Practical and efficient Bayesian model fitting with Variational Bayesian Monte Carlo (PyVBMC)

Luigi Acerbi

Department of Computer Science
University of Helsinki
Finnish Center for Artificial Intelligence FCAI

UNIVERSITY OF HELSINKI

FCAI Finnish Center for Artificial Intelligence

Data Science in Action @ UniPd
17 Dec 2022

What this is all about

By the end of this lecture/tutorial, we will:

Perform Bayesian inference on a real dataset and model from neuroscience

- Recap the basics of statistical modelling
- Define the psychometric model used in cognitive & neuroscience
- Explain the Bayesian approach to model fitting
- Briefly introduce variational inference algorithms
- Set up and run PyVBMC on a real dataset

1. A recap of statistical modelling
   - Of models and likelihoods
   - The psychometric function

2. Bayesian model fitting
   - Refresher of Bayesian inference
   - Bayesian inference for model fitting

3. Computing the posterior distribution
   - Setting things up
   - Inference algorithms
   - Making use of a Bayesian posterior

4. Hands-on tutorial

What is a model?

The best material model of a cat is another, or preferably the same, cat.

Wiener, *Philosophy of Science* (1945) (with Rosenblueth)

What is a mathematical model?

- Quantitative stand-in for a theory

What is a mathematical model?

- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\left(\text{data}|\boldsymbol{\theta}\right)$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector

What is a mathematical model?

- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\,(\text{data}|\boldsymbol{\theta})$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector

- **Why?**

What is a mathematical model?

- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\left(\mathrm{data}|\boldsymbol{\theta}\right)$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector

- **Why?** Description, prediction, and explanation

What is a mathematical model?

- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\,(\text{data}|\boldsymbol{\theta})$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector

- **Why?** Description, prediction, and explanation
- Defining $p(\text{data}|\boldsymbol{\theta})$ is the core of model building

What is a mathematical model?

- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\,(\text{data}|\boldsymbol{\theta})$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector

- **Why?** Description, prediction, and explanation
- Defining $p(\text{data}|\boldsymbol{\theta})$ is the core of model building
  - Wait, what?

What is a mathematical model?

- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\,(\text{data}|\boldsymbol{\theta})$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector
- **Why?** Description, prediction, and explanation
- Defining $p(\text{data}|\boldsymbol{\theta})$ is the core of model building
  - Wait, what?
- **How?** Think about the data generation process!

What is a mathematical model?

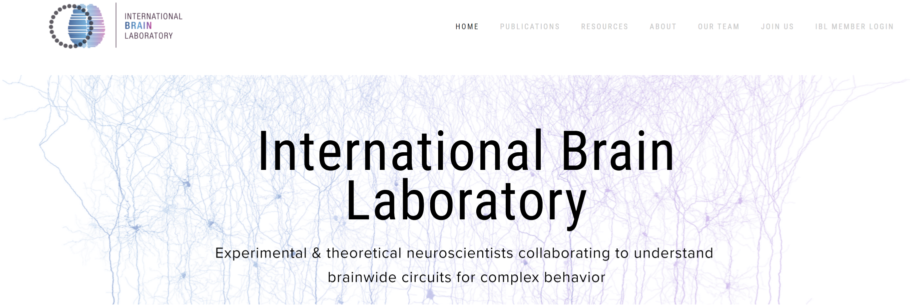
- Quantitative stand-in for a theory
- A *family of probability distributions* over possible datasets:

$$p\,(\text{data}|\boldsymbol{\theta})$$

  - data is a dataset with n data points (e.g., trials)
  - $\boldsymbol{\theta}$ is a parameter vector
- **Why?** Description, prediction, and explanation
- Defining $p(\text{data}|\boldsymbol{\theta})$ is the core of model building
  - Wait, what?
- **How?** Think about the data generation process!

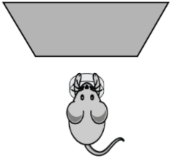
We need some data

Data from International Brain Laboratory (IBL)



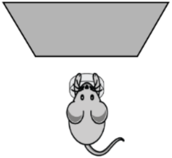
`https://www.internationalbrainlab.com`

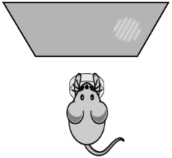
IBL Task



(IBL et al., *eLife*, 2021)

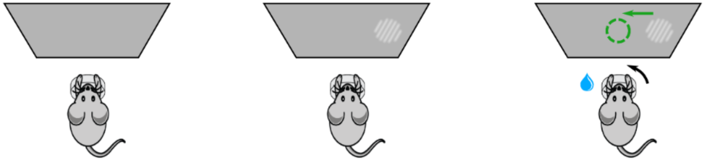
IBL Task





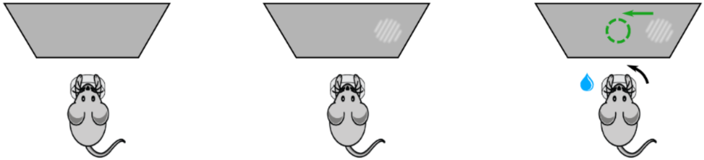
(IBL et al., *eLife*, 2021)

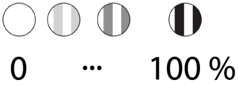
IBL Task



(IBL et al., *eLife*, 2021)

