

Aula 7 – Astro-estatística

Tópicos:

- 1 Revisão de probabilidades
- 2 Inferência bayesianas e análise de dados
- 3 Inferência dos parâmetros de um modelo
- 4 Inferência dos parâmetros de uma distribuição gaussiana
- 5 Inferência dos parâmetros de uma distribuição gaussiana bidimensional
- 6 Aplicação: a estatística de Riess et al. (1998)
- 7 Dark Energy Task Force (DETF)
- 8 MCMC: Monte Carlo Markov Chain
- 9 O método de Monte Carlo (revisão)
- 10 Referências
 - ▶ Sivia e Skilling, *Data Analysis - A Bayesian Tutorial*
 - ▶ Wall e Jenkins, *Practical Statistics for Astronomers*
 - ▶ I. M. Sobol, *The Monte Carlo Method*

- A Cosmologia contemporânea tem usado extensivamente a estatística Bayesiana.
- Mais de que discutir os méritos e deméritos das estatísticas Bayesiana e clássica (frequentista) vamos procurar entendê-las.
- A aula de hoje é sobre os fundamentos da estatística Bayesiana e como ela pode ser aplicada à análise de dados em cosmologia.

- O que é a probabilidade? *Formalização numérica de seu grau de intensidade em qualquer crença.* (Independentemente dessa definição subjetiva, é crítico que duas pessoas com as mesmas informações chegarão às mesmas probabilidades.)
- Axiomas da probabilidade:
 - 1 Qualquer evento aleatório tem uma probabilidade $P(A)$ entre 0 e 1;
 - 2 Um evento certo tem $P(A) = 1$;
 - 3 Se A e B são eventos mutuamente exclusivos, então $P(A \text{ ou } B) = P(A) + P(B)$.
- Esses axiomas são suficientes para todo o desenvolvimento da teoria matemática de probabilidade.

Revisão: condicionalidade e independência

- Dois eventos são ditos *independentes* se a probabilidade de um não é afetada pelo que possamos saber a respeito da outra. Nesse caso, segue (de modo não trivial) dos axiomas que:

$$P(A \text{ e } B) = P(A) \times P(B) \quad (3)$$

- No caso em que a independência não ocorre o que se quer saber é a *probabilidade condicional*: a probabilidade de A dado que conhecemos B . A definição disso é:

$$P(A|B) = \frac{P(A \text{ e } B)}{P(B)} \quad (4)$$

(tudo o que estiver a direita da barra “|” é suposto conhecido)

- Se A e B forem independentes a informação sobre se B ocorreu ou não não deve afetar nossas crenças sobre A . Assim sendo $P(A|B) = P(A)$ e temos novamente que $P(A \text{ e } B) = P(A) P(B)$

- Suponha que existam várias probabilidades para o evento B , então temos que:

$$P(A) = \sum_i P(A|B_i)P(B_i) \quad (5)$$

- A pode ser o parâmetro do modelo que se quer encontrar e os B 's, por exemplo, parâmetros instrumentais que não são de interesse.
- Conhecendo as probabilidades $P(B_i)$ podemos nos livrar desses *parâmetros irrelevantes* ou *nuisance parameters* através de uma soma ou integração.
- Esse processo é conhecido como *marginalização*.

Revisão: o teorema de Bayes

- O teorema de Bayes é uma igualdade simples que vem da afirmação de que $P(A \text{ e } B) = P(B \text{ e } A)$:

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}, \quad (6)$$

no qual o denominador é a probabilidade total.

- Esse teorema é útil quando interpretado como uma regra para indução: os dados e o evento A são considerados como sucessores de B , o grau de crença anterior a realização do experimento.

