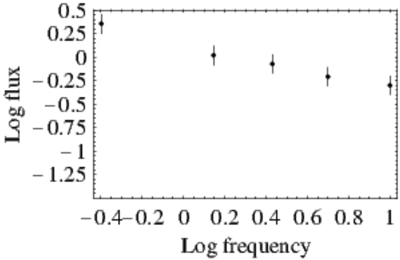

1.4
1.2
K
1
0.8
0.6
0.4 0.6 0.8 1 1.2 1.4
g
7
6
5
4
3
2
1
0
probHKL
0.7 0.75 0.8 0.85 0.9 0.95 1
K
1.4
1.2
K
1
0.8
0.6
0.4 0.6 0.8 1 1.2 1.4
g
8
6
4
2
0
probHgL
0.7 0.75 0.8 0.85 0.9 0.95 1
g

Análise de verossimilhança bayesiana

-

Exemplo W&J P. 145





-
-

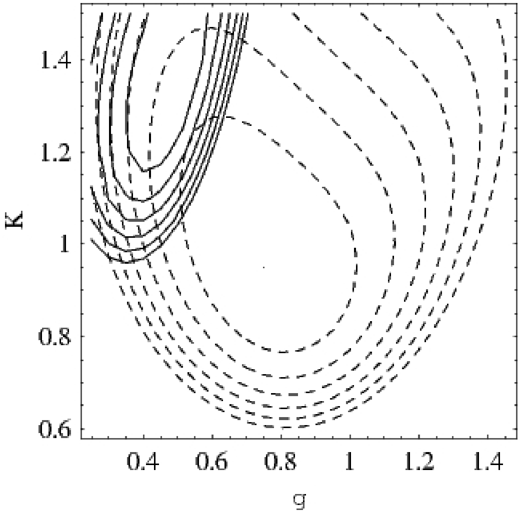
$$\mathcal{L} = \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\epsilon K f^{-\gamma}} exp\left[-\frac{[S_i - (\beta + K f_i^{-\gamma})]^2}{2(\epsilon K f_i^{-\gamma})^2}\right]$$

-

$$prob(\beta) = \frac{1}{\sqrt{2\pi\epsilon}} exp\left[-\frac{(\beta - 0.4)^2}{2\epsilon^2}\right]$$

Análise de verossimilhança bayesiana

Exemplo W&J P. 145



MCMC: Markov Chain Monte Carlo

- Amostrar uma distribuição de probabilidades arbitrária
- Eficiente para amostrar distribuições complexas
- Computacionalmente intensivo

MCMC: Markov Chain Monte Carlo

- Objetivo do MCMC: gerar uma sequência de pontos no espaço de parâmetros cuja distribuição é $P(x)$

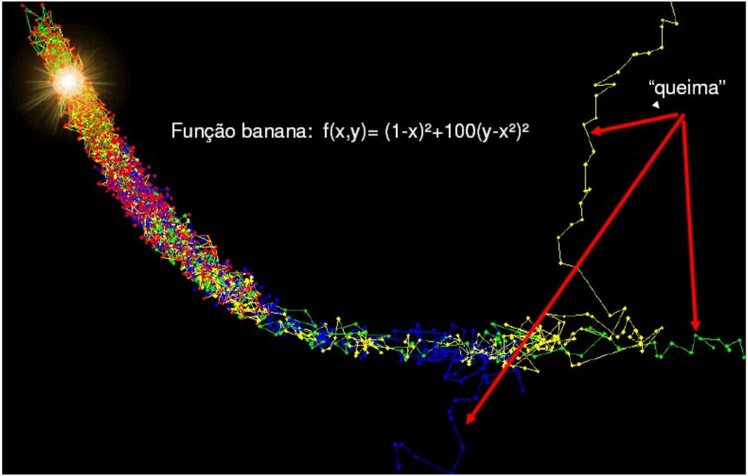
- Processo de Markov: a próxima amostra depende da atual mas não das anteriores

- A geração de novas amostras deve obedecer o princípio do "balanceamento detalhado": a probabilidade de se ir de um ponto x_1 no espaço de parâmetros a um ponto x_2 é igual à probabilidade de se ir de x_2 a x_1

MCMC: Markov Chain Monte Carlo

- O algoritmo de Metropolis:
  - Comece num ponto x_1 e determine um novo ponto x_2:
    $$x_2 = x_1 + \epsilon$$
  - O valor de ϵ deve ser sorteado, via o método de Monte Carlo, com uma distribuição que é usualmente gaussiana - *random walk*.
  - Se $P(x_2) > P(x_1)$ aceita o novo ponto
  - Se $P(x_2) < P(x_1)$ aceita com probabilidade $P(x_2)/P(x_1)$ (Via Monte Carlo)
  - Se um novo ponto é aceito, faz-se $x_1 \leftarrow x_2$
  - Determina-se um novo x_2 e repete-se o processo

