Fitting a model: VB & MCMC

Lionel Rigoux

Max Planck Institute for Metabolism Research

Translational Neuro-Circuitry Group

Max-Planck-Institut
für Stoffwechselforschung

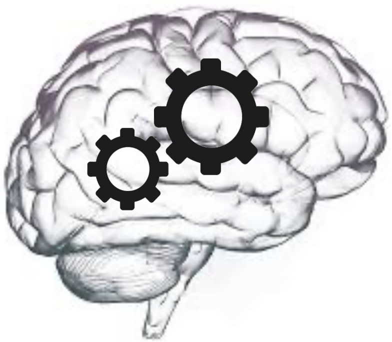
Overview

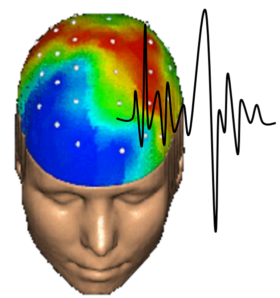
$\mathbf{p(y|\theta, m)}$

forward problem

generative model

$\mathbf{m, p(\theta)}$





observation

$\mathbf{y}$

inverse problem

$\mathbf{p(\theta|y, m)}$ $\mathbf{p(m|y)}$

Bayes rule

Joint distribution

$$p(y, \theta|m)$$

$$p(\theta|y, m) = \frac{\overbrace{p(y|\theta, m)p(\theta|m)}}{\underbrace{\int p(y|\theta, m)p(\theta|m)d\theta}}$$

Expectation

$$E[p(y|\theta, m)]_{p(\theta|m)}$$

Marginal likelihood

$$\int p(y, \theta|m)\, d\theta$$

Model evidence

$$p(y|m)$$

Overview

Compute the posterior and the evidence for a model

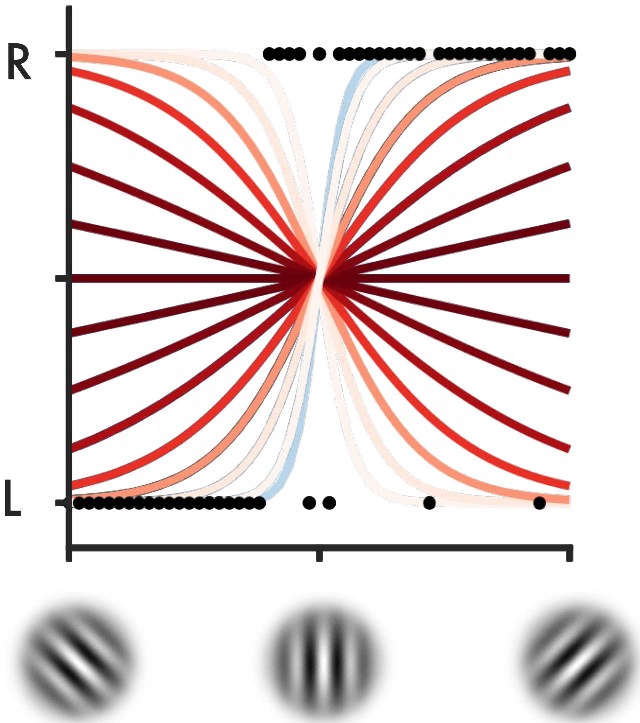
- Monte-Carlo (sampling) methods
- Variational methods

Good practices

Online supplementary material

github.com/lionel-rigoux/tutorial-bayesian-inference

Example: logistic regression



Sensitivity to orientation?

Bias?

Model prediction

$$\mathbf{p(y = 1|u, \theta, \beta) = sig(\theta u + \beta) = s}$$

Likelihood

$$\log \mathbf{p}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum \boldsymbol{y} \log \boldsymbol{s} + (\mathbf{1} - \boldsymbol{y}) \log(\mathbf{1} - \boldsymbol{s})$$

Prior

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu_\theta}, \boldsymbol{\sigma_\theta^2}) \qquad \boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{0})$$

$$\log \mathbf{p}(\boldsymbol{\theta}) = -\frac{1}{2}\left[\frac{(\boldsymbol{\theta} - \boldsymbol{\mu_\theta})^2}{\boldsymbol{\sigma_\theta^2}} + \log 2\pi\sigma^2\right]$$

Joint

$$\log \mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}) =$$

$$\sum \boldsymbol{y} \log \boldsymbol{s} + (\mathbf{1} - \boldsymbol{y}) \log(\mathbf{1} - \boldsymbol{s}) - \frac{(\boldsymbol{\theta} - \boldsymbol{\mu_\theta})^2}{\mathbf{2}\ \boldsymbol{\sigma_\theta^2}} + \boldsymbol{cst}$$

Example: logistic regression

Joint

$$\log \mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \sum \boldsymbol{y} \log \boldsymbol{s} + (\mathbf{1} - \boldsymbol{y}) \log(\mathbf{1} - \boldsymbol{s}) - \frac{(\boldsymbol{\theta} - \boldsymbol{\mu_\theta})^2}{\mathbf{2}\ \boldsymbol{\sigma_\theta^2}} + cst$$

$$\mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}) \propto \prod \ \mathbf{s}^{\ \mathbf{y}} \, (\mathbf{1} - \boldsymbol{s})^{\ \mathbf{1} - \boldsymbol{y}} \ \mathbf{e}^{-\frac{(\boldsymbol{\theta} - \boldsymbol{\mu_\theta})^2}{\mathbf{2}\ \boldsymbol{\sigma_\theta^2}}}$$

Posterior

$$\mathbf{p}(\boldsymbol{\theta}, \boldsymbol{\beta} | \mathbf{y}) \propto \mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta})$$

$$\mathbf{MAP} = \underset{\theta, \beta}{\operatorname{argmax}}\, \mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta})$$