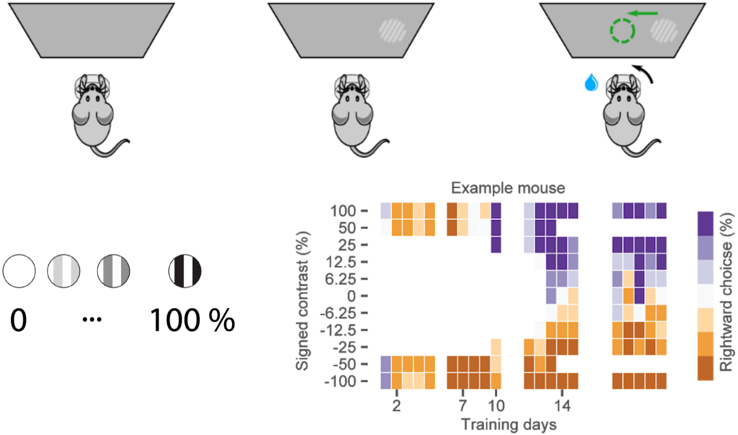
IBL Task





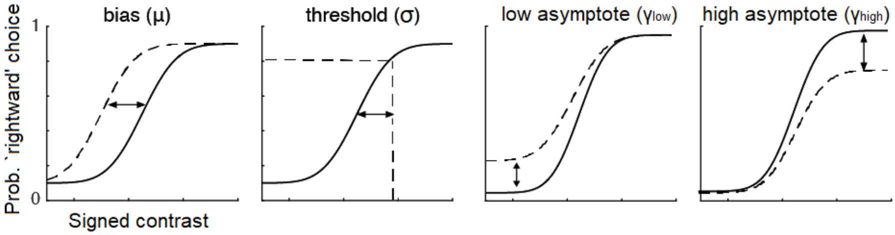
(IBL et al., *eLife*, 2021)

IBL Task



(IBL et al., *eLife*, 2021)

The psychometric function



- Data: (signed contrast, choice) for each trial
- Parameters $\boldsymbol{\theta}$: $(\mu,\ \sigma,\ \gamma_{\text{low}},\ \gamma_{\text{high}})$

$$p(\text{rightward choice}|s, \boldsymbol{\theta}) = \gamma_{\text{low}} + (1 - \gamma_{\text{low}} - \gamma_{\text{high}}) \cdot F(s; \mu, \sigma)$$

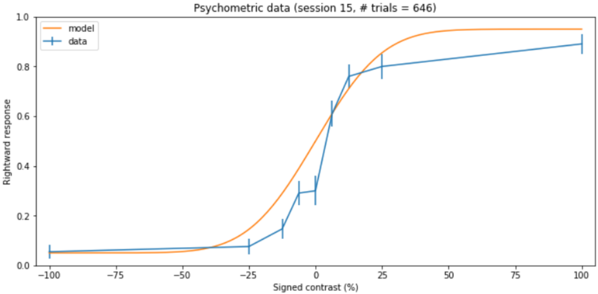
The psychometric function (alt version)

- Default decision process $F(s; \mu, \sigma)$
- Lapses with probability $\lambda \in [0, 1]$ (*lapse rate*)
- If lapse, respond 'rightward' with probability $\gamma \in [0, 1]$ (*lapse bias*)
- Parameters $\boldsymbol{\theta}$: $(\mu,\ \sigma,\ \lambda,\ \gamma)$

$$p(\text{rightward choice}|s, \boldsymbol{\theta}) = \lambda\gamma + (1 - \lambda) \cdot F(s; \mu, \sigma)$$

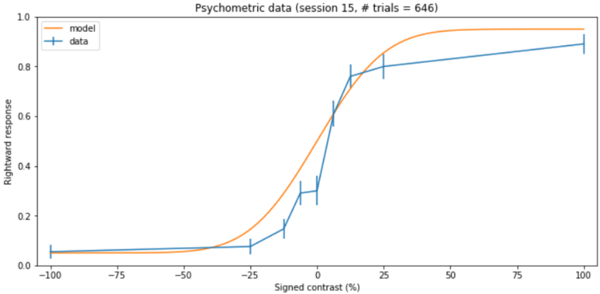
Metric for model fitting

We need a quantity to measure *goodness of fit*



Metric for model fitting

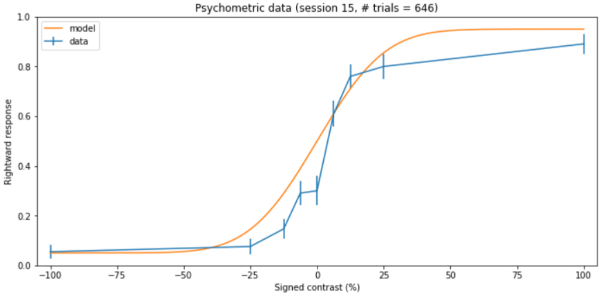
We need a quantity to measure *goodness of fit*



- Mean squared error?

Metric for model fitting

We need a quantity to measure *goodness of fit*



- ~~Mean squared error?~~
- The likelihood $p(\text{data}|\boldsymbol{\theta}) \equiv L(\boldsymbol{\theta}; \text{data})$

The (log) likelihood

- For numerical reasons we work with $\log p(\text{data}|\boldsymbol{\theta}) \equiv LL(\boldsymbol{\theta}; \text{data})$

The (log) likelihood

- For numerical reasons we work with $\log p(\text{data}|\boldsymbol{\theta}) \equiv LL(\boldsymbol{\theta}; \text{data})$
- Simplest case (conditionally independent trials):

$$\begin{aligned}\log p\left(\text{data}|\boldsymbol{\theta}\right) &= \log \prod_{i=1}^{n} p_i \left(\boldsymbol{y}^{(i)}|\boldsymbol{s}^{(i)}, \boldsymbol{\theta}\right) \\ &= \sum_{i=1}^{n} \log p_i \left(\boldsymbol{y}^{(i)}|\boldsymbol{s}^{(i)}, \boldsymbol{\theta}\right)\end{aligned}$$

The (log) likelihood

- For numerical reasons we work with $\log p(\text{data}|\boldsymbol{\theta}) \equiv LL(\boldsymbol{\theta}; \text{data})$
- Simplest case (conditionally independent trials):

$$\begin{aligned}\log p\left(\text{data}|\boldsymbol{\theta}\right) &= \log \prod_{i=1}^{n} p_i\left(\boldsymbol{y}^{(i)}|\boldsymbol{s}^{(i)}, \boldsymbol{\theta}\right) \\ &= \sum_{i=1}^{n} \log p_i\left(\boldsymbol{y}^{(i)}|\boldsymbol{s}^{(i)}, \boldsymbol{\theta}\right)\end{aligned}$$

