Método dos processos Gaussianos: Análise de dados em cosmologia

Kevin Mota da Costa

UFES

Julho/2023

Definição de regressão linear

Dado um conjunto de vetores $\{\mathbf{x}, \mathbf{y}\}$ se pretende encontrar um função tal que $\mathbf{y} = f(\mathbf{x}) + \epsilon$

Definição de regressão linear

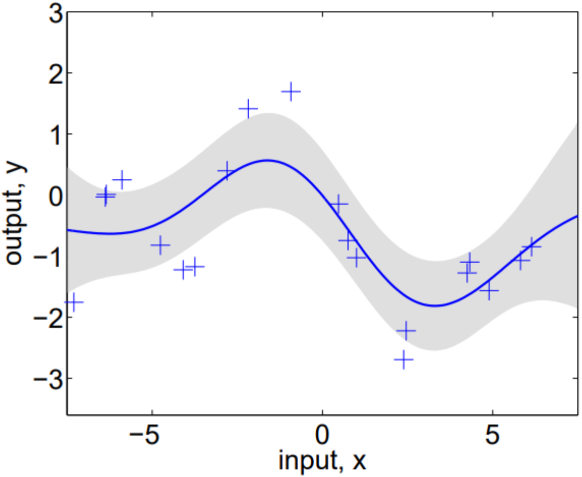


Figure: C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning

Teorema do limite central

O teorema do limite central da probabilidade afirma que dado um conjunto $X_1, X_2, \ldots, X_n$ de n variáveis aleatórias, então

$$\lim_{n\ grande} S_n \sim \mathcal{N}(\mu, \Sigma) \tag{1}$$

Assumimos então

$$\epsilon \sim \mathcal{N}(\mu, \Sigma) \tag{2}$$

Likelihood

A definição do modelo para o ruido junto com a definição da função $f(\mathbf{x})$ da origem ao conceito de *likelihood*

$$P(\mathbf{y}|\mathbf{x},\mathbf{w}) = \prod_{i=1}^{n} P(\mathbf{y}_i|\mathbf{x}_i,\mathbf{w}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{1}{2}\frac{(\mathbf{y}_i - \mathbf{x}_i{}^T\mathbf{w})^2}{\sigma_n{}^2}\right)$$

$$P(\mathbf{y}|\mathbf{x},\mathbf{w}) = \frac{1}{(2\pi\sigma_n{}^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma_n{}^2}|\mathbf{y} - \mathbf{X}^T\mathbf{w}|\right) = \mathcal{N}(\mathbf{X}^T\mathbf{w}, \sigma_n{}^2 I_n)$$

A *likelihood* é a distribuição de probabilidade das observações dado o modelo

Inferencia bayesiana

A inferência bayesiana é baseada na distribuição de probabilidade posterior sobre os parâmetros atavés do Teorema de Bayes

$$\text{posterior} = \frac{\textit{likelihood} \cdot \text{prior}}{\text{marginal likelihood}}$$

$$P(\mathbf{w}|\mathbf{y},\mathbf{x}) = \frac{P(\mathbf{y}|\mathbf{x},\mathbf{w}) \cdot P(\mathbf{w})}{P(\mathbf{y}|\mathbf{x})}$$

$$P(\mathbf{y}|\mathbf{x}) = \int P(\mathbf{y}|\mathbf{x},\mathbf{w}) d\mathbf{w}$$

$$P(\mathbf{w}|\mathbf{y},\mathbf{x}) \sim \mathcal{N}\left(\frac{1}{\sigma_n^2} A^{-1}\mathbf{x}\mathbf{y}, A^{-1}\right)$$

$$A = \sigma_n^{-2}\mathbf{x}\mathbf{x}^T + \Sigma^{-1}$$

Inferencia bayesiana

Para realizar predições sobre os o vetor de entrada $\mathbf{x}^*$ dados os vetores de treinamento $\{x, y\}$ aplicamos

$$P(f^*|x^*, x, y) = \int P(f^*|x^*, w)P(w|x, y)dw$$

$$P(f^*|x^*, x, y) = \mathcal{N}\left(\frac{1}{{\sigma_n}^2}{x^*}^T A^{-1} X y, {x^*}^T A^{-1} x^*\right)$$