Model evidence

$$\mathbf{p}(\mathbf{y}) = \int \mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}) \, \boldsymbol{d\theta d\beta}$$

Sampling (Monte Carlo)

Monte-Carlo methods

Expectation (theoretical mean)

$$E[z] = \sum p(z)z = \sum_{z=1}^{6} \frac{1}{6} z = 3.5$$

Variance (theoretical distance to the mean)

$$E\left[(z-3.5)^2\right] = \sum p(z)(z-3.5)^2 = 2.9167$$

Expectation ≈ Empirical mean

$$E[z] \approx \frac{1}{n} \sum_{i=1}^{n} z_i \qquad z_i \sim p(z)$$

$$E[f(z)] \approx \frac{1}{n} \sum_{i=1}^{n} f(z_i)$$

Law of Large Numbers

Monte-Carlo methods

Model evidence

Arithmetic estimator

$$\mathbf{p(y)} = \mathbf{E}[\mathbf{p(y|\theta)}]_{\mathbf{p(\theta)}} \approx \frac{\mathbf{1}}{\mathbf{n}}\sum \mathbf{p}(\mathbf{y}|\boldsymbol{\theta}_i)$$

Samples from prior

$$\boldsymbol{\theta}_i \sim \mathbf{p(\theta)}$$

Harmomic estimator, Gibb's estimator,

Annealed importance sampling, etc.

Posterior moments

Mean

$$\boldsymbol{\mu} = \mathbf{E}[\boldsymbol{\theta}]_{\mathbf{p(\theta|y)}} \approx \frac{\mathbf{1}}{\mathbf{n}}\sum \boldsymbol{\theta}_i$$

Variance

$$\boldsymbol{\Sigma} = \mathbf{E}\left[(\boldsymbol{\theta} - \boldsymbol{\mu})^2\right]_{\mathbf{p(\theta|y)}} \approx \frac{\mathbf{1}}{\mathbf{n}}\sum (\boldsymbol{\theta}_i - \hat{\boldsymbol{\mu}})^2$$

Samples from posterior

$$\boldsymbol{\theta}_i \sim \mathbf{p(\theta|y)}$$

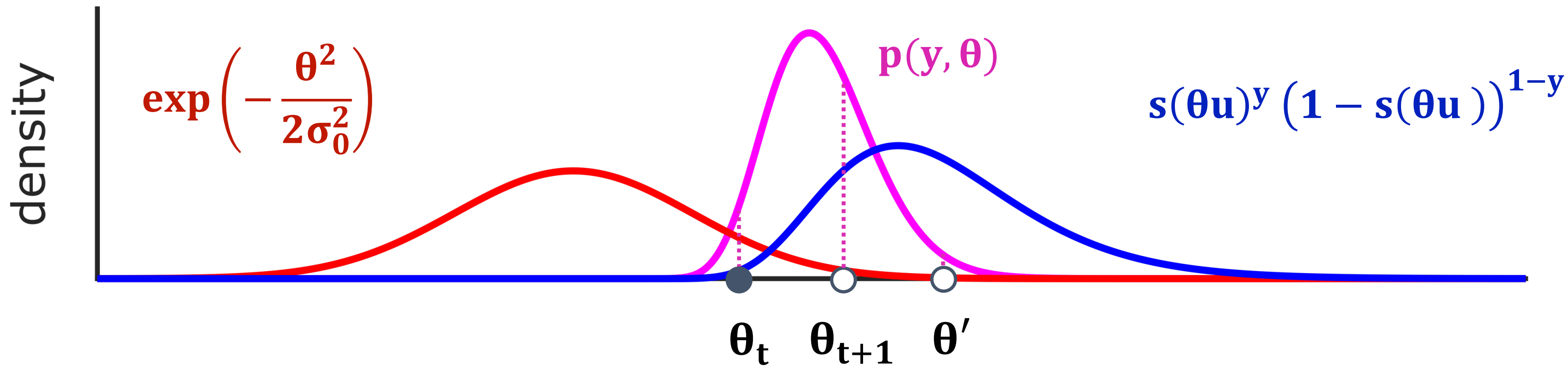
A little game

The joint as an un-normalized posterior:

$$\mathbf{p(\theta|y) \propto p(\theta)\, p(y|\theta) = p(\theta, y)}$$

- is not a probability over parameters
- gives the relative plausibility of parameter values

Metropolis-Hastings algorithm



Current state

$$p(y,\theta_t) = p(\theta_t)\,p(y|\theta_t)$$

Proposal

$$\theta' \sim q(\theta|\theta_t)$$

$$p(y,\theta') = p(\theta')\,p(y|\theta')$$

$$\alpha = \frac{p(y,\theta')}{p(y,\theta_t)}$$

$\alpha \geq 1$

Jump to proposed value

$$\theta_{t+1} = \theta'$$

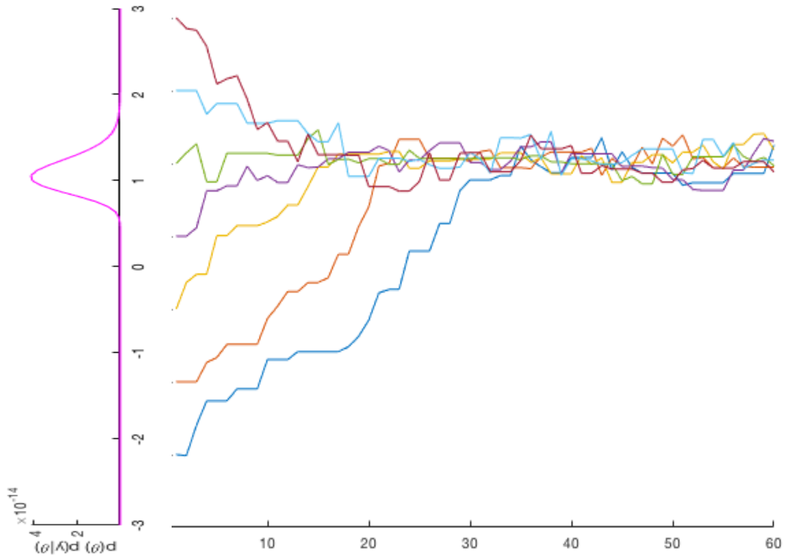
$\alpha < 1$

Draw $x \sim U(0,1)$

- if $\alpha > x$, jump
  $\theta_{t+1} = \theta'$
- else, stay in place
  $\theta_{t+1} = \theta_t$

Did I sample right?