- Consideremos um conjunto de dados D que queremos explicar com um modelo H que tem alguns parâmetros x
- O teorema de Bayes pode ser escrito como

$$P(x|D, H) = \frac{P(D|x, H) \times P(x|H)}{P(D|H)},$$

onde cada termo tem um nome:

- ▶ $P(x|D, H)$: **Posterior**, ou probabilidade *a posteriori* dos parâmetros
- ▶ $P(D|x, H)$: Função de **verossimilhança** dos dados (*likelihood*)
- ▶ $P(x|H)$: **Prior**, ou probabilidade *a priori* dos valores dos parâmetros
- ▶ $P(D|H)$: **Evidência** dos dados

$$P(D|H) = \int P(D|x, H) \times P(x|H) dx$$

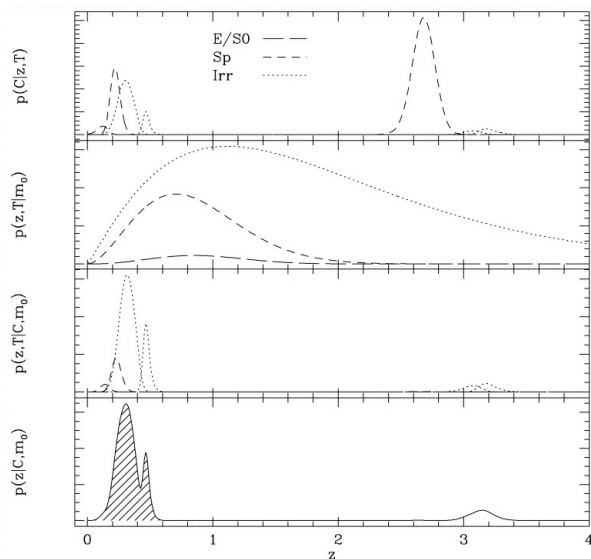
- Apesar da aparente trivialidade, o teorema de Bayes permite: uma abordagem lógica de alguns problemas importantes para a análise e interpretação de dados:
 - ▶ Estimativa de parâmetros de modelos
 - ▶ Comparação de modelos
 - ▶ Testes de modelos

- Estimativa de parâmetros de modelos:

$$P(x|D, H) = \frac{P(D|x, H) \times P(x|H)}{P(D|H)} \propto P(D|x, H) \times P(x|H)$$

- ▶ Os parâmetros são selecionados maximizando o posterior
- ▶ A evidência não é importante nesse tipo de problema
- ▶ Para um prior uniforme ($P(x|H) = \text{const.}$) a melhor estimativa é a que maximiza a verossimilhança (\rightarrow método da máxima verossimilhança)
- ▶ Erros podem ser estimados como níveis de confiança do posterior

Exemplo: *Bayesian Photometric Redshift Estimation* (Benitez 2000)



- Verossimilhança
 C -cores; z -redshift; T -tipo
espectral
- Prior
 m_0 -magnitude na banda de
referência
- Posterior
- Posterior marginalizado

- Estimativa de parâmetros de modelos:

$$P(x|D, H) \propto P(D|x, H) \times P(x|H)$$

a evidência não é importante neste tipo de problema pois não depende de x

- Vamos resumir o resultado da determinação de um parâmetro com dois números:
a melhor estimativa e uma medida de sua confiabilidade (o “erro”)
→ método da máxima verossimilhança

Estimativa de parâmetros de modelos

- Estimativa de parâmetros de modelos:

$$P(x|D, H) \propto P(D|x, H) \times P(x|H)$$

- Como o posterior é uma medida do nosso grau de plausibilidade no valor de um parâmetro, a melhor estimativa é o valor x_0 que maximiza o posterior:

$$\left. \frac{dP(x|D, H)}{dx} \right|_{x_0} = 0$$

e a condição de máximo

$$\left. \frac{d^2 P(x|D, H)}{dx^2} \right|_{x_0} < 0$$

- Cuidado:** o posterior pode ser multimodal.

Problema:

- Temos um conjunto de dados sobre uma variável x , $D = \{x_k\}$ com $k = 1, \dots, N$; vamos supor que estão distribuídos como uma gaussiana de largura σ ; Qual é o melhor valor de μ e qual é nossa confiança nele?
- Modelo para os dados:
 $x_k = \mu + \text{ruído gaussiano de largura } \sigma$

$$P(x_k | \mu, \sigma, H) \propto \exp \left[-\frac{(x_k - \mu)^2}{2\sigma^2} \right]$$