O método dos processos gaussianos

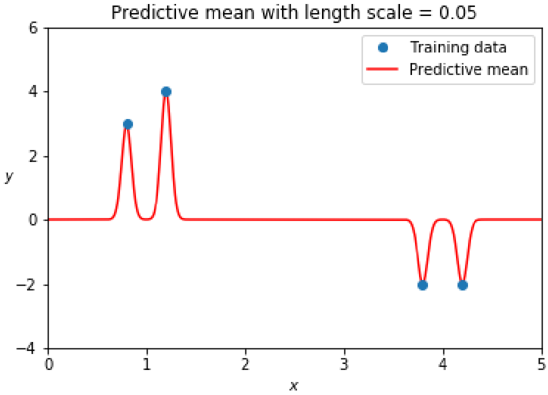
Um processo gaussiano (unidimensional) é uma família de variáveis aleatórias $\{f(x)\}x \in \chi$, tal que qualquer coleção finita em um subconjunto de $\{f(x)\}$ tem uma distribuição conjunta gaussiana.

$$f \sim GP(\mu, \Sigma) \tag{3}$$

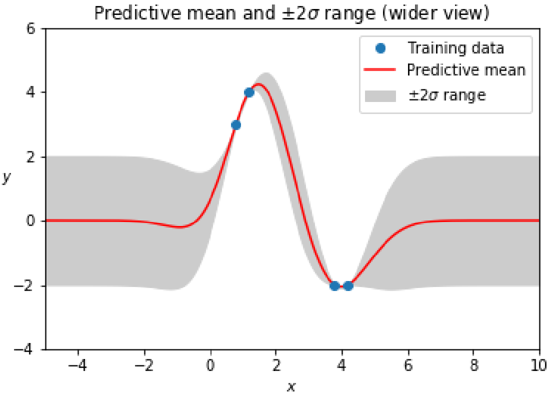
Um processo gaussiano e definido unicamente pela de média μ e pela covariância Σ

Função de Média

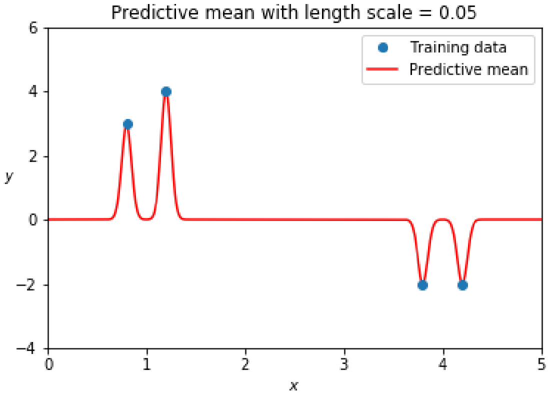
A função de média define a tendência global da funcao f^*, agindo - principalmente - onde não existem muitos pontos de input



Função de Média



A função de covariância calcula a similaridade entre dois pontos, definindo o comportamento do ajuste principalmente em pontos locais