All sampling methods requires some “post-processing” and an extensive diagnostic to ensure the samples are representative.



1) Run multiple chains
2) Check:

- Convergence (eg. Geweke)
- Mixing (eg. Gelman-Rubin)
- Autocorrelation (decimation)
- Step size (Goldilocks principle)

Multivariate case

Write conditional posteriors

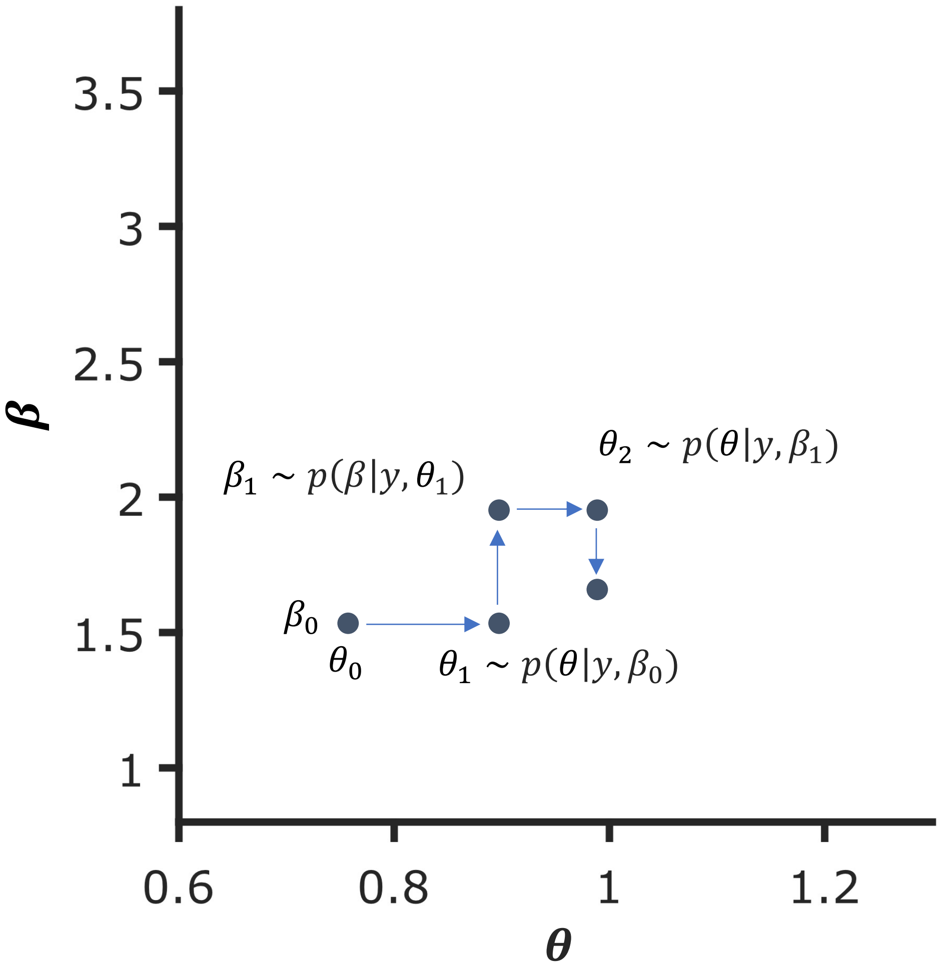
$$\mathbf{p(\theta|y,\beta) = \frac{p(y,\theta,\beta)}{p(y,\beta)}}$$

$$\mathbf{p(\beta|y,\theta) = \frac{p(y,\theta,\beta)}{p(y,\theta)}}$$

Iterative sampling

$$\mathbf{\theta_t \sim p(\theta|y,\beta_{t-1})}$$

$$\mathbf{\beta_t \sim p(\beta|y,\theta_t)}$$



Multivariate case

Using the law of large numbers:

Posterior mean

$$\mathbf{E}[\boldsymbol{\theta}|\mathbf{y}] \approx \mathbf{mean}(\boldsymbol{\theta}_\mathbf{t})$$