- Model building: Write function with
  - Input: $\boldsymbol{\theta}$ and data
  - Output: $\log p(\text{data}|\boldsymbol{\theta})$

1. A recap of statistical modelling
   - Of models and likelihoods
   - The psychometric function

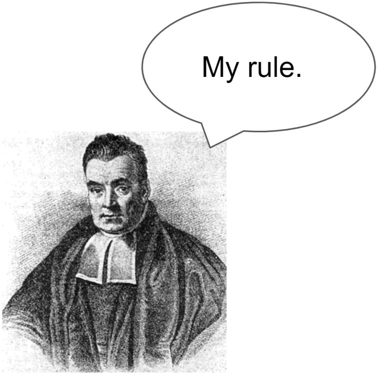
2. Bayesian model fitting
   - Refresher of Bayesian inference
   - Bayesian inference for model fitting

3. Computing the posterior distribution
   - Setting things up
   - Inference algorithms
   - Making use of a Bayesian posterior

4. Hands-on tutorial

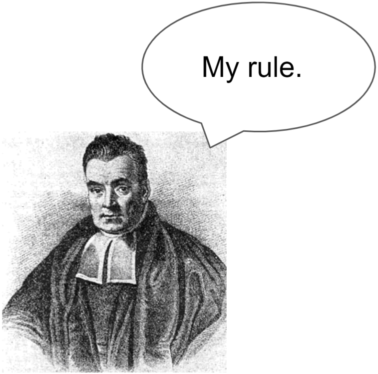
What is Bayesian inference?

What is Bayesian inference?



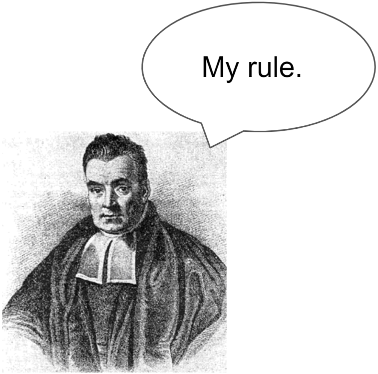
$$p(\boldsymbol{\theta}|\mathrm{data}) = \frac{p(\mathrm{data}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathrm{data})}$$

What is Bayesian inference?



$$\overbrace{p(\boldsymbol{\theta}|\text{data})}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\boldsymbol{\theta})}^{\text{likelihood}}\,\overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\text{data})}_{\text{evidence}}}$$

What is Bayesian inference?



$$\overbrace{p(\boldsymbol{\theta}|\text{data})}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\boldsymbol{\theta})}^{\text{likelihood}}\,\overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\text{data})}_{\text{evidence}}}$$

$$p(\text{data}) = \int p(\text{data}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

What's special in Bayesian inference for model fitting?

The output of Bayesian inference is a probability distribution (posterior) over model parameters:

$$p(\boldsymbol{\theta}|\text{data})$$

Instead, other methods (like maximum-likelihood estimation or loss minimization) only return a single best point estimate $\boldsymbol{\theta}_\star$.

What's special in Bayesian inference for model fitting?

The output of Bayesian inference is a probability distribution (posterior) over model parameters:

$$p(\boldsymbol{\theta}|\text{data})$$

Instead, other methods (like maximum-likelihood estimation or loss minimization) only return a single best point estimate $\boldsymbol{\theta}_\star$.

Questions:

1. How do we compute $p(\boldsymbol{\theta}|\text{data})$?
2. What do we do once we have $p(\boldsymbol{\theta}|\text{data})$?
3. Why should we bother?

What's special in Bayesian inference for model fitting?

The output of Bayesian inference is a probability distribution (posterior) over model parameters:

$$p(\boldsymbol{\theta}|\text{data})$$

Instead, other methods (like maximum-likelihood estimation or loss minimization) only return a single best point estimate $\boldsymbol{\theta}_\star$.

Questions:

1. How do we compute $p(\boldsymbol{\theta}|\text{data})$?
2. What do we do once we have $p(\boldsymbol{\theta}|\text{data})$?
3. Why should we bother?

Why Bayesian inference?

$$\overbrace{p(\boldsymbol{\theta}|\text{data})}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\boldsymbol{\theta})}^{\text{likelihood}} \overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\text{data})}_{\text{evidence}}} \qquad p(\text{data}) = \int p(\text{data}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Why Bayesian inference?

$$\overbrace{p(\boldsymbol{\theta}|\text{data})}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\boldsymbol{\theta})}^{\text{likelihood}}\,\overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\text{data})}_{\text{evidence}}} \qquad p(\text{data}) = \int p(\text{data}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

- Uncertainty quantification
- Optimal experiment design
- Robustness
- Interpretability

Why Bayesian inference?

$$\overbrace{p(\boldsymbol{\theta}|\text{data})}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\boldsymbol{\theta})}^{\text{likelihood}}\ \overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\text{data})}_{\text{evidence}}} \qquad p(\text{data}) = \int p(\text{data}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

- Uncertainty quantification
- Optimal experiment design
- Robustness
- Interpretability
- Hyperparameter tuning
- Model selection

Why Bayesian inference?

$$\overbrace{p(\boldsymbol{\theta}|\text{data})}^{\text{posterior}} = \frac{\overbrace{p(\text{data}|\boldsymbol{\theta})}^{\text{likelihood}}\ \overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\text{data})}_{\text{evidence}}} \qquad p(\text{data}) = \int p(\text{data}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

- Uncertainty quantification
- Optimal experiment design
- Robustness
- Interpretability
- Hyperparameter tuning
- Model selection
- Better predictions