Método dos processos gaussianos para regressão

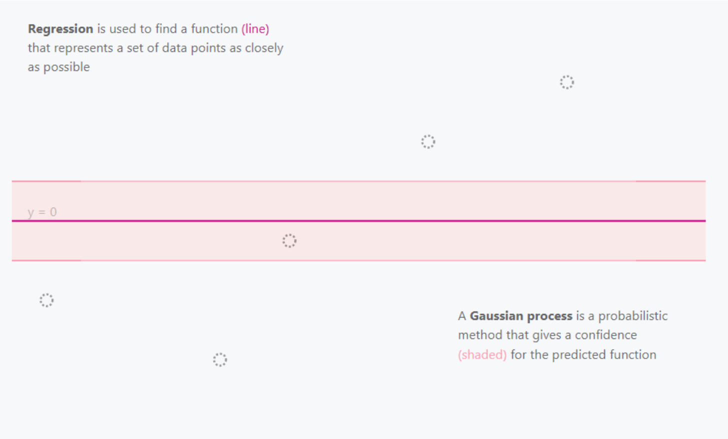


Figure: https://distill.pub/2019/visual-exploration-gaussian-processes/

Método dos processos gaussianos para regressão

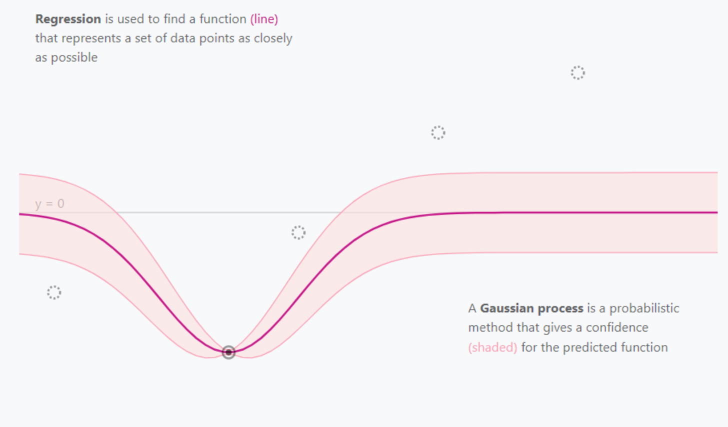


Figure: C. E. Rasmussen C. K. I. Williams, Gaussian Processes for Machine Learning