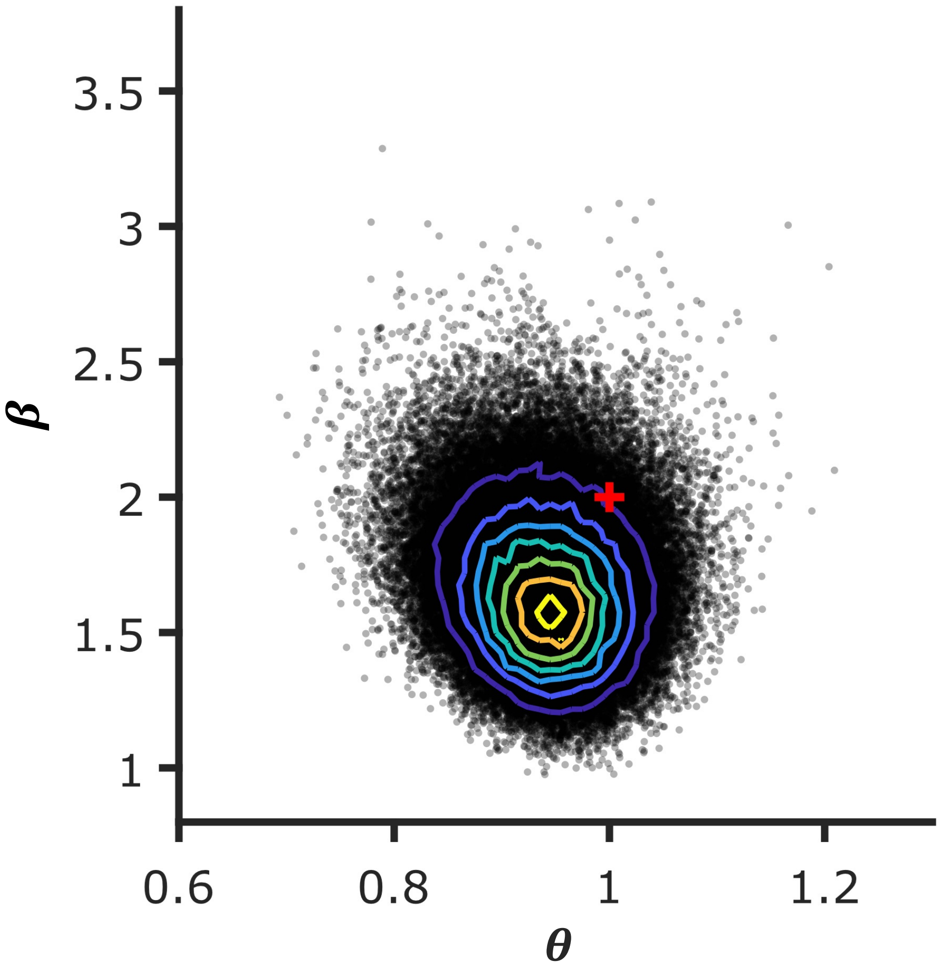
$$\mathbf{E}[\boldsymbol{\beta}|\mathbf{y}] \approx \mathbf{mean}(\boldsymbol{\beta}_\mathbf{t})$$

Posterior variance

$$\mathbf{E}\left[(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})^2\middle|\mathbf{y}\right] \approx \mathbf{var}(\boldsymbol{\theta}_\mathbf{t})$$

$$\mathbf{E}\left[\left(\boldsymbol{\beta}-\overline{\boldsymbol{\beta}}\right)^2\middle|\mathbf{y}\right] \approx \mathbf{var}(\boldsymbol{\beta}_\mathbf{t})$$

Covariance, etc.



Monte-Carlo inference

Monte-Carlo methods rely on sampling to estimate the posterior and the model evidence.

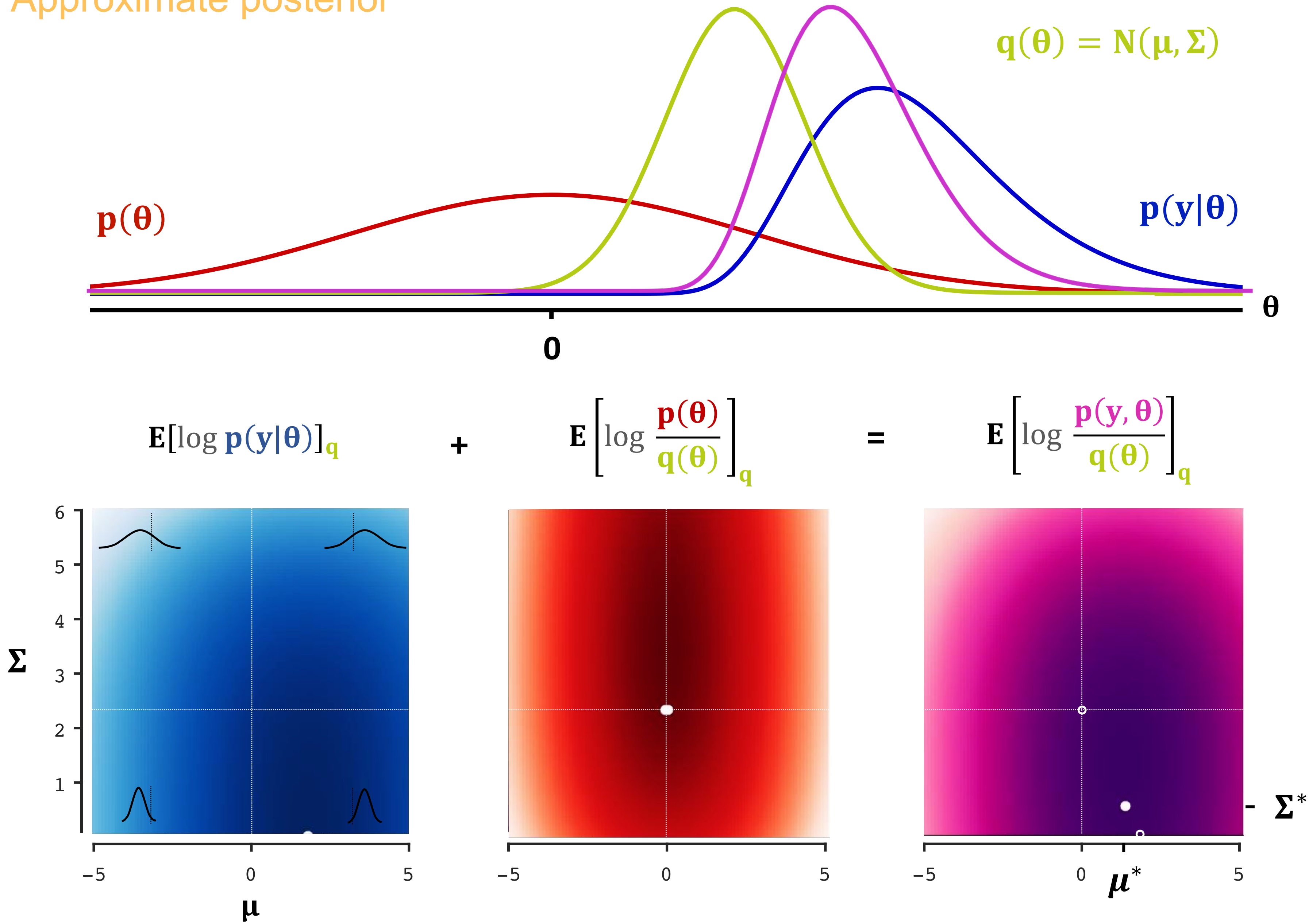
The Law of Large Numbers guarantees that the sufficient statistics of the samples will converge to the true posterior moments.