- Teorema de Bayes:

$$P(\mu|D, \sigma, H) = \frac{P(D|\mu, \sigma, H)P(\mu|\sigma, H)}{P(D|H)}$$

- A verossimilhança dos dados é

$$P(D|\mu, \sigma, H) = P(\{x_k\}|\mu, \sigma, H)$$

- Supondo que as medidas $\{x_k\}$ são independentes; nesse caso, a probabilidade dos dados é o produto das probabilidades das medidas:

$$P(\{x_k\}|\mu, \sigma, H) = \prod_{k=1}^N P(x_k|\mu, \sigma, H)$$

O prior

- Vamos atribuir uma pdf uniforme para o prior de μ :

$$P(\mu|\sigma, H) = P(\mu|H) = (\mu_{\max} - \mu_{\min})^{-1}$$

se $\mu_{\min} < \mu < \mu_{\max}$, e

$$P(\mu|\sigma, H) = P(\mu|H) = 0$$

fora desse intervalo

Estimativa de parâmetros de modelos: Gaussiana 1D

- Usando o teorema de Bayes e considerando o logaritmo do posterior,

$$\mathcal{L} \equiv \ln P(\mu|D, \sigma, H) = \text{constante} - \sum_{k=1}^N \frac{(x_k - \mu)^2}{2\sigma^2}$$

onde a constante inclui todos os termos que não dependem de μ

- Note a semelhança do termo acima com o nosso “querido” χ^2
- O posterior é nulo fora do intervalo $[\mu_{min}, \mu_{max}]$: pontos fora do intervalo não são considerados – vamos supor que nenhum ponto é excluído

Estimativa de parâmetros de modelos: Gaussiana 1D

- Logaritmo do posterior:

$$\mathcal{L} = \ln P(\mu|D, \sigma, H) = \text{constante} - \sum_{k=1}^N \frac{(x_k - \mu)^2}{2\sigma^2}$$

- A melhor estimativa, μ_0 , é a que maximiza o posterior e, portanto,

$$\left. \frac{d\mathcal{L}}{d\mu} \right|_{\mu_0} = \sum_{k=1}^N \frac{(x_k - \mu_0)}{\sigma^2} = 0$$

- ou

$$\mu_0 = \frac{1}{N} \sum_{k=1}^N x_k$$

isto é, a melhor estimativa é igual à média aritmética dos valores das medidas

- Note que a melhor estimativa não depende dos erros (σ)

Estimativa de parâmetros de modelos: Gaussiana 1D

- A confiança em μ_0 depende da largura do posterior, cuja variância (ou *desvio quadrático médio*, rms) é

$$\langle (\mu - \mu_0)^2 \rangle = \int (\mu - \mu_0)^2 P(\mu|D, H) d\mu = \frac{\sigma^2}{N}$$

- A raiz quadrada da variância é o *desvio padrão* de μ
- A largura do posterior está associada ao inverso da derivada segunda de L :

$$\left. \frac{d^2 \mathcal{L}}{d\mu^2} \right|_{\mu_0} = - \sum_{k=1}^N \frac{1}{\sigma^2} = - \frac{N}{\sigma^2}$$

- O resultado final da inferência pode então ser expresso como

$$\mu = \mu_0 \pm \frac{\sigma}{\sqrt{N}}$$

Estimativa de parâmetros de modelos: Gaussiana 1D

- Qual o papel do prior nesse caso?
- Se o intervalo $[\mu_{min}, \mu_{max}]$ cobre todos os dados, a inferência não depende de μ_{min} e μ_{max} e maximizar o posterior é equivalente a maximizar a verossimilhança
- O método da máxima verossimilhança equivalente ao uso de um prior uniforme, não informativo: nesse caso o valor de μ_0 depende só dos dados
- Imagine agora que você ache que μ_0 está num intervalo relativamente estreito: nesse caso sua inferência vai depender mais fortemente do prior

Estimativa de parâmetros de modelos: Gaussiana 2D

- Vamos supor agora que nosso modelo tem 2 parâmetros, X e Y
- O melhor valor dos parâmetros (X_0, Y_0) é obtido maximizando o posterior ou seu logaritmo:

$$\left. \frac{\partial \mathcal{L}}{\partial X} \right|_{X_0} = 0 \quad \text{e} \quad \left. \frac{\partial \mathcal{L}}{\partial Y} \right|_{Y_0} = 0$$