1 A recap of statistical modelling
- Of models and likelihoods
- The psychometric function

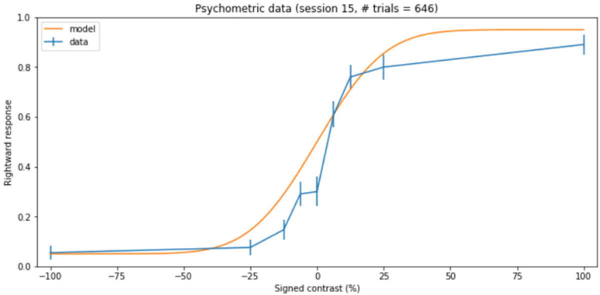
2 Bayesian model fitting
- Refresher of Bayesian inference
- Bayesian inference for model fitting

3 Computing the posterior distribution
- Setting things up
- Inference algorithms
- Making use of a Bayesian posterior

4 Hands-on tutorial

Data and model

- Data: IBL mouse behavioral data
- Model: psychometric function



Choose your prior

- In Bayesian inference you need a prior over parameters, $p(\boldsymbol{\theta})$

Choose your prior

- In Bayesian inference you need a prior over parameters, $p(\boldsymbol{\theta})$
- Common choice: independent priors $p(\boldsymbol{\theta}) = \prod_{d=1}^{D} p(\theta_d)$

Choose your prior

- In Bayesian inference you need a prior over parameters, $p(\boldsymbol{\theta})$
- Common choice: independent priors $p(\boldsymbol{\theta}) = \prod_{d=1}^{D} p(\theta_d)$
  - Choose the prior $p(\theta_d)$ for each parameter
  - Independent prior does not mean that the posterior is independent!

Choose your prior

- In Bayesian inference you need a prior over parameters, $p(\boldsymbol{\theta})$
- Common choice: independent priors $p(\boldsymbol{\theta}) = \prod_{d=1}^{D} p(\theta_d)$
  - Choose the prior $p(\theta_d)$ for each parameter
  - Independent prior does not mean that the posterior is independent!
- Remember that the prior is a probability distribution $\int p(\theta) d\theta = 1$

Choose your prior

- In Bayesian inference you need a prior over parameters, $p(\boldsymbol{\theta})$
- Common choice: independent priors $p(\boldsymbol{\theta}) = \prod_{d=1}^{D} p(\theta_d)$
  - Choose the prior $p(\theta_d)$ for each parameter
  - Independent prior does not mean that the posterior is independent!
- Remember that the prior is a probability distribution $\int p(\theta) d\theta = 1$
- Okay, but how do I pick a prior for each parameter?
  - Bounded parameter: Uniform, ...
  - Unbounded parameter: Gaussian, Student's t...
  - Would deserve a separate lecture

Inference algorithms

- A general-purpose inference algorithm
  - takes as input an inference problem (likelihood, prior,...)
  - returns an approximate posterior

Inference algorithms

- A general-purpose inference algorithm
  - takes as input an inference problem (likelihood, prior,...)
  - returns an approximate posterior
- Example families of algorithms
  1. Markov Chain Monte Carlo (MCMC)
  2. Variational inference
  3. Others

Variational inference

- Approximate $p(\boldsymbol{\theta}|\text{data})$ with $q_{\phi}(\boldsymbol{\theta})$

Variational inference

- Approximate $p(\boldsymbol{\theta}|\text{data})$ with $q_{\phi}(\boldsymbol{\theta})$
- Minimize Kullback-Leibler divergence between q and p

Variational inference

- Approximate $p(\boldsymbol{\theta}|\text{data})$ with $q_{\phi}(\boldsymbol{\theta})$
- Minimize Kullback-Leibler divergence between q and p

Outputs:

- An approximate posterior $q_{\phi}(\boldsymbol{\theta})$
- A lower bound to the log marginal likelihood, $\text{ELBO}(\phi)$

Variational inference

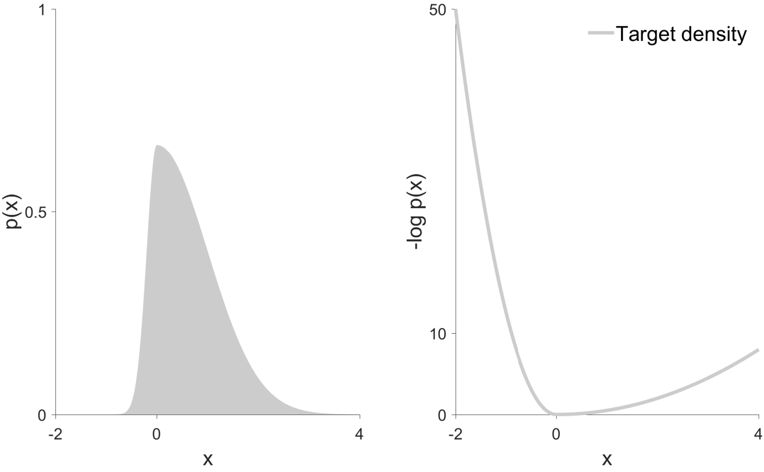
- Approximate $p(\boldsymbol{\theta}|\text{data})$ with $q_{\phi}(\boldsymbol{\theta})$
- Minimize Kullback-Leibler divergence between q and p

Outputs:

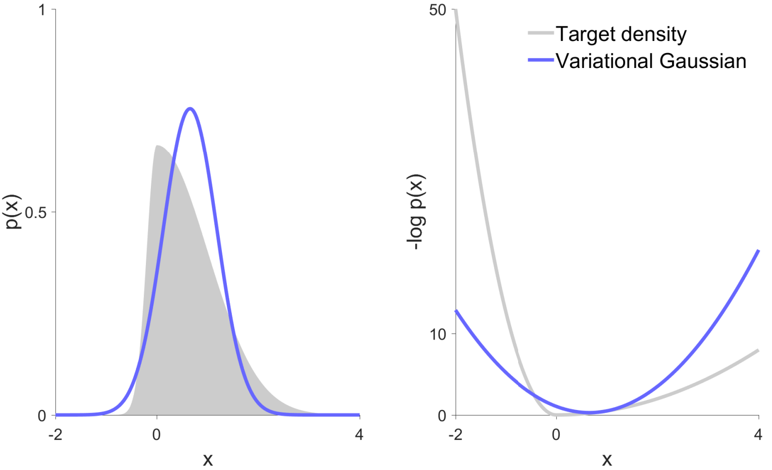
- An approximate posterior $q_{\phi}(\boldsymbol{\theta})$
- A lower bound to the log marginal likelihood, $\text{ELBO}(\phi)$