MCMC: Markov Chain Monte Carlo



MCMC: Markov Chain Monte Carlo

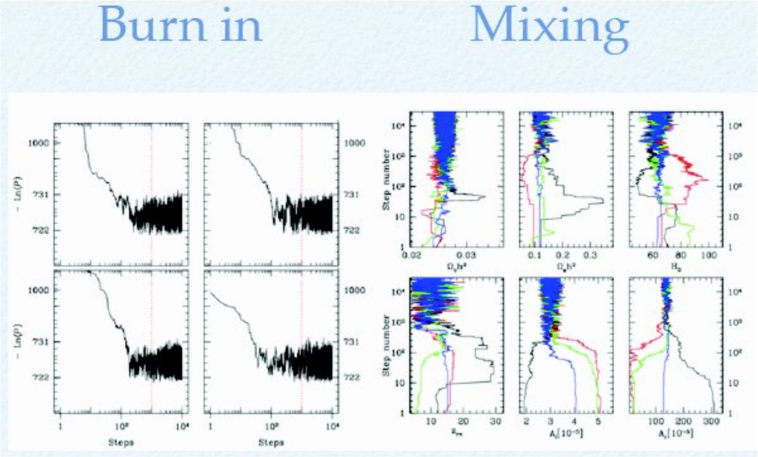
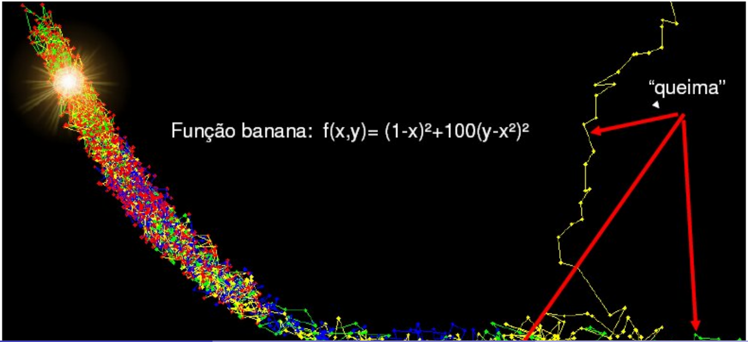


Figure : Diagnóstico de MCMC.

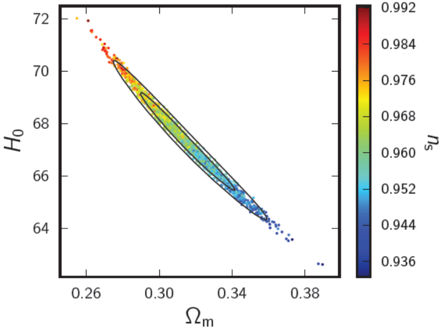
MCMC: Markov Chain Monte Carlo

- Na prática:
  - Trabalha-se com várias cadeias (>4)
  - Desprezam-se as primeiras amostras de cada cadeia ("queima")
  - Aplicam-se testes para verificar se as cadeias convergiram
  - Verifica-se a taxa de aceitação de pontos. Valores ente 15% e 50% são considerados saudáveis.



MCMC: Markov Chain Monte Carlo

- Obtendo-se amostras do posterior via MCMC pode-se resumir a inferência via:
  - Melhor solução: média, mediana, moda
  - Regiões de confiança ("erros")



Bootstrap e jacknife: bootstrap

- Bootstrap e jacknife são técnicas para estimar incertezas nos parâmetros obtidos nos modelos.
- Bootstrap: Dada uma amostra com N pontos:
  1. "Rotule cada ponto"
  2. Gere uma nova amostra com N pontos tomados aleatoriamente da amostra original
  3. Recalcule os parâmetros com a nova amostra
  4. Repita a operação o máximo número de vezes possível.

- A distribuição dos parâmetros gerados por esse procedimento é uma excelente aproximação para as incertezas envolvidas no processo.

Bootstrap e jacknife: jacknife

- Jacknife: Suponha que estamos interessados numa função $f(X_1, X_2, ..., X_N)$ que é um bom estimador de uma parâmetro α:

$$\hat{\alpha} = f(X_1, X_2, ..., X_N)$$

- A j-ésima estimativa parcial é obtida removendo-se o j-ézimo elemento no conjunto de dados:

$$\hat{\alpha} = f(X_1, X_2, ...X_{j-1}, X_{j+1}, ..., X_N)$$

  o que nos dá N estimativas parciais.

- O próximo (e crucial) passo é definir os pseudo valores:

$$\hat{\alpha}_j^* = N\hat{\alpha} - (N-1)\hat{\alpha}_j$$

- Finalmente o estimativa via *jacknife* de α será a média dos pseudo valores

$$\hat{\alpha}^* = \frac{1}{N}\sum_{j=1}^{N} \hat{\alpha}_j^*$$

Bootstrap e jacknife: jacknife

- Bootstrap e jacknife são os exemplos mais comuns de técnicas de validação (*cross-validation*) nas quais o conjunto de dados é dividido em partes e os resultados da análise da partes são comparados por consistência. Há uma série de variantes sofisticadas dessa ideia.