

# 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

$$\mathbf{P}(\text{science} \mid \text{data})$$

data:

- Observations
- Results of a simulation
- ...

# Bayesian problem setting

$$\mathbf{P}(\text{science} \mid \text{data}, \text{background info})$$

background information:

≡ What you know before getting any data

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

# Bayesian problem setting

$\mathbf{P}(\text{science, nuisance parameters} \mid \text{data, background info})$

Nuisance parameters:

$\equiv$  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)

$$\mathbf{P}( \theta_S, \theta_N | D, I )$$


≡ Posterior probability

# Bayes theorem

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

# Bayes theorem

Likelihood



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



# Bayes theorem

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

Prior



# Bayes theorem

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

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

Evidence

/

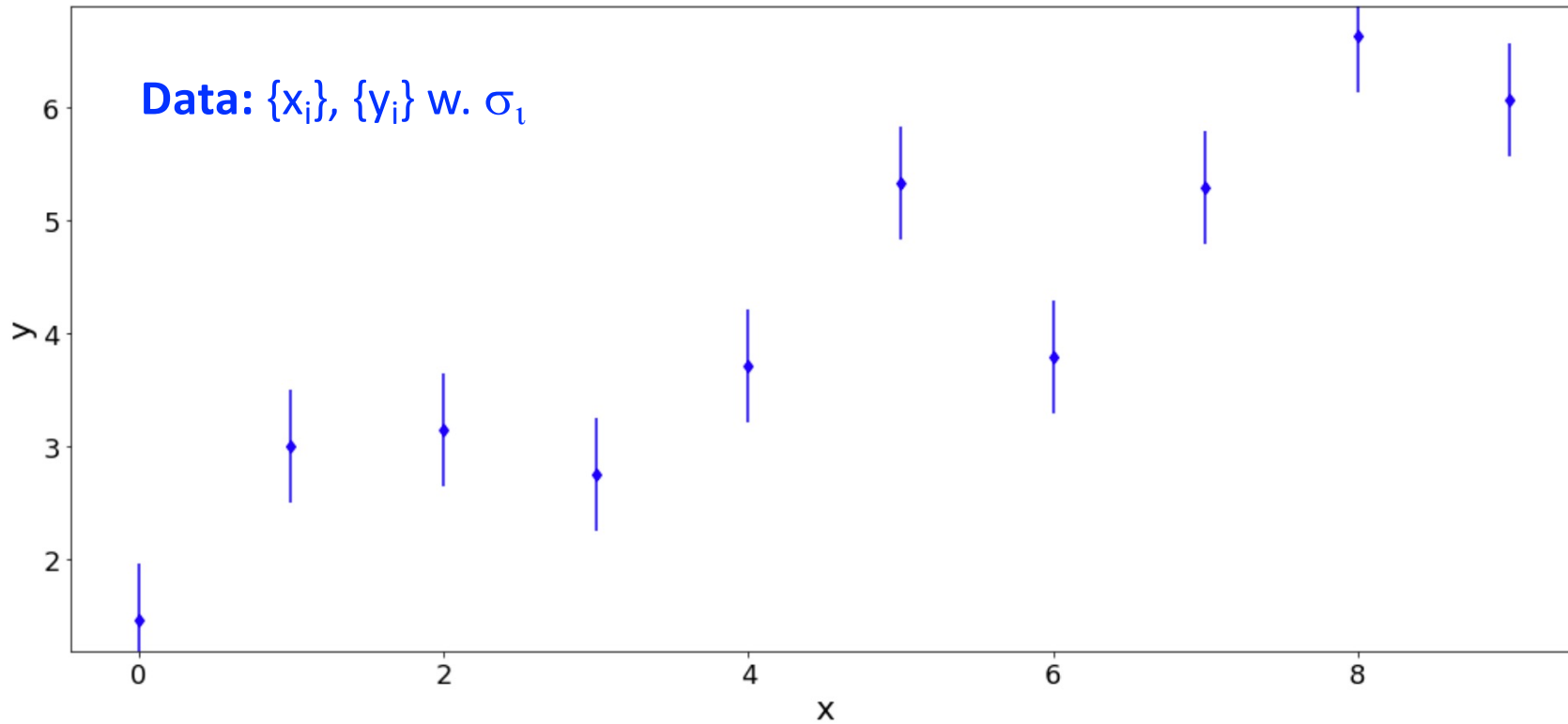
Fully Marginalized likelihood

# 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:  $M(\boldsymbol{\theta}) : y_M(x) = \theta_0 + \theta_1 x$

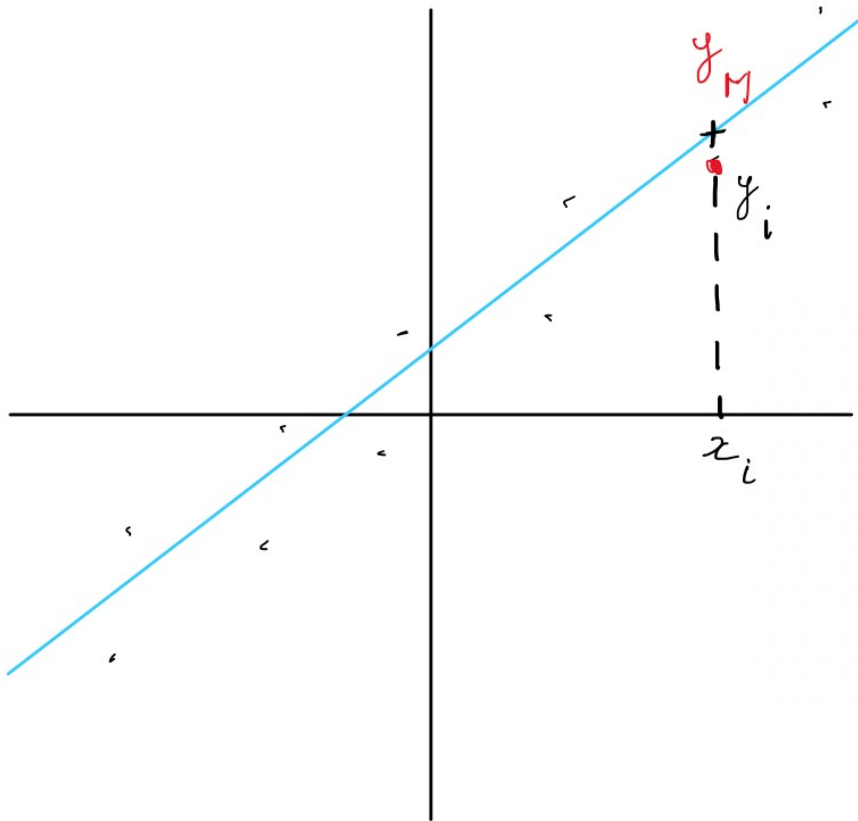
# “Bayesian” Regression

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

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

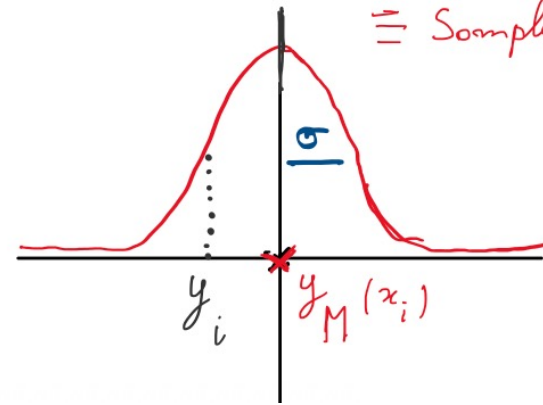
2. Likelihood: 
$$\ln(P(D \mid \boldsymbol{\theta})) = -\frac{1}{2} \sum_{i=1}^N \left( \ln(2\pi\sigma_i^2) + \frac{(y_i - (\theta_0 + \theta_1 x_i))^2}{\sigma_i^2} \right)$$

# “Bayesian” Regression



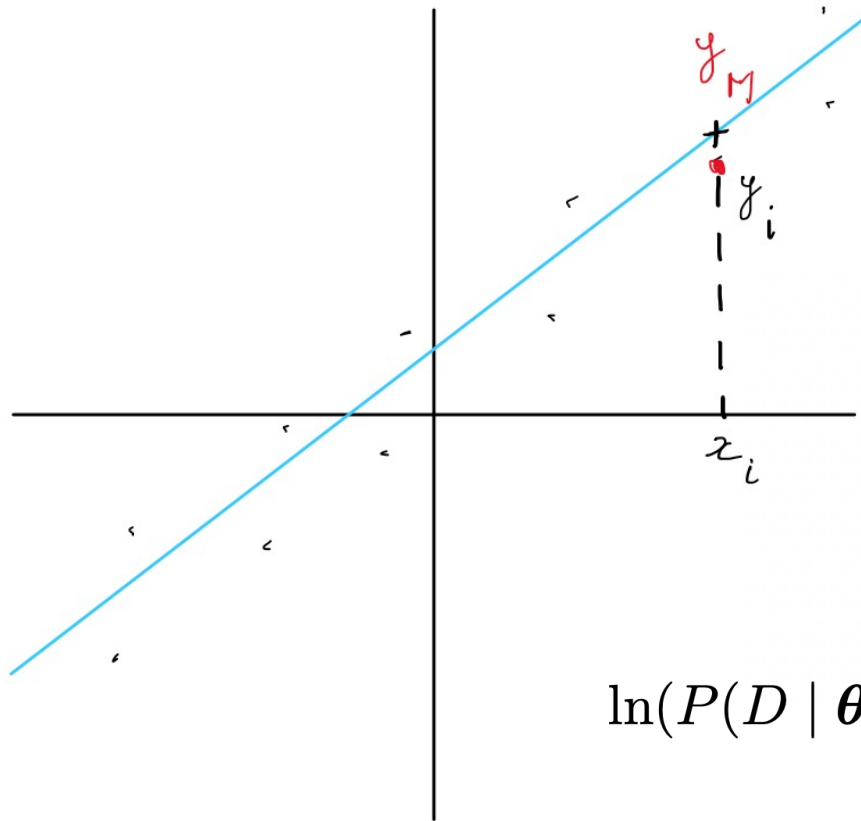
$$\mathcal{N}(y_M(x_i, \theta), \sigma)$$

$\equiv$  Sample distrib.