VI casts Bayesian inference into optimization + integration

Variational inference: example

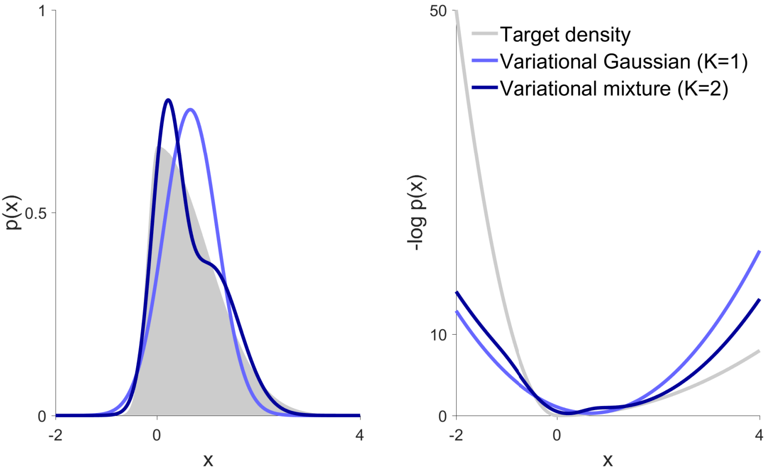


Variational inference: example



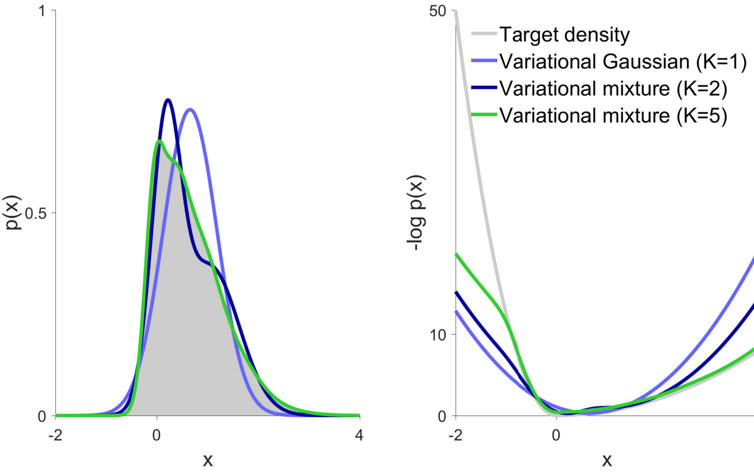
$$q_\phi(x) = \mathcal{N}\left(x, \mu, \sigma^2\right) \qquad \phi = \left(\mu, \sigma^2\right)$$

Variational inference: example



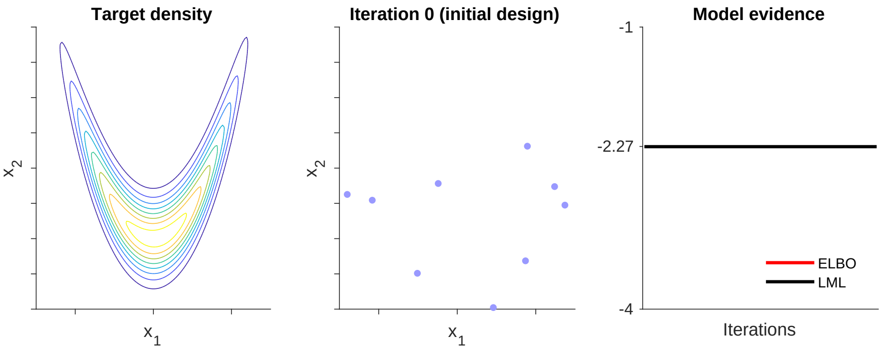
$$q_\phi(x) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x, \mu_k, \sigma_k^2\right) \qquad \phi = \left(w_k, \mu_k, \sigma_k^2\right)_{k=1}^{K}$$

Variational inference: example



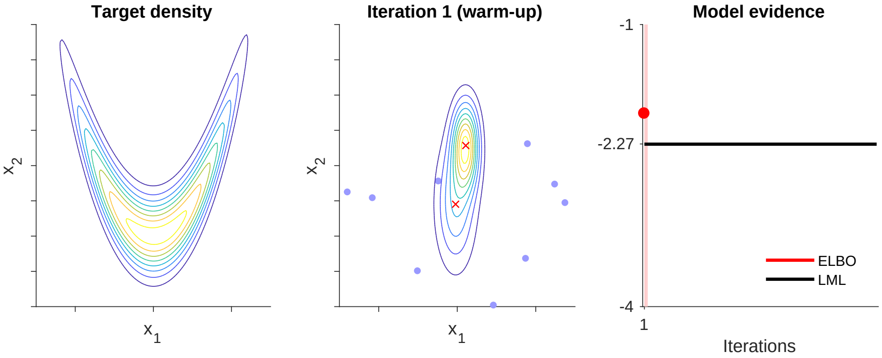
$$q_\phi(x) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x, \mu_k, \sigma_k^2\right) \qquad \phi = \left(w_k, \mu_k, \sigma_k^2\right)_{k=1}^{K}$$

Variational Bayesian Monte Carlo (VBMC)