Regressão bayesiana

Método dos processos gaussianos para regressão

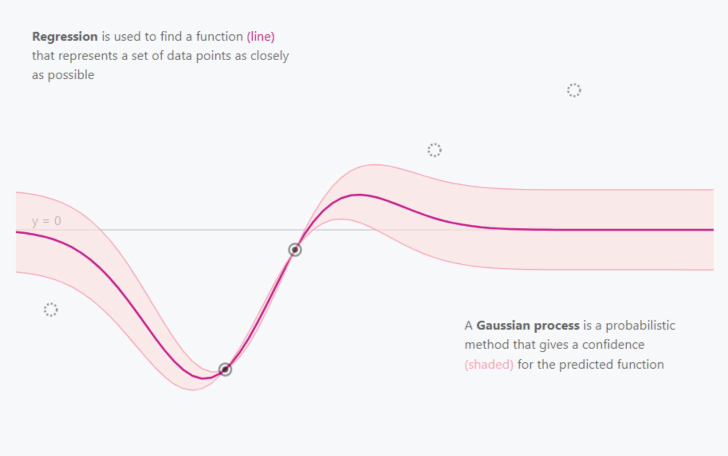


Figure: https://distill.pub/2019/visual-exploration-gaussian-processes/

Regressão bayesiana

Método dos processos gaussianos para regressão

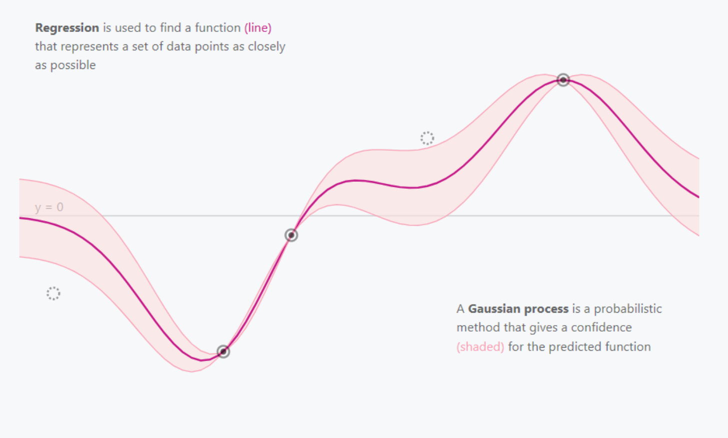


Figure: https://distill.pub/2019/visual-exploration-gaussian-processes/

Regressão bayesiana

Método dos processos gaussianos para regressão

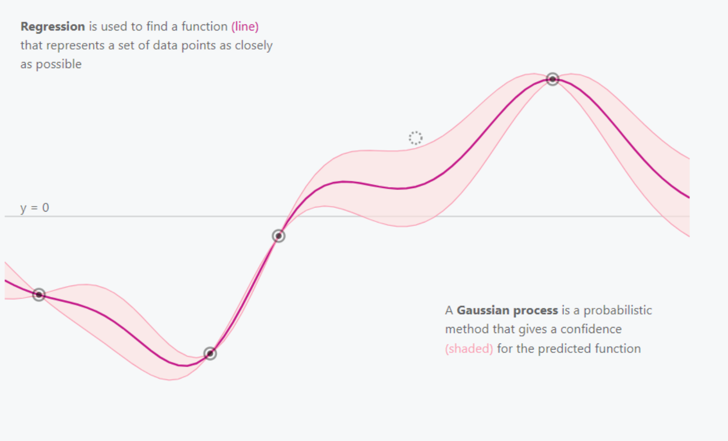


Figure: https://distill.pub/2019/visual-exploration-gaussian-processes/

Regressão bayesiana

Método dos processos gaussianos para regressão

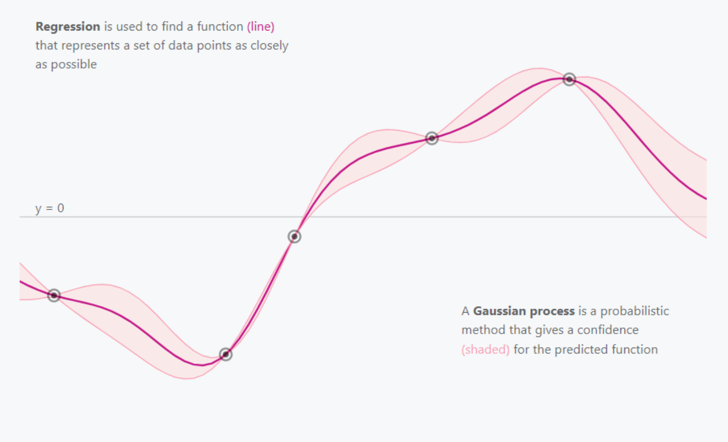


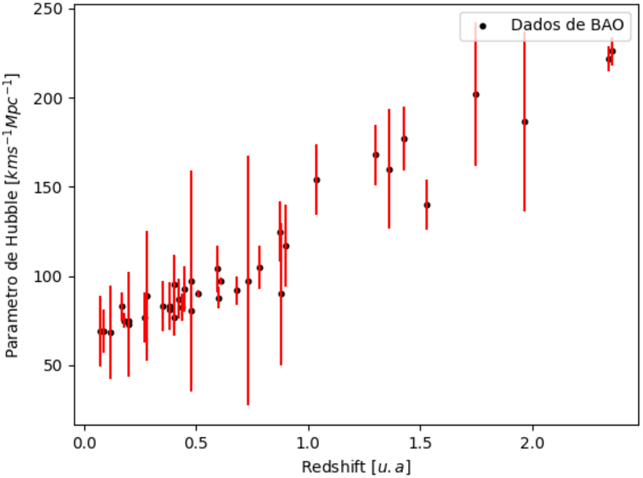