- Aproximando L por uma gaussiana em torno de (X_0, Y_0) temos:

$$\begin{aligned} \mathcal{L} = \mathcal{L}(X_0, Y_0) + \frac{1}{2} \left[\left(\frac{\partial^2 \mathcal{L}}{\partial X^2} \right)_0 (X - X_0)^2 + \left(\frac{\partial^2 \mathcal{L}}{\partial Y^2} \right)_0 (Y - Y_0)^2 \right. \\ \left. + 2 \left(\frac{\partial^2 \mathcal{L}}{\partial X \partial Y} \right)_0 (X - X_0)(Y - Y_0) \right] + \dots = \\ = \mathcal{L}_0 + Q + \dots \end{aligned}$$

Estimativa de parâmetros de modelos: Gaussiana 2D

- Fazendo uma aproximação gaussiana em torno de (X_0, Y_0) , temos:

$$\mathcal{L} = \mathcal{L}_0 + Q$$

sendo que Q , em notação matricial, fica:

$$Q = \mathbf{W} \mathbf{F}(\mathbf{W}_0) \mathbf{W}^T$$

onde

$\mathbf{W} = [X - X_0 \quad Y - Y_0]$ - vetor linha com os parâmetros

$$\mathbf{F}(\mathbf{W}_0) = - \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$

e

$$A = \left(\frac{\partial^2 \mathcal{L}}{\partial X^2} \right)_0 \quad B = \left(\frac{\partial^2 \mathcal{L}}{\partial Y^2} \right)_0 \quad C = \left(\frac{\partial^2 \mathcal{L}}{\partial X \partial Y} \right)_0$$

- $\mathbf{F}(\mathbf{W}_0)$ é a matriz 2×2 das derivadas segundas e é chamada de Matriz de Fisher (ou o Hessiano calculado nos pontos de máximo)

Estimativa de parâmetros de modelos: Gaussiana 2D

- Se estamos interessados apenas em X , podemos marginalizar sobre Y :

$$P(X|D, H) = \int P(X, Y|D, H) dY$$

com $P(X, Y|D, H) \propto \exp(\mathcal{L}) \propto \exp(-Q/2)$

- Isto é diferente de fixar $Y = Y_0$!
- Pode-se verificar que nesse caso temos que

$$P(X|D, H) \propto \exp \left[\frac{1}{2} \left(\frac{AB - C^2}{B} \right) (X - X_0)^2 \right]$$

- ou seja, a melhor estimativa de X é X_0 e o erro associado é

$$\sigma_x = \sqrt{\frac{-B}{AB - C^2}}$$

- Erro em X_0 :

$$\sigma_x = \sqrt{\frac{-B}{AB - C^2}}$$

- Erro em Y_0 :

$$\sigma_y = \sqrt{\frac{-A}{AB - C^2}}$$

- *Covariância* entre X e Y é

$$\begin{aligned} \langle (X - X_0)(Y - Y_0) \rangle &= \int (X - X_0)(Y - Y_0) P(X, Y | D, H) dX dY = \\ &= \sigma_{xy}^2 = \frac{C}{AB - C^2} \end{aligned}$$

Matriz de covariância:

$$\begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} = \frac{1}{AB - C^2} \begin{bmatrix} -B & C \\ C & -A \end{bmatrix} = - \begin{bmatrix} A & C \\ C & B \end{bmatrix}^{-1}$$

- Note que $AB - C^2$ é o determinante da matriz das derivadas
- A matriz de covariância é o negativo do inverso da matriz de derivadas segundas
- $C = 0$ implica em $\sigma_{xy} = 0$: os parâmetros não são correlacionados
- $C = \pm\sqrt{AB}$: a matriz de derivadas segundas é singular (determinante = 0); nesse caso X e Y estão relacionados linearmente
- O coeficiente de correlação: $\rho = \sigma_{xy}/(\sigma_x\sigma_y)$

Estimativa de parâmetros de modelos: Gaussiana 2D

A elipse de erro:

- Os parâmetros de elipse de erro são calculadas da seguinte forma (Coe 2009, arXiv:0906.4123v1):

$$a^2 = \frac{\sigma_x^2 + \sigma_y^2}{2} + \sqrt{\frac{(\sigma_x^2 - \sigma_y^2)^2}{4} + \sigma_{xy}^2}$$

$$b^2 = \frac{\sigma_x^2 + \sigma_y^2}{2} - \sqrt{\frac{(\sigma_x^2 - \sigma_y^2)^2}{4} + \sigma_{xy}^2}$$

$$\tan 2\theta = \frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}$$

- O tamanho desses eixos é então multiplicado por um coeficiente $\alpha = \sqrt{\Delta\chi^2}$ que depende do nível de confiança:

σ	CL	$\Delta\chi^2$	α
1	68.3%	2.3	1.52
2	95.4%	6.17	2.48
3	99.7%	11.8	3.44

- A área dessa elipse é dada por:

$$A = \pi(\Delta\chi^2)ab = \pi\sigma_x\sigma_y\sqrt{1 - \rho}$$

Aplicação: a estatística de Riess et al. (1998)

- Riess et al., 1998, ApJ, 116, 1009:
Observational evidence from SNe for an accelerating universe and a cosmological constant
- Modelo cosmológico Λ CDM
- Dados $\{\mu_0\}$: conjunto dos módulos de distância para uma amostra de N SNe Ia:

$$\mu = 5 \log D_I(z) + 25$$

onde

$$D_I = D_I(z, \Omega_m, \Omega_\lambda, H_0)$$

- Parâmetros do modelo: $\Omega_m, \Omega_\lambda, H_0$

Aplicação: a estatística de Riess et al. (1998)

Qual é o melhor modelo cosmológico que se pode inferir dos dados de SN?

- Posterior dos parâmetros, dado os dados $\{\mu_0\}$

$$P(\Omega_m, \Omega_\lambda, H_0 | \{\mu_0\}) = \frac{P(\{\mu_0\} | \Omega_m, \Omega_\lambda, H_0) P(\Omega_m, \Omega_\lambda, H_0)}{P(\{\mu_0\})}$$

- Como não queremos usar vínculos *a priori* sobre os parâmetros, vamos supor que o prior $P(\Omega_m, \Omega_\lambda, H_0)$ é constante

- Logo,

$$P(\Omega_m, \Omega_\lambda, H_0 | \{\mu_0\}) \propto P(\{\mu_0\} | \Omega_m, \Omega_\lambda, H_0)$$

- O melhor modelo, nesse caso, é o que maximiza a verossimilhança

Aplicação: a estatística de Riess et al. (1998)

Verossimilhança do modelo: $P(\{\mu_0\}|\Omega_m, \Omega_\lambda, H_0)$

- Supondo que as medidas são independentes:

$$P(\{\mu_0\}|\Omega_m, \Omega_\lambda, H_0) = \prod_{k=1}^N P(\mu_k|\Omega_m, \Omega_\lambda, H_0)$$

- Supondo que os erros têm distribuição gaussiana:

$$P(\mu_i|\Omega_m, \Omega_\lambda, H_0) = \frac{1}{\sqrt{2\pi(\sigma_{\mu,i}^2 + \sigma_v^2)}} \times \\ \exp \left\{ -\frac{[\mu(z_i, \Omega_m, \Omega_\lambda, H_0) - \mu_i]^2}{2(\sigma_{\mu,i}^2 + \sigma_v^2)} \right\}$$

onde $\sigma_{\mu,i}$ e σ_v são os erros no módulo de distância e no redshift (medido em unidades de módulo de distância)

Aplicação: a estatística de Riess et al. (1998)

Verossimilhança do modelo: $P(\{\mu_0\}|\Omega_m, \Omega_\lambda, H_0)$

- Logo,

$$P(\{\mu_0\}|\Omega_m, \Omega_\lambda, H_0) \propto \exp\left(-\frac{\chi^2}{2}\right)$$

onde

$$\chi^2 = \sum_{i=1}^N \frac{[\mu(z_i, \Omega_m, \Omega_\lambda, H_0) - \mu_i]^2}{\sigma_{\mu,i}^2 + \sigma_v^2}$$

- O posterior normalizado será:

$$P(\Omega_m, \Omega_\lambda, H_0|\{\mu_0\}) = \frac{\exp\left(-\frac{\chi^2}{2}\right)}{\int dH_0 \int d\Omega_\lambda \int \exp\left(-\frac{\chi^2}{2}\right) d\Omega_m}$$