Problems

- computationally expensive
- does not scale well with the number of parameters
- hard to tune and diagnose
- no direct measure of model evidence

Variational Methods

Approximate posterior



Evidence LOwer Bound

candidate distribution $\mathbf{q}(\boldsymbol{\theta})$

Jensen‘s inequality

$$\log \mathbf{p}(\mathbf{y}) = \log \int \mathbf{p}(\mathbf{y}, \boldsymbol{\theta})\, d\theta$$

$$= \log \int \frac{\mathbf{p}(\mathbf{y}, \boldsymbol{\theta})}{\mathbf{q}(\boldsymbol{\theta})} \mathbf{q}(\boldsymbol{\theta})\, d\theta$$

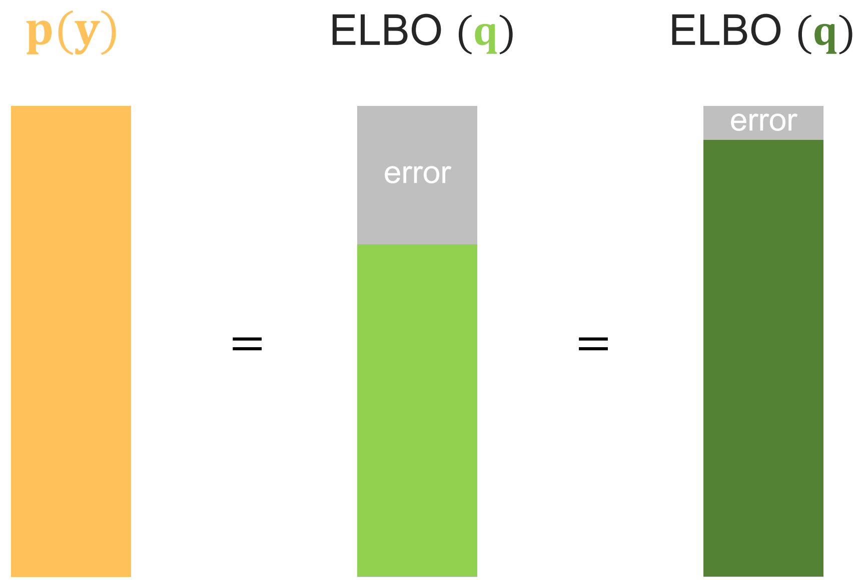
$$= \log \mathbf{E}\left[\frac{\mathbf{p}(\mathbf{y}, \boldsymbol{\theta})}{\mathbf{q}(\boldsymbol{\theta})}\right]_{\mathbf{q}(\boldsymbol{\theta})}$$

$$= \mathbf{E}\left[\log \frac{\mathbf{p}(\mathbf{y}, \boldsymbol{\theta})}{\mathbf{q}(\boldsymbol{\theta})}\right]_{\mathbf{q}(\boldsymbol{\theta})} + \mathbf{KL}[\mathbf{q}(\boldsymbol{\theta})||\mathbf{p}(\boldsymbol{\theta}|\mathbf{y})]$$

ELBO error

$< \mathbf{p}(\mathbf{y})$ $> \mathbf{0}$

Evidence LOwer Bound



integral problem **optimization problem**

Maximizing the ELBO

$$\log p(y) \approx \max \mathbf{E}\left[\log \frac{p(y,\theta)}{q(\theta)}\right]_{q(\theta)}$$

Variational Laplace

Using exponential family

$$q(\theta) = N(\mu, \Sigma)$$

Analytical approximation

$$\mathbf{ELBO} \approx \mathbf{ELBO}_{\mathrm{Laplace}}$$

Find maximum

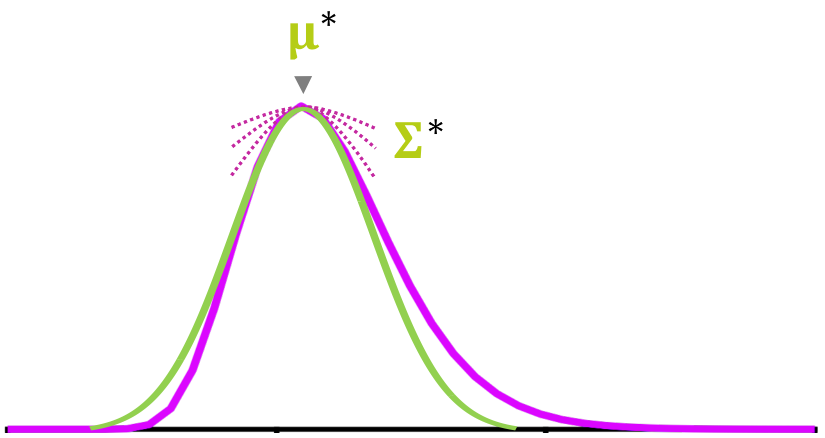
$$\frac{d}{dq(\theta)}\mathbf{ELBO}_{\mathrm{Laplace}} = 0$$