Figure: https://distill.pub/2019/visual-exploration-gaussian-processes/

Regressão bayesiana

Apesar da independência de um modelo de regressão, os processos gaussianos são muito dependentes de 3 fatores extremamente sensíveis

- Kernel
- Média
- Hiperparâmetros

Utilizarei os dados de parâmetro de Hubble



Regressão bayesiana

Para os dados de Parâmetro de Hubble, uilizarei duas médias:

- LCDM:

$$H(z) = H_0\sqrt{\Omega_{m0}(1+z)^3 + \Omega_{k0}(1+z)^2 + (1 - \Omega_{m0} - \Omega_{k0})}$$

- Zero:

$$H(z) = 0$$

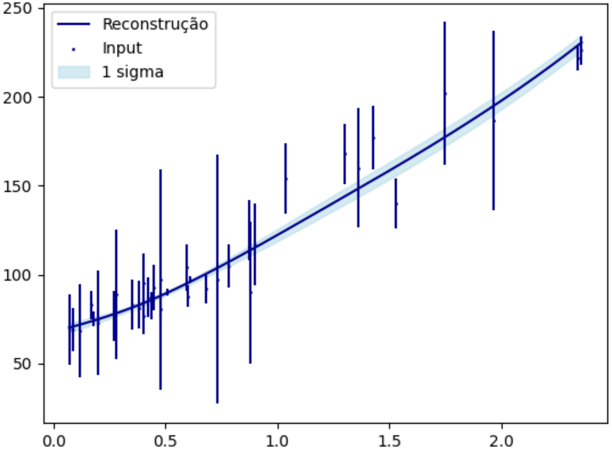
O Kernel Matern é definido por

$$\mathrm{Matern}_{\nu}(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}||x - x'||}{\rho} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}||x - x'||}{\rho} \right)$$