- Os parâmetros são obtidos maximizando essa expressão

Aplicação: a estatística de Riess et al. (1998)

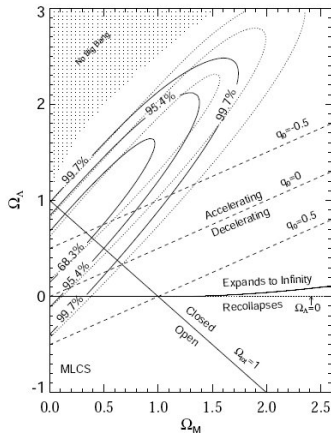


FIG. 6.—Joint confidence intervals for (Ω_M, Ω_A) from SNe Ia. The solid contours are results from the MLCS method applied to well-observed SNe Ia light curves together with the snapshot method (Riess et al. 1998b) applied to incomplete SNe Ia light curves. The dotted contours are for the same objects excluding the unclassified SN 1997ck ($z = 0.97$). Regions representing specific cosmological scenarios are illustrated. Contours are closed by their intersection with the line $\Omega_M = 0$.

Aplicação: a estatística de Riess et al. (1998)

Exemplos de análises

- Pode-se examinar o comportamento da solução para Ω_m e Ω_λ , independentemente do valor de H_0 , marginalizando sobre esse parâmetro:

$$P(\Omega_m, \Omega_\lambda | \{\mu_0\}) = \int P(\Omega_m, \Omega_\lambda, H_0 | \{\mu_0\}) dH_0$$

- Pode-se estimar a probabilidade de Ω_λ ser positivo, marginalizando sobre Ω_m e só considerando os valores positivos de Ω_λ :

$$P(\Omega_\lambda > 0 | \{\mu_0\}) = \int_0^\infty d\Omega_\lambda \int P(\Omega_m, \Omega_\lambda | \{\mu_0\}) d\Omega_m$$

- Para combinar os resultados desses experimento com os obtidos por CMB ou medidas independentes de H_0 , por exemplo poderíamos simplesmente colocar esses resultados como priores para os parâmetros.

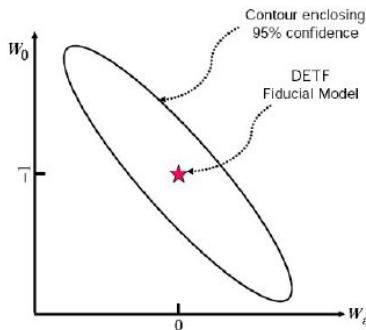
Dark Energy Task Force

Report of the ark Energy Task Force: *arXiv:astro-ph/0609591*

- Equação de estado da energia escura: $p = w\rho$

$$w(a) = w_0 + (1 - a)w_a$$

- w_0 : valor atual de w ; w_a : parametriza a evolução de w (Constante cosmológica: $w_0 = -1$, $w_a = 0$)
- Figura de mérito: inverso da área contida pelo limite de confiança de 95% no plano $w_0 \times w_a$



- Propriedades úteis da matriz de Fisher:
 - ▶ Marginalização: remova da matriz de covariância a linha e a coluna da variável a ser marginalizada e calcule o inverso para obter a nova matriz de Fisher
 - ▶ Fixar parâmetro: remova a variável correspondente da matriz de Fisher
 - ▶ Prior de um parâmetro: se o prior for uma gaussiana de largura σ , adicione $1/\sigma^2$ ao termo diagonal correspondente a esta variável na matriz de Fisher
 - ▶ Combinar conjuntos de dados diferentes: simplesmente adicione as matrizes de Fisher: $F = F_1 + F_2$; qualquer marginalização deve ser feita após a adição
- Por outro lado, todo o formalismo da matriz de Fisher supõe posteriores gaussianos, unimodais. Como lidar com posteriores mais complexos?
- MCMC (*Markov Chain Monte Carlo*)