Solution

$$\mu^* = \operatorname{argmax} p(y, \theta) = \mathbf{MAP}$$

$$\Sigma^* = -\left[\left.\frac{\partial^2}{\partial\theta^2}\right|_{\mu^*} \log p(y, \theta)\right]^{-1}$$

$$\log p(y) \approx \log p(y, \mu^*) + \frac{1}{2}\left[\log|\Sigma^*| + n_\theta \log(2\pi)\right]$$



Multivariate posterior

Mean field approximation

$$\mathbf{q}(\boldsymbol{\theta}, \boldsymbol{\beta}) \approx \mathbf{q}(\boldsymbol{\theta})\mathbf{q}(\boldsymbol{\beta})$$

Variational energy

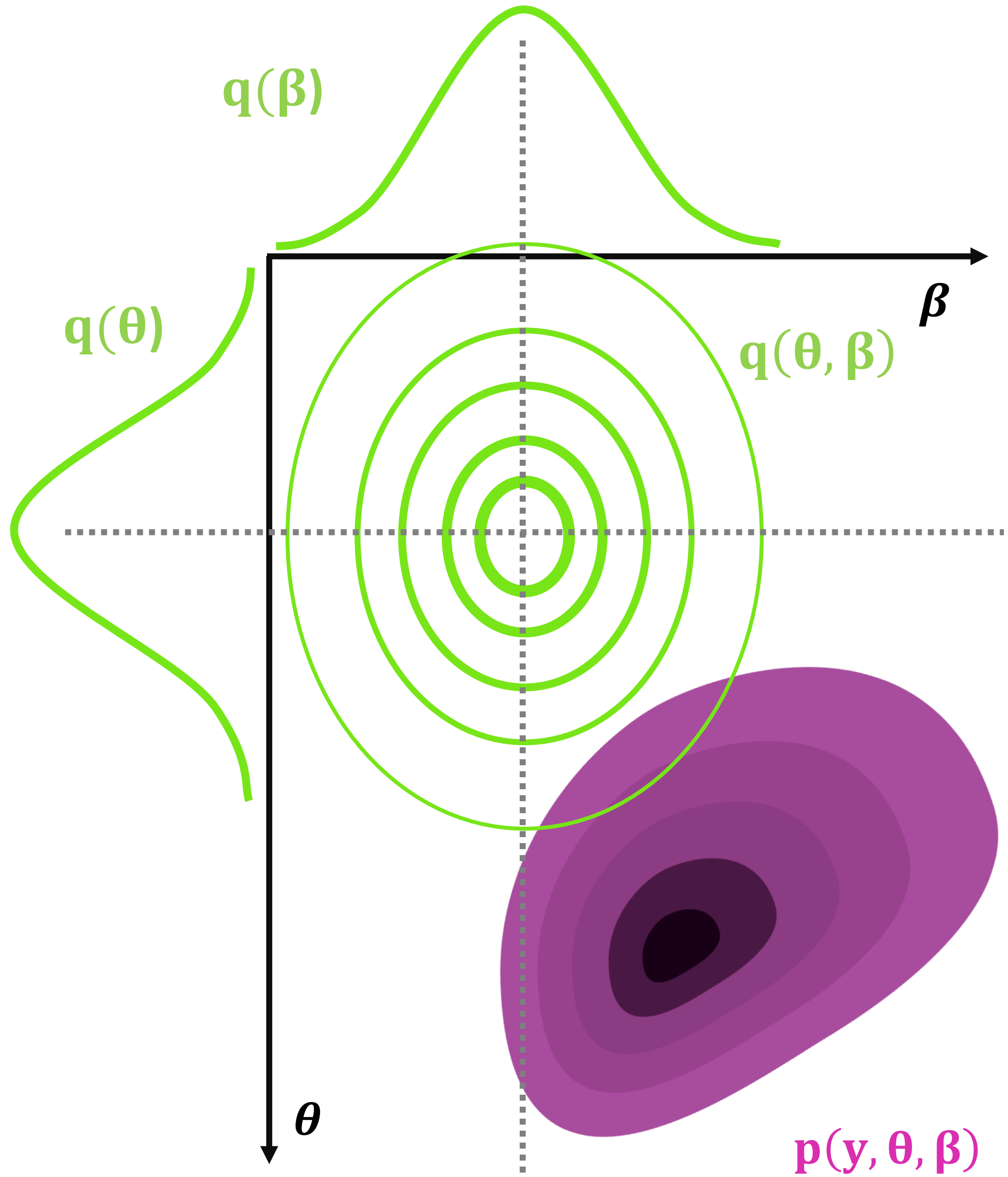
$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{E}[\log \mathbf{p}(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta})]_{\mathbf{q}(\boldsymbol{\beta})}$$

$$\approx \log \mathbf{p}\left(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\mu}_{\boldsymbol{\beta}}\right) + \ \ldots$$

Iterative optimization

$$\boldsymbol{\mu}_{\mathbf{i}} = \operatorname{argmax} \mathbf{I}(\boldsymbol{\theta}_{\mathbf{i}})$$

$$\boldsymbol{\Sigma}_{\mathbf{i}} = -\left[\left. \frac{\partial^2}{\partial \boldsymbol{\theta}_{\mathbf{i}}^2} \right|_{\boldsymbol{\mu}_{\mathbf{i}}} \mathbf{I}(\boldsymbol{\theta}_{\mathbf{i}}) \right]^{-1}$$