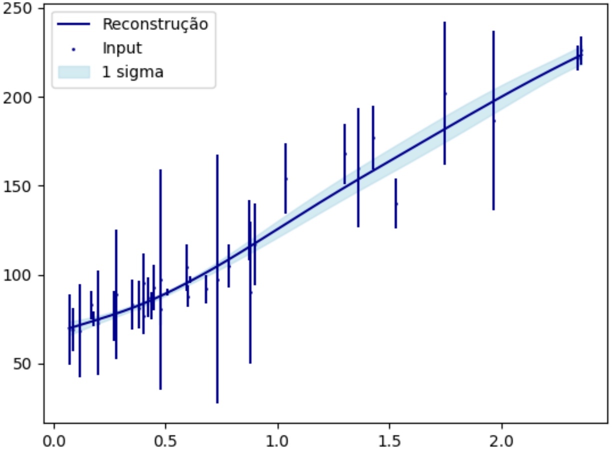
Onde

- ν é um parâmetro semi-inteiro
- Γ é a função gamma
- K_{ν} é a função de Bessel modificada de ordem ν

Reconstrução para o Kernel Matern 3/2, com média LCDM

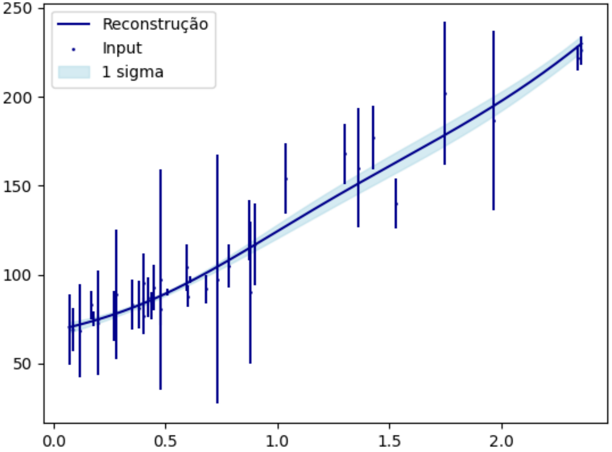


Reconstrução para o Kernel Matern 3/2, com média Zero

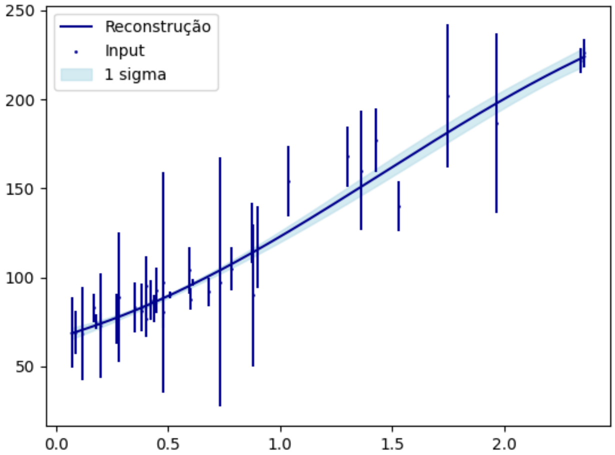


Regressão bayesiana

Reconstrução para o Kernel Matern 9/2, com média LCDM



Reconstrução para o Kernel Matern 9/2, com média Zero



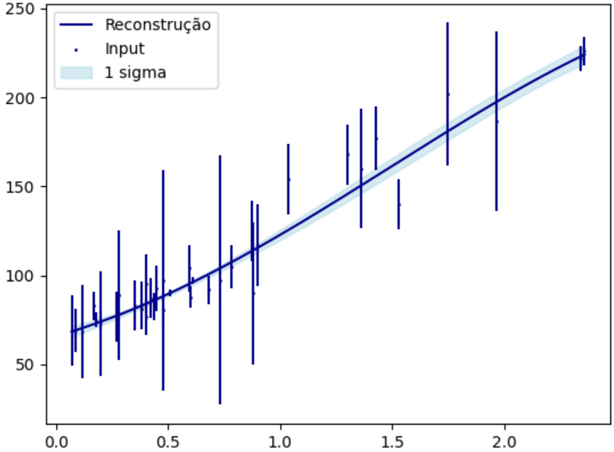
O Kernel Squared Exponential

$$SE(x, x') = \sigma^2 \exp\left(-\frac{||x - x'||^2}{2\ell^2}\right)$$

Onde

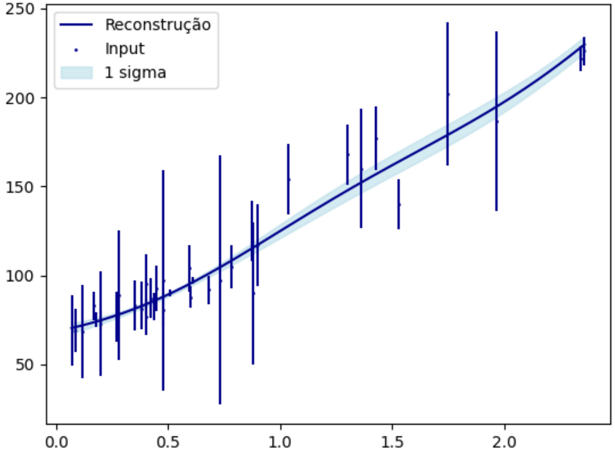
- σ é o parâmetro de variância
- ℓ é o parâmetro de escala