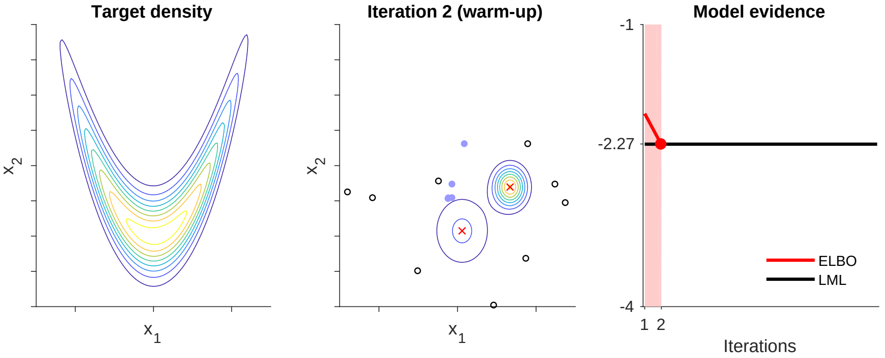
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



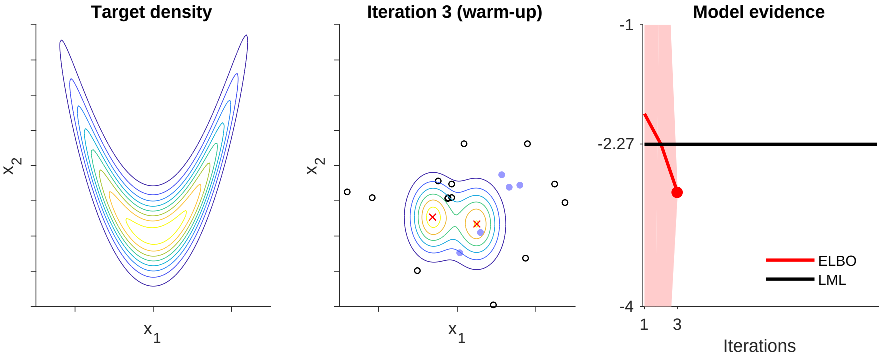
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



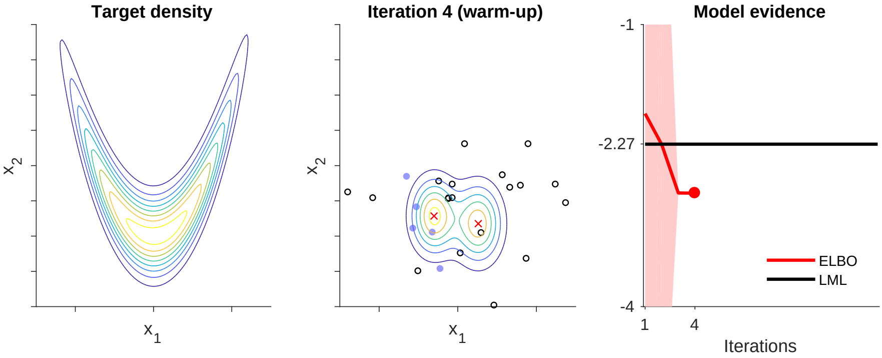
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



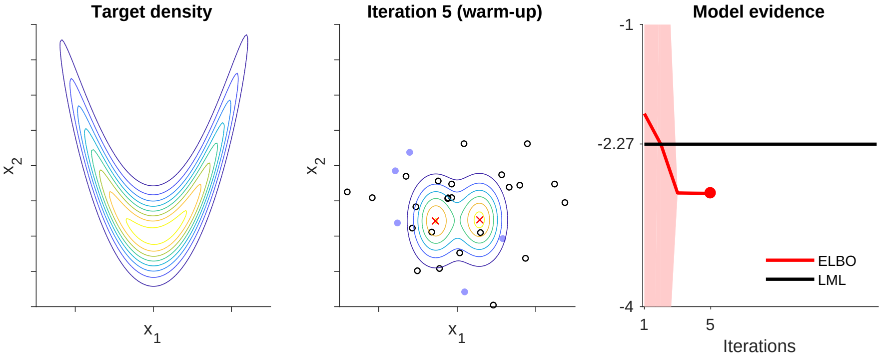
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



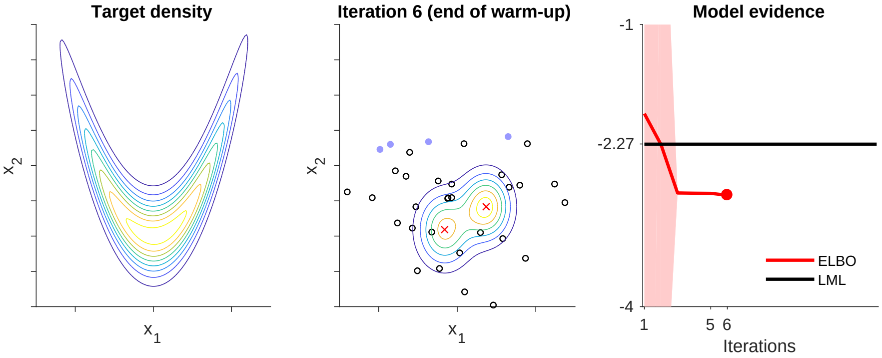
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



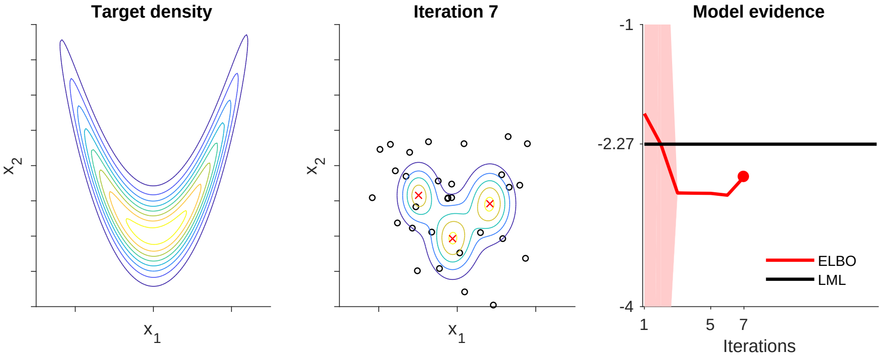
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



