Fitting a model: VB & MCMC

Lionel Rigoux

Translational Neuro-Circuitry (TNC) Cologne

Translational Neuromodeling Unit (TNU) Zürich

Max-Planck-Institut für Stoffwechselforschung

online supplementary material

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

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

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

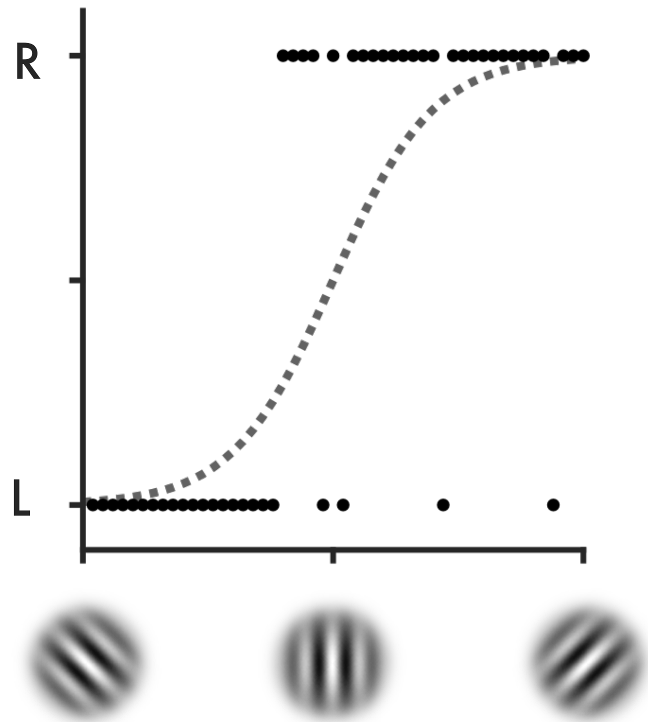
Marginal likelihood

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

Model evidence

$$p(y|m)$$

Example: logistic regression



Sensitivity to orientation?

Bias?

Sampling (Monte Carlo)

Monte-Carlo methods

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

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

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

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

Law of
Large Numbers

Monte-Carlo methods

Model evidence

$$p(y) = E[p(y|\theta)]_{p(\theta)} \approx \frac{1}{n}\sum_{i=1}^{n} p(y|\theta_i)$$

$$\theta_i \sim p(\theta)$$

Posterior moments

$$\mu = E[\theta]_{p(\theta|y)} \approx \frac{1}{n}\sum_{i=1}^{n} \theta_i$$

$$\theta_i \sim p(\theta|y)$$

$$\Sigma = E\left[(\theta - \mu)^2\right]_{p(\theta|y)} \approx \frac{1}{n}\sum_{i=1}^{n} (\theta_i - \hat{\mu})^2$$

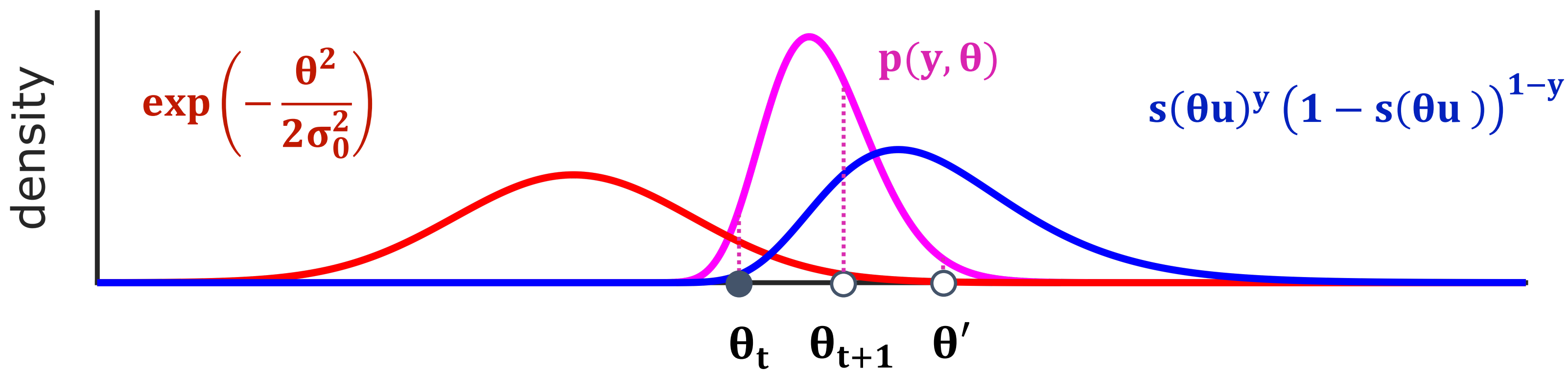
A little game

The joint as an un-normalized posterior:

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

- 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

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

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

- If $\alpha > x$, jump

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

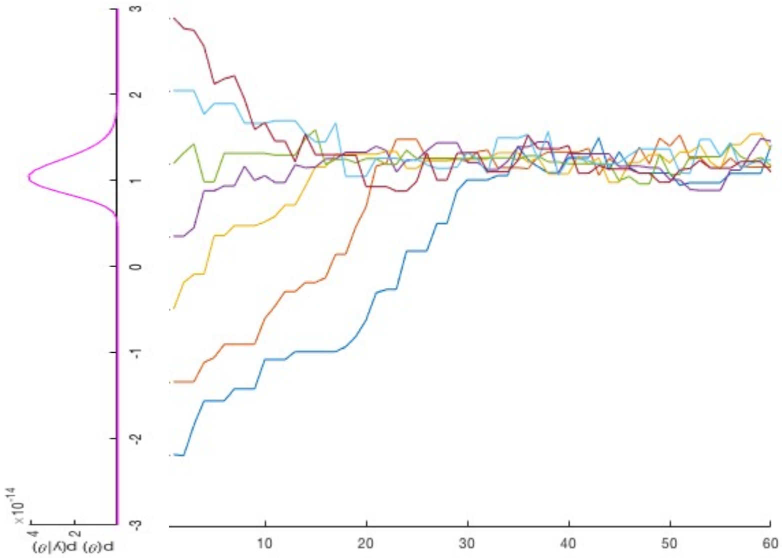
$\alpha < 1$

- 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
- Step size (Goldilocks principle)

Multivariate case

write conditional posteriors

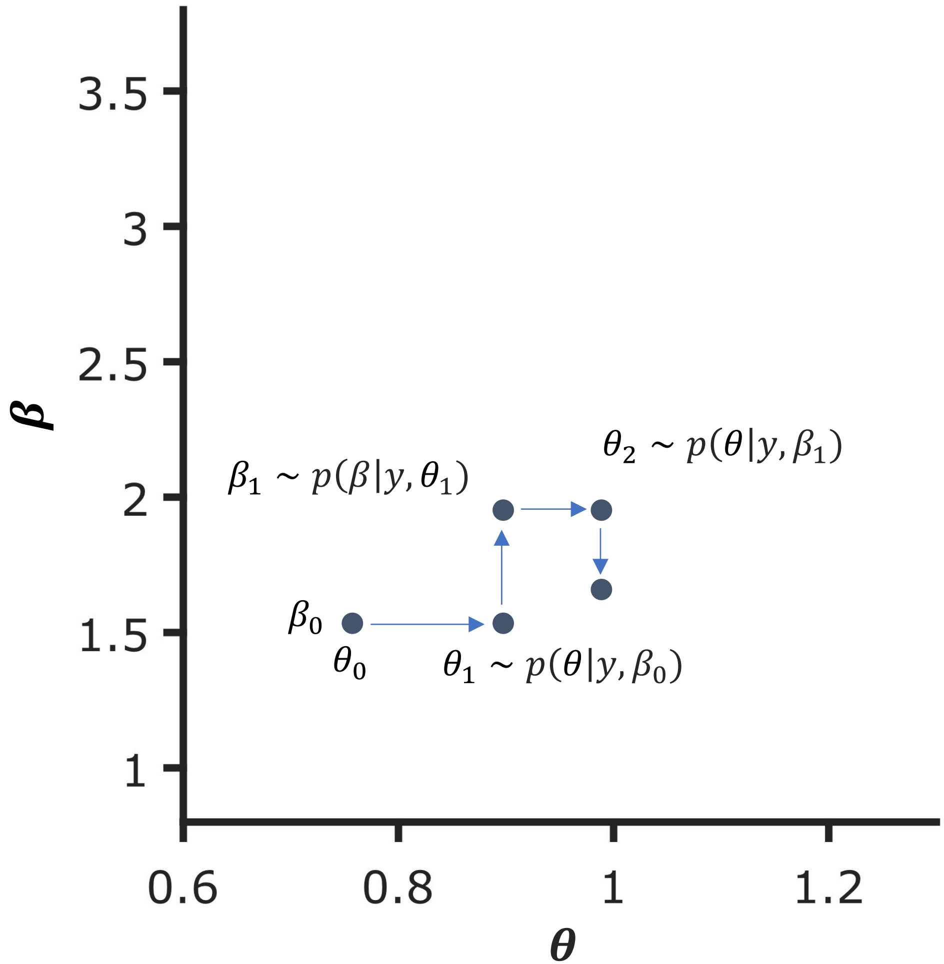
$$\mathbf{p(\theta|y,\beta) = \frac{p(y,\theta,\beta)}{p(y,\beta)}}$$