Multivariate posterior

Mean field approximation

$$q(\theta, \varphi) \approx q(\theta)q(\beta)$$

Maximise Variational energy

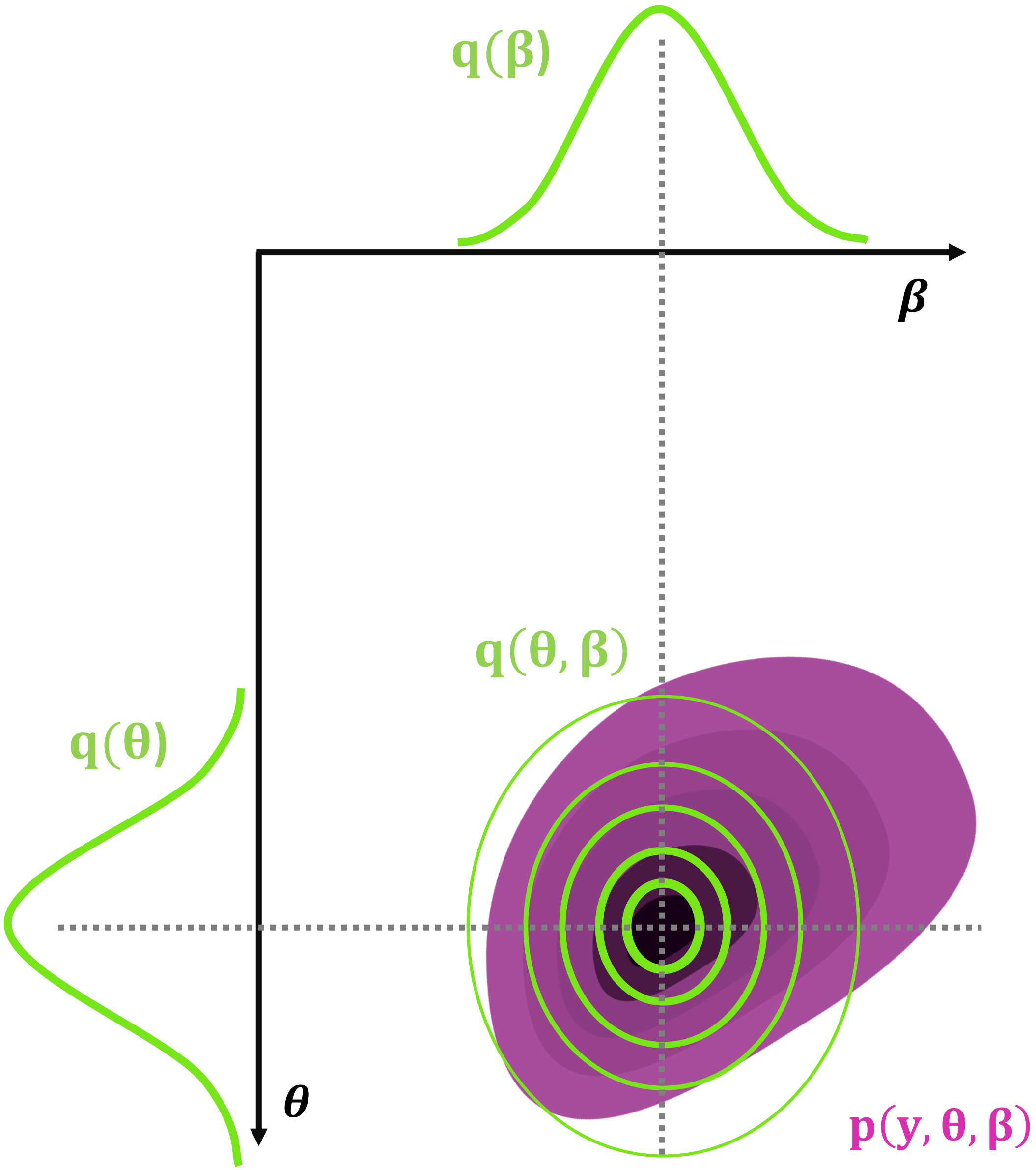
$$I(\theta) = E[\log p(y, \theta, \beta)]_{q(\beta)}$$

$$\approx \log p(y, \theta, \mu_\beta) + \ \ldots$$

Iterative optimization

$$\mu_i = \operatorname{argmax} I(\theta_i)$$

$$\Sigma_i = -\left[\left. \frac{\partial^2}{\partial \theta_i^2} \right|_{\mu_i} I(\theta_i) \right]^{-1}$$



Maximizing the ELBO

$$\log \mathbf{p}(\mathbf{y}) \approx \max \mathbf{E}\left[\log \frac{\mathbf{p}(\mathbf{y},\boldsymbol{\theta})}{\mathbf{q}(\boldsymbol{\theta})}\right]_{\mathbf{q}(\boldsymbol{\theta})}$$

Stochastic gradient

Using samplable distribution

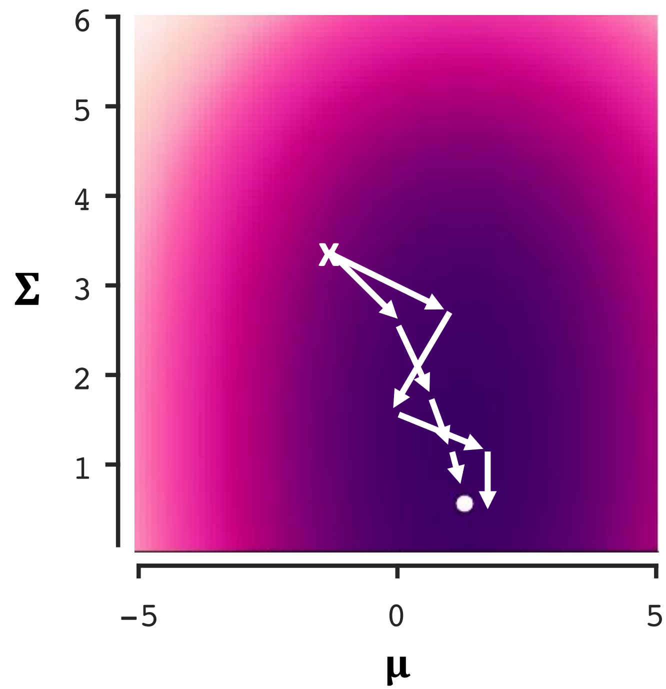
$$\mathbf{q}(\boldsymbol{\theta}) = \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Gradient

$$\nabla \mathbf{ELBO} = \mathbf{E}\left[\nabla \log \mathbf{q}(\boldsymbol{\theta}) \left(\log \frac{\mathbf{p}(\mathbf{y},\boldsymbol{\theta})}{\mathbf{q}(\boldsymbol{\theta})}\right)\right]_{\mathbf{q}(\boldsymbol{\theta})}$$