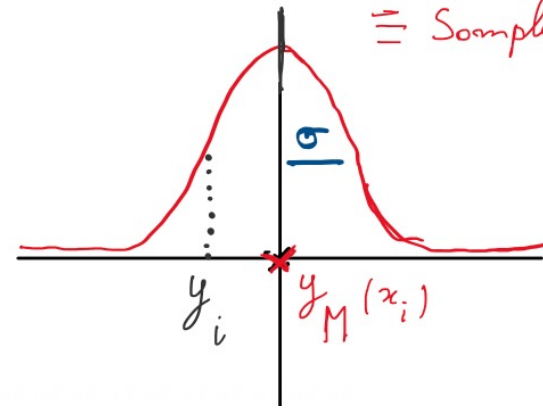
$$p(y_i \mid \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -0.5 \left( \frac{y_i - y_M(x_i; \theta)}{\sigma} \right)^2 \right]$$

# “Bayesian” Regression



$$\mathcal{N}(y_M(x_i, \theta), \sigma)$$

$\equiv$  Sample distrib.

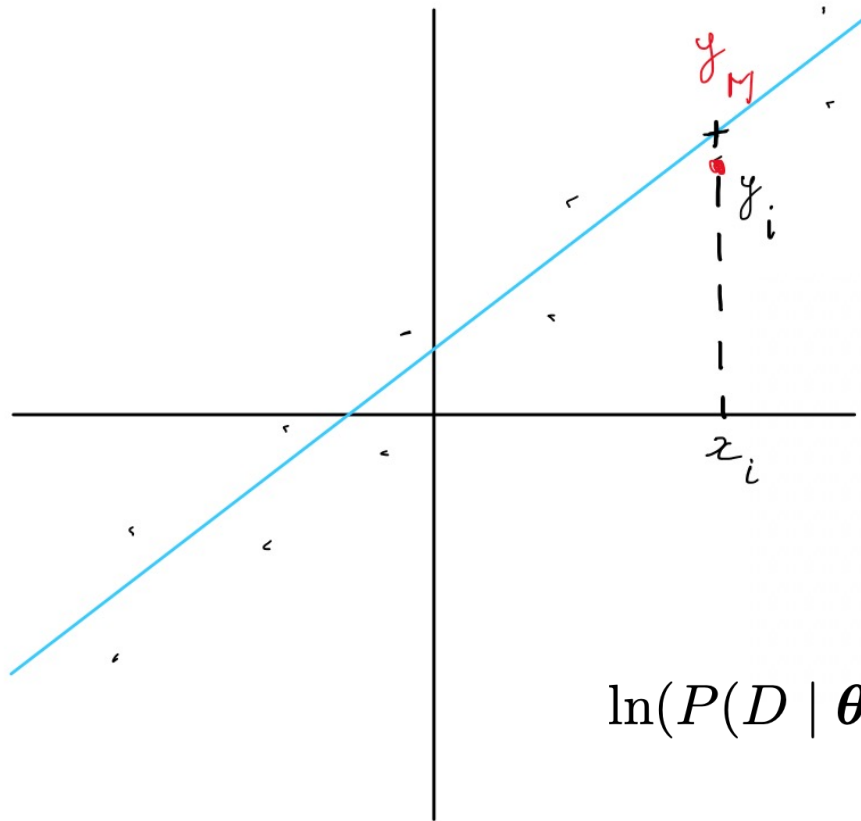


$$p(y_i \mid \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -0.5 \left( \frac{y_i - y_M(x_i; \theta)}{\sigma} \right)^2 \right]$$

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

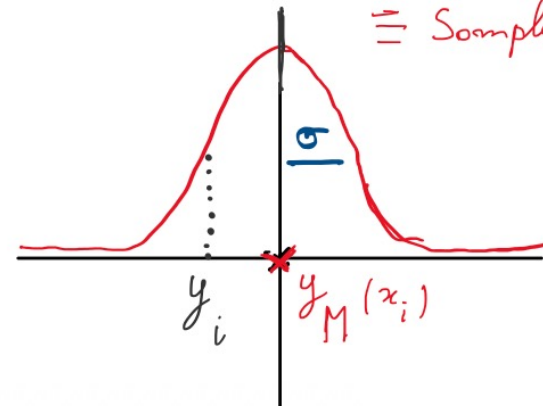


# “Bayesian” Regression



$$\mathcal{N}(y_M(x_i, \theta), \sigma)$$

$\equiv$  Sample distrib.



$$p(y_i \mid \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -0.5 \left( \frac{y_i - y_M(x_i; \theta)}{\sigma} \right)^2 \right]$$

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

Heteroscedastic errors

$\sigma_i^2$

# “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  $P(\mathbf{q} \mid D)$ )~~
- 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

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