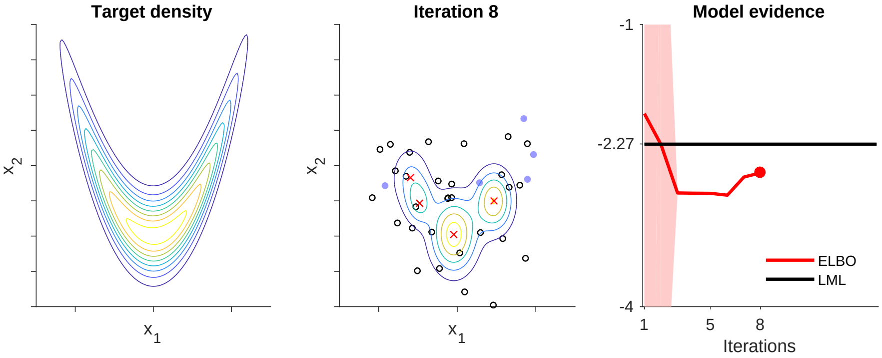
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



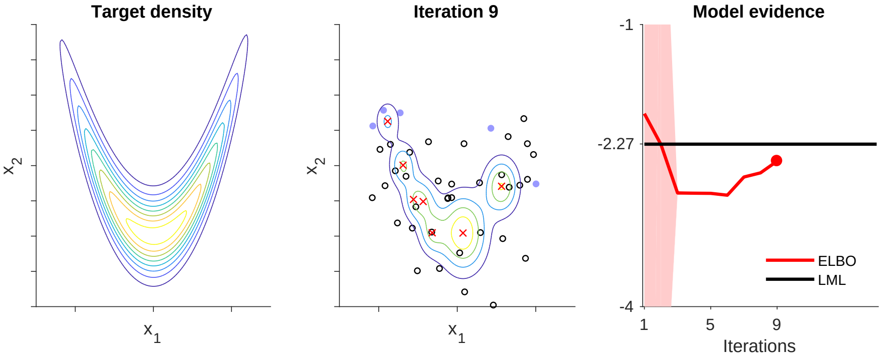
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



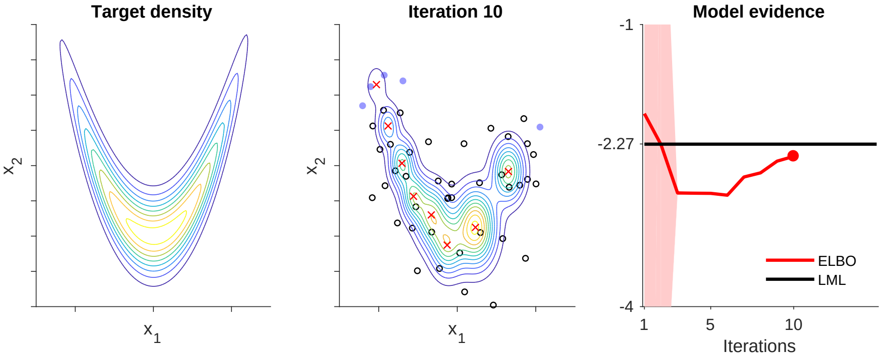
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



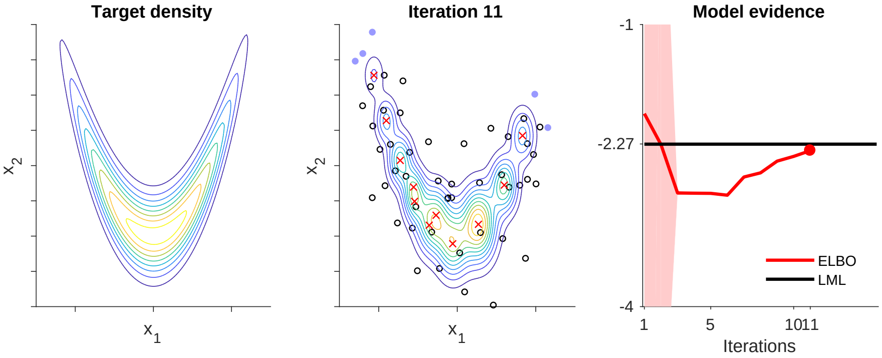
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



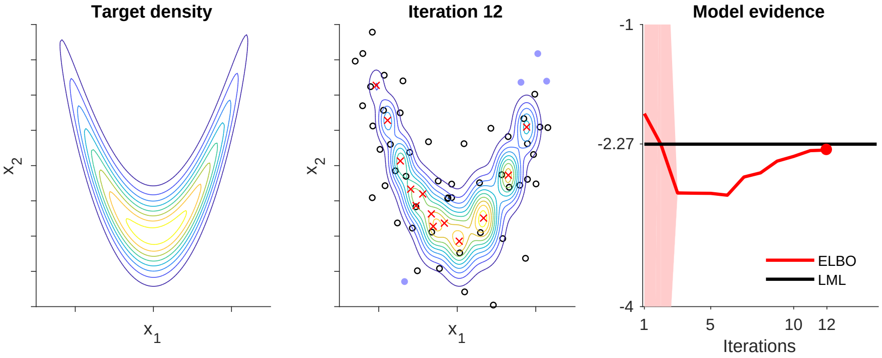
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