Reconstrução para o Squared Exponential, com média Zero

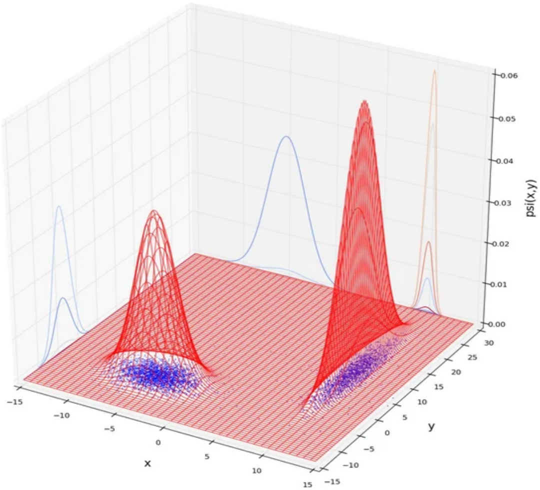


Regressão bayesiana

Reconstrução para o Squared Exponential, com média LCDM



Inferencia Estatística



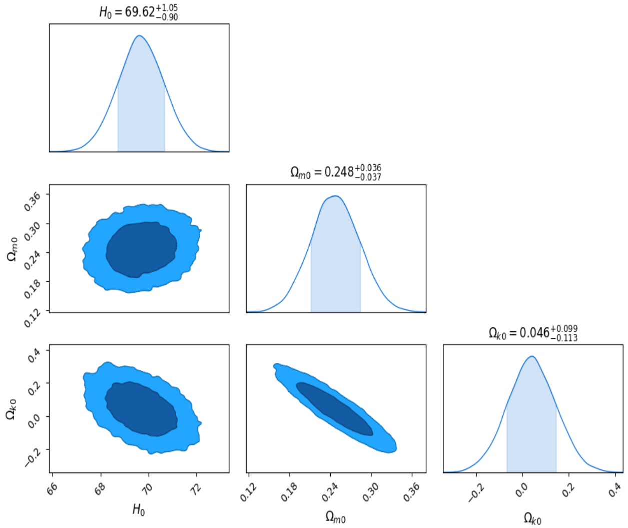


Figure: Resultado para os dados observacionais

Inferencia Estatística

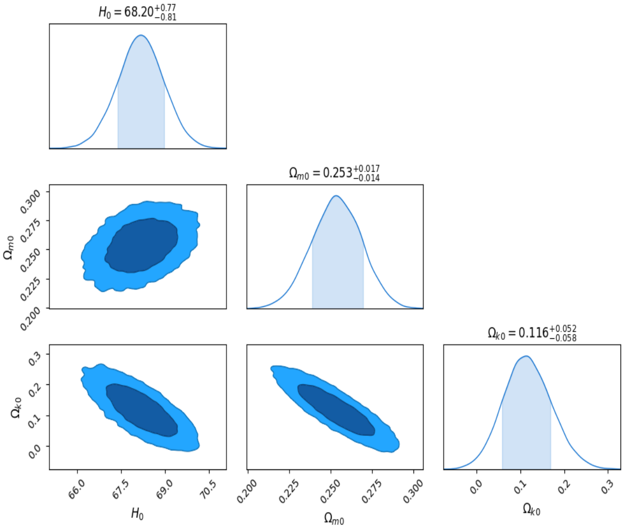


Figure: Resultado para o Kernel de Cauchy, com média LCDM

Inferencia Estatística

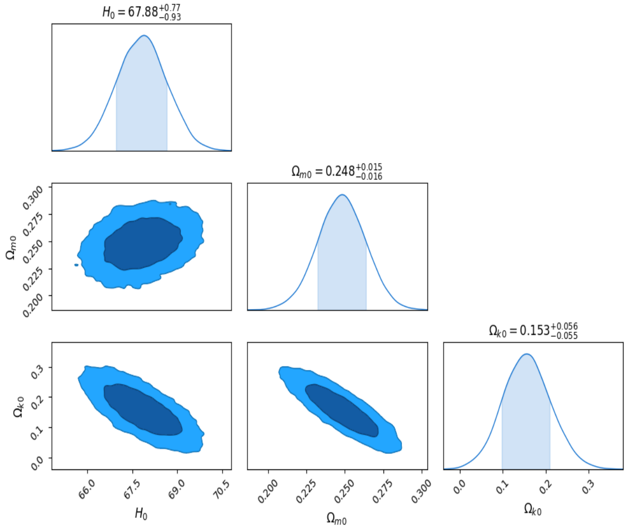


Figure: Resultado para o Kernel de Cauchy, com média Zero

Inferencia Estatística

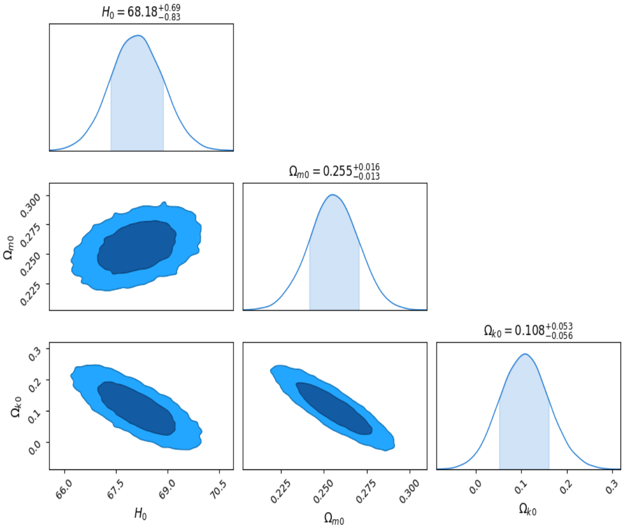


Figure: Resultado para o Kernel de Matern 1.5, com média LCDM

Inferencia Estatística

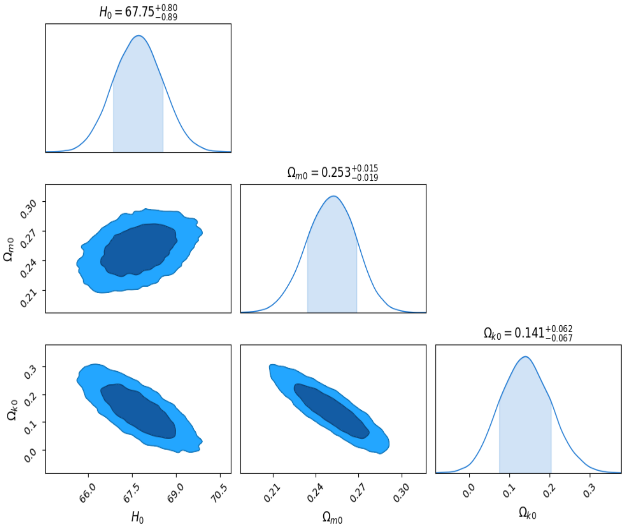


Figure: Resultado para o Kernel de Matern 1.5, com média Zero

Bibliografia I

[1] Carl. Rasmussen and Christopher. Williams Gaussian Processes for Machine Learning