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

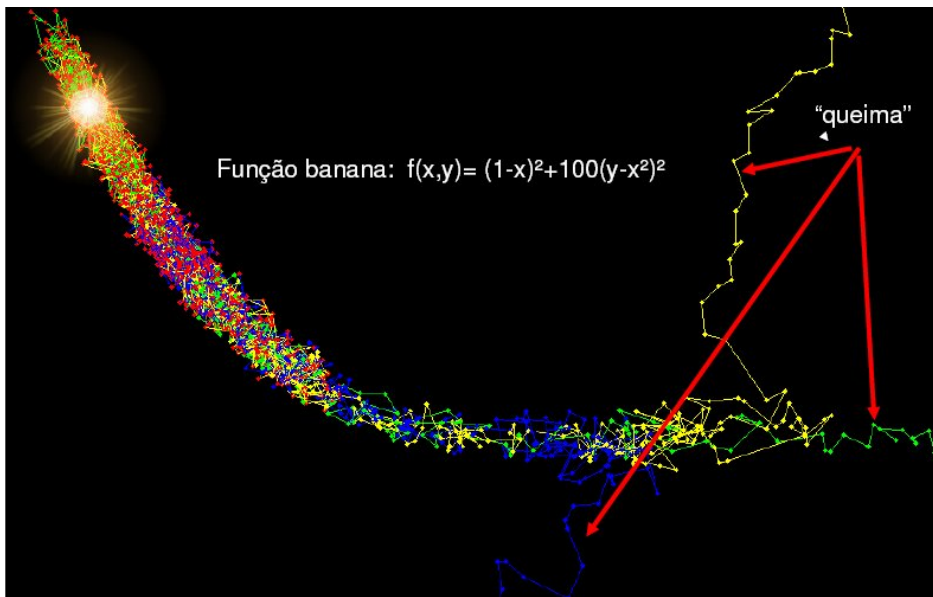
- O algoritmo de Metropolis:

- ▶ Comece num ponto x_1 e determine um novo ponto x_2 :

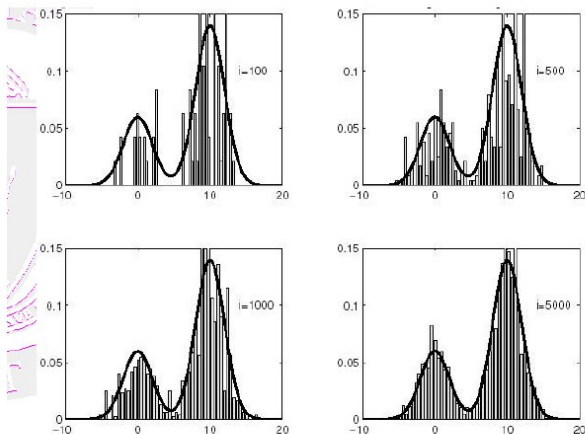
$$x_2 = x_1 + \epsilon$$

- ▶ 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



$$q(x^*|x) \sim N(x^{(i)}, 100)$$

$$p(x) \sim 0.3 \exp(-0.2x^2) + 0.7 \exp(-0.2(x-10)^2)$$

Burn in

Mixing

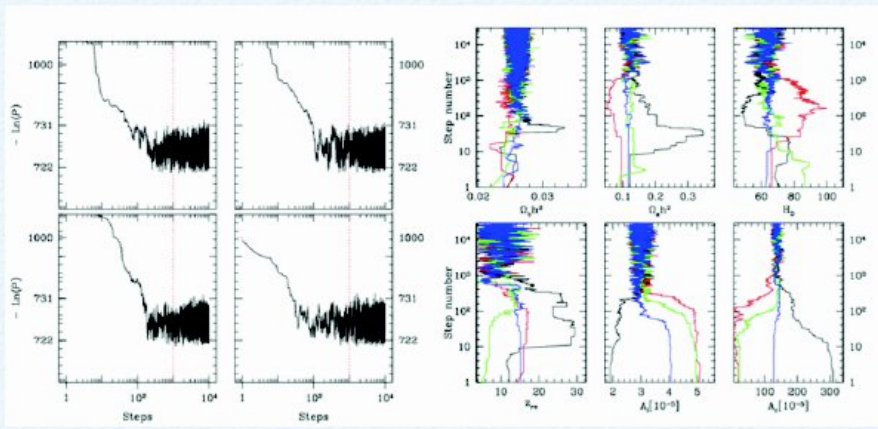
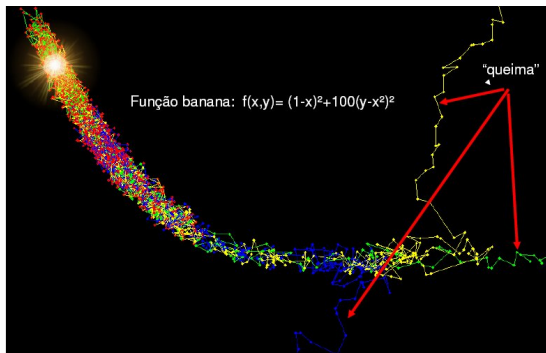