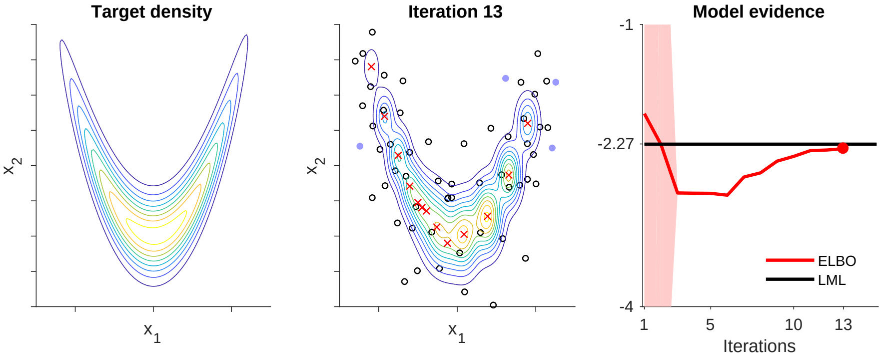
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



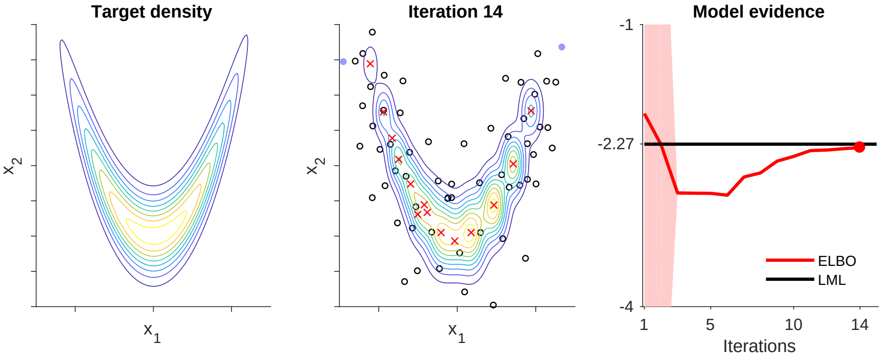
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



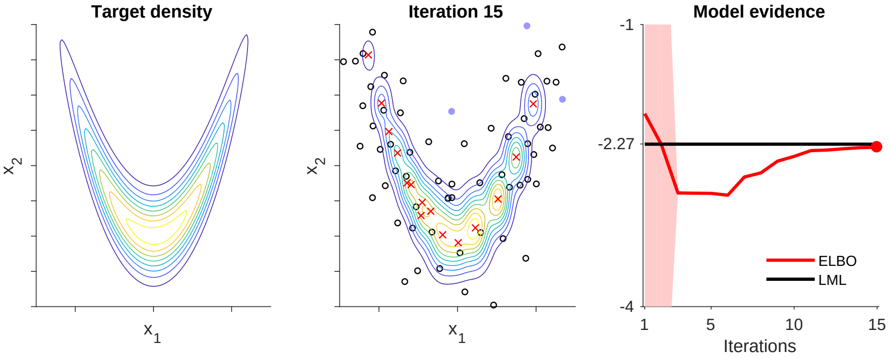
Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



Acerbi, *NeurIPS* (2018; 2020)

Variational Bayesian Monte Carlo (VBMC)



Acerbi, *NeurIPS* (2018; 2020)

OK suppose we have a posterior what now

OK suppose we have a posterior what now

- Visualize the posterior distribution
- Represent uncertainty (e.g., credible intervals)
- Make posterior predictions ("Bayesian fit") and compare to data

1. A recap of statistical modelling
   - Of models and likelihoods
   - The psychometric function
2. Bayesian model fitting
   - Refresher of Bayesian inference
   - Bayesian inference for model fitting
3. Computing the posterior distribution
   - Setting things up
   - Inference algorithms
   - Making use of a Bayesian posterior
4. Hands-on tutorial

It's Bayes time!

Let's set up and run PyVBMC

`github.com/acerbilab/pyvbmc`

What we learnt

By the end of this lecture/tutorial, we will:

Perform Bayesian inference on a real dataset and model from neuroscience

- Recap the basics of statistical modelling
- Define the psychometric model used in cognitive & neuroscience
- Explain the Bayesian approach to model fitting
- Briefly introduce variational inference algorithms
- Set up and run PyVBMC on a real dataset

This was a lot

This was a lot

You deserve a cat picture

This was a lot

You deserve a cat picture

- Bayesian model fitting could fill the entire day
- This tutorial is just the first steps on the Bayesian way

Final slide

Contacts:

- Email: `luigi.acerbi@helsinki.fi`
- Twitter: `@AcerbiLuigi`

Acknowledgments:

- The PyVBMC development team
- FCAI

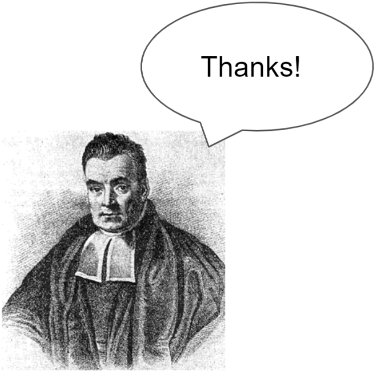
Code:

- PyVBMC: `github.com/acerbilab/pyvbmc`

Final slide

Contacts:

- Email: `luigi.acerbi@helsinki.fi`
- Twitter: `@AcerbiLuigi`



Acknowledgments:

- The PyVBMC development team
- FCAI

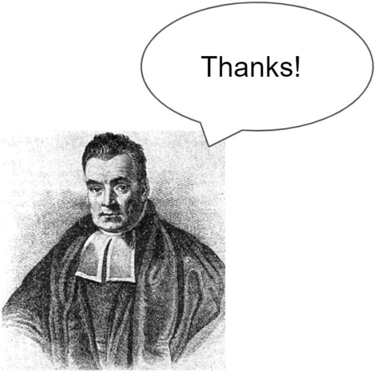
Code:

- PyVBMC: `github.com/acerbilab/pyvbmc`

Final slide

Contacts:

- Email: `luigi.acerbi@helsinki.fi`
- Twitter: `@AcerbiLuigi`



Acknowledgments:

- The PyVBMC development team
- FCAI

Code:

- PyVBMC: `github.com/acerbilab/pyvbmc`

Questions?