Solution

Ascend (MC approximation)



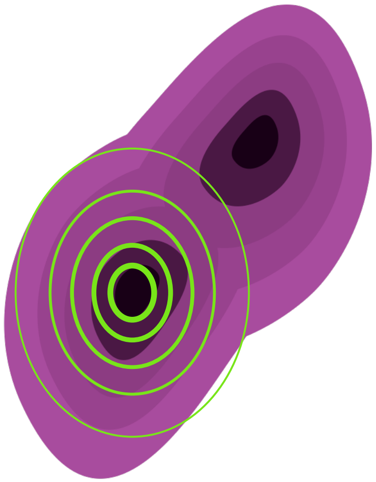
Variational inference

Summarize the posterior to its sufficient statistics (mean, variance) and optimize those values wrt the ELBO.

This requires multiple approximations (Jensen/Free-energy, Gaussian posterior, Laplace, mean-field) to be tractable.

Problems

- does not converge to the true posterior
- can get stuck in local optimum



Take home message

Model evidence (normalization factor of the posterior) is in general intractable and calls for numerical methods.

- ✓ Sampling methods give a computationally expensive estimation of the true posterior.
- ✓ Variational methods are fast & scalable computations of an approximation of the posterior.
- ✓ Other techniques in development: Deep Bayesian Inversion

Software

Variational

- VBA-toolbox
- TAPAS
- SPM

Sampling

- STAN
- BUGS
- JAGS
- hBayesDM
- hddm

VBA Toolbox

282 published papers

85 demos (tutorial, Q-learning, HGF, DCMs, etc)

Online wiki + Q&A

Simulation

Inversion (single subject, hierarchical)

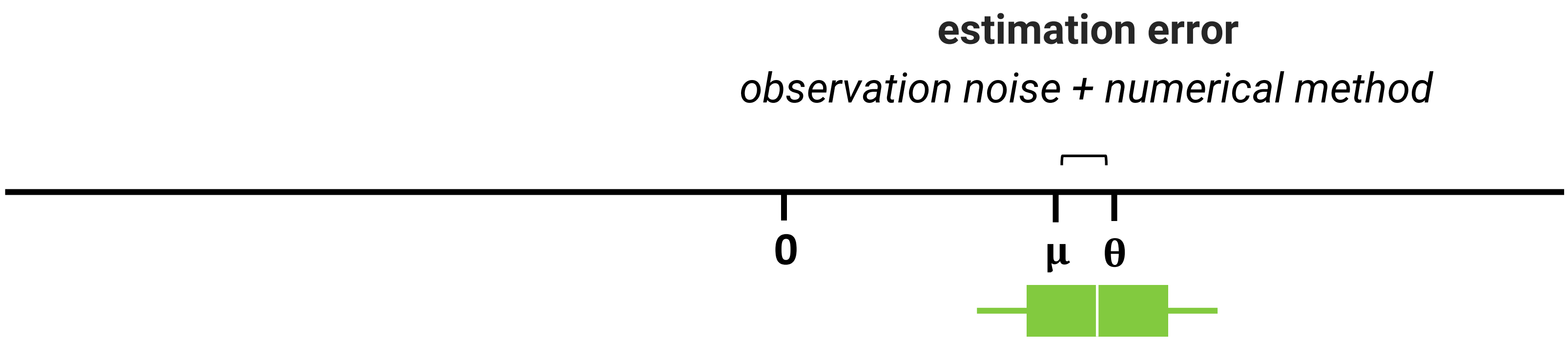
Model selection (families, btw groups, btw conditions)

Visual diagnostics

Design optimization, multisession, multimodal observations, …

Need only the model description!

Validating your pipeline: parameter identifiability



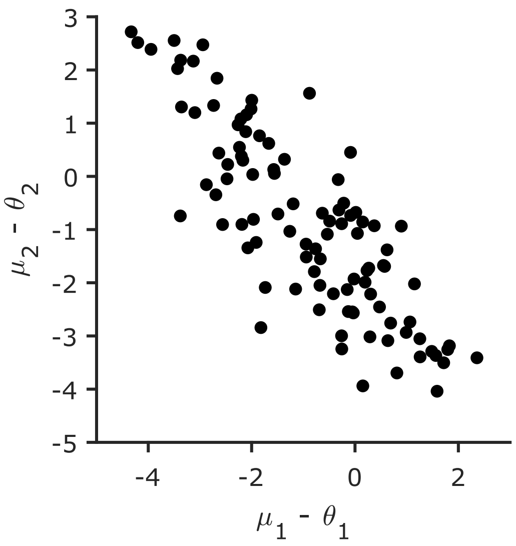
simulate data using your design with a realistic $\boldsymbol{\theta}$

do check if model predictions do emerge

invert your model (find $\boldsymbol{\mu}$)

compute estimation error ($\boldsymbol{\mu} - \boldsymbol{\theta}$)

- check effect of prior mean
- check effect of prior variance
- assess overfitting
- check for posterior cov / error correlation



Thank you!

Online supplementary material

github.com/lionel-rigoux/tutorial-bayesian-inference

- interactive app
- code of all algorithms
- selected references

VBA-Toolbox

mbb-team.github.io/VBA-toolbox

Easy and reproducible writing workflow

pandemics.gitlab.io