Bayesian statistical inference

Regression

Associated notebooks:

06-Bayesian_inference_MCMC/Bayes_basics_short.ipynb

06-Bayesian_inference_MCMC/Bayes_simple_modeling.ipynb

Bayesian problem setting

P(science)

Science:

- Mass of a planet
- Rotation P of asteroid
- Super massive BH mass
- ...

Bayesian problem setting

P(science | data)

data:

- Observations
- Results of a simulation
- …

Bayesian problem setting

P(science | data, background info)

background information:

≡ What you know before getting any data

- Physical range (e.g. $M > 0$)
- Previous measurement
- …

Bayesian problem setting

P(science, nuisance parameters | data, background info)

Nuisance parameters:

≡ parameters you are not interested in

- Secular motion of a star during a transit
- Dust extinction in SN distance measurement
- …

Bayesian problem setting

P(science, nuisance parameters | data, background info)

$$\boxed{\mathbf{P}(\ \boldsymbol{\theta}_S,\ \boldsymbol{\theta}_N \mid D, I\ )}$$

≡ Posterior probability

Bayes theorem

$$\boxed{P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{P(D)}}$$

Bayes theorem

Likelihood

$$P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})}{P(D)}$$

Bayes theorem

Prior

$$P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta}) \boxed{P(\boldsymbol{\theta})}}{P(D)}$$

Bayes theorem

$$P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})}{P(D)}$$

$$P(D) = \int P(D \mid \boldsymbol{\theta})\, P(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$$

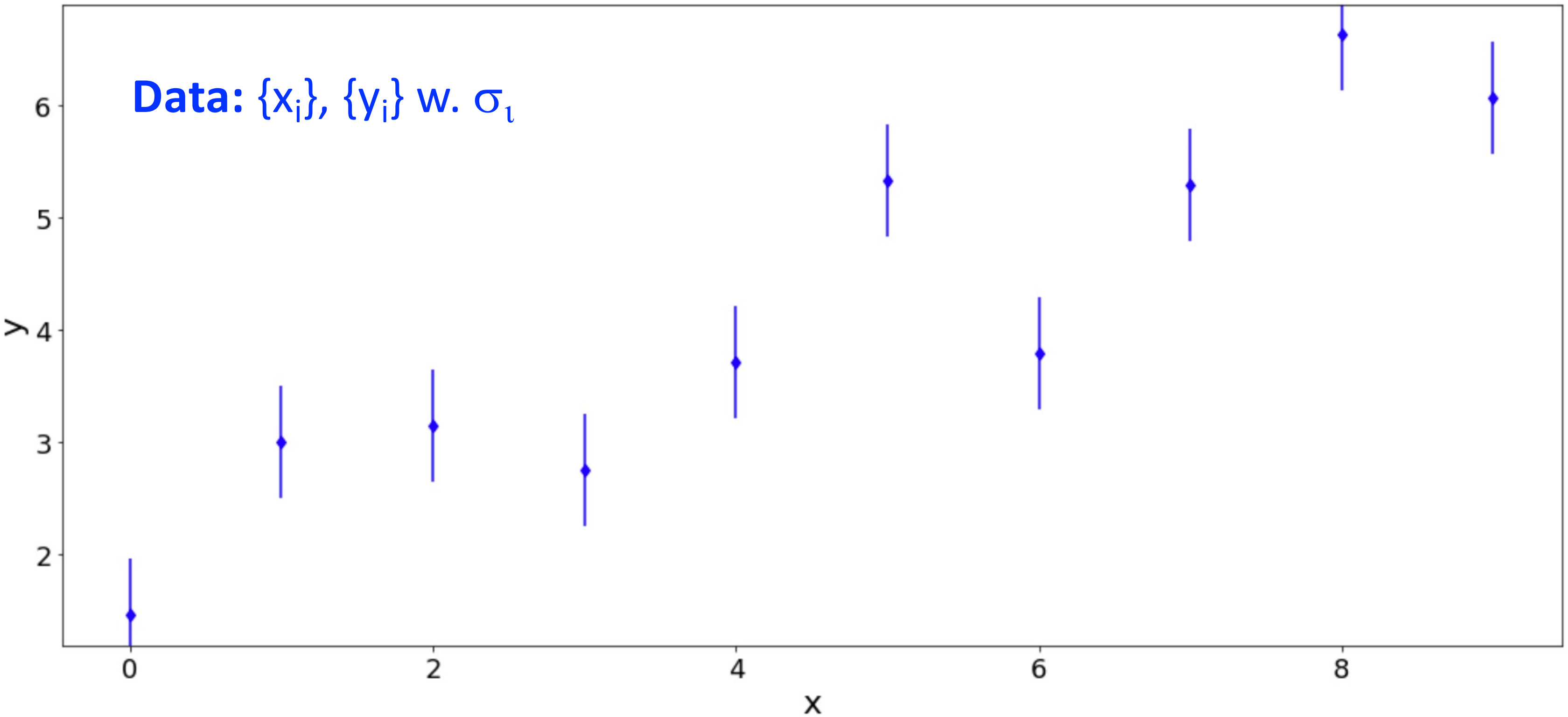
Evidence
/
Fully Marginalized likelihood

Bayes theorem

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Regression in the Bayesian framework

Let's assume that: $y_i \sim N(y_M(x_i;\boldsymbol{\theta}), \sigma)$



“Bayesian” Regression

$$P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{P(D)}$$

1. Model choice: $M(\boldsymbol{\theta}) : y_M(x) = \theta_0 + \theta_1 \, x$

"Bayesian" Regression

$$P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})}{P(D)}$$

1. Model choice: $M(\boldsymbol{\theta}) : y_M(x) = \theta_0 + \theta_1 \, x$

2. Likelihood: $\ln(P(D \mid \boldsymbol{\theta})) = -\frac{1}{2} \sum_{i=1}^{N} \left(\ln(2\pi\sigma_i^2) + \frac{(y_i - (\theta_0 + \theta_1 \, x_i))^2}{\sigma_i^2} \right)$

“Bayesian” Regression

$$P(\boldsymbol{\theta} \mid D, I) = \frac{P(D \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})}{P(D)}$$

1. Model choice: $M(\boldsymbol{\theta}) : y_M(x) = \theta_0 + \theta_1\, x$

2. Likelihood: $\ln(P(D \mid \boldsymbol{\theta})) = -\frac{1}{2} \sum_{i=1}^{N} \left(\ln(2\pi\sigma_i^2) + \frac{(y_i - (\theta_0 + \theta_1\, x_i))^2}{\sigma_i^2} \right)$

3. Prior:
   - ~~Conjugate (allows one to get analytic form of P(q | D))~~
   - Empirical: based on previous measurement
   - Flat: constant between 2 bounds (but can be informative)
   - Non / less informative

Go to Sect. IV.2 of the Notebook

“Bayesian” vs “Frequentist” regression

Frequentist:	***Bayes:***
Optimization with some merit function	*Sampling* of the likelihood
Search for *best* (fit) model *parameters*	PDF on parameters
“*Ignore*” the priors	*Accounts* explicitly for the priors