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

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

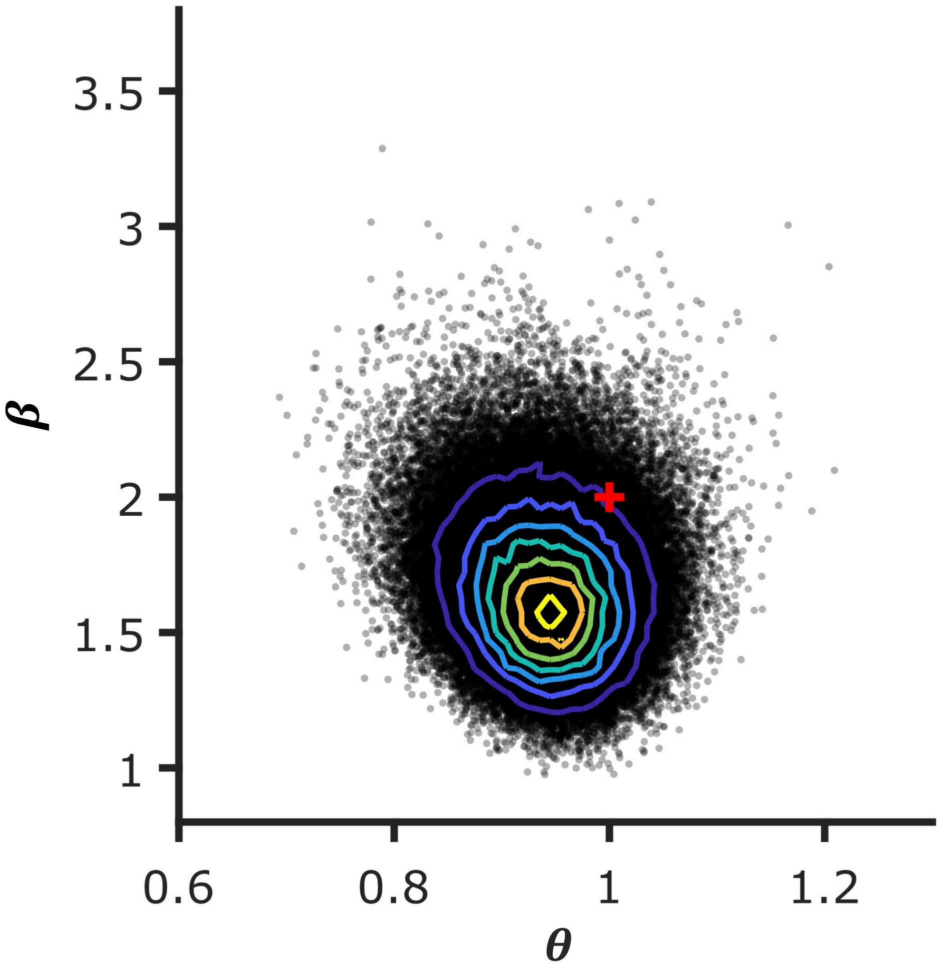
$$E[\theta|y] \approx mean(\theta_t)$$

$$E[\beta|y] \approx mean(\beta_t)$$

- Posterior variance

$$E\left[(\theta-\bar{\theta})^2|y\right] \approx var(\theta_t)$$

$$E\left[\left(\beta-\bar{\beta}\right)^2\middle|y\right] \approx var(\beta_t)$$



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
- no direct measure of model evidence
- hard to tune and diagnose

Variational Methods

The ELBO

model evidence

$$\log p(y) = \log \int p(y|\theta)p(\theta)\, d\theta$$

$$= \log \int p(y|\theta)p(\theta)\frac{q(\theta)}{q(\theta)}\, d\theta \quad = \log E\left[p(y|\theta)\frac{p(\