Figure: Diagnóstico de MCMC.

MCMC: Markov Chain Monte Carlo

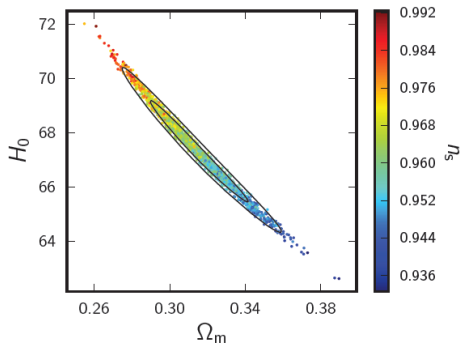
- Na prática:

- ▶ Trabalha-se com várias cadeias (>4)
- ▶ Desprezam-se as primeiras amostras de cada cadeia (“queima”)
- ▶ Aplicam-se testes para verificar se as cadeias convergiram



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”)



Geradores de Distribuição de Monte Carlo

Referência para essa subseção: I. M. Sobol: *The Monte Carlo Method*

Geração de uma variável aleatória contínua

- Vamos supor que queremos gerar os valores de uma variável aleatória ξ distribuída no intervalo (a, b) com densidade $p(x)$.
- Pode ser demonstrado que os valores de ξ podem ser encontrados a partir da equação:

$$\int_a^{\xi} p(x)dx = \gamma \quad (7)$$

- Para provar isso consideremos a função:

$$y = \int_a^x p(x)dx$$

- Das propriedades gerais da PDF decorre que:

$$y(a) = 0, \quad y(b) = 1,$$

- e que a derivada:

$$y'(x) = p(x) > 0.$$

Geração de uma variável aleatória contínua

- Seleccionemos agora um intervalo arbitrário (a', b') , dentro de (a, b) , onde

$$y(a') < y < y(b')$$

e

$$a' < x < b'$$

- Consequentemente, se ξ pertence ao intervalo $a' < x < b'$, então γ pertence ao intervalo $y(a') < y < y(b')$, e vice-versa.
- Assim:

$$P\{a' < \xi < b'\} = P\{y(a') < \gamma < y(b')\}$$

- Como γ é uniformemente distribuído em $(0, 1)$,

$$P\{y(a') < \gamma < y(b')\} = y(b') - y(a') = \int_{a'}^{b'} p(x) dx$$

- Assim sendo:

$$P\{a' < \xi < b'\} = \int_{a'}^{b'} p(x) dx$$

- e isso significa precisamente que a variável aleatória ξ , que é a raiz da equação 7, tem uma função densidade de probabilidade $p(x)$

Geração de uma variável aleatória contínua

- **Exemplo:** Uma variável aleatória η será considerada *uniformemente distribuída no intervalo* (a, b) se sua densidade no intervalo for constante:

$$p(x) = 1/(b - a) \text{ para } a < x < b.$$

- Para sortear o valor de η aplica-se a equação 7:

$$\int_a^\eta \frac{dx}{b - a} = \gamma$$

- Resolvendo-se a integral:

$$\frac{\eta - a}{b - a} = \gamma$$

- então:

$$\eta = a + \gamma(b - a) \tag{8}$$

Geração de uma distribuição normal.



$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta} e^{-\frac{x^2}{2}} dx = \gamma$$

- Dada uma tabela de valores de ζ calculados anteriormente (*look-up table*) é simples demonstrar que:

$$\zeta' = a + \sigma\zeta \tag{9}$$

será também normal, com média = a e variância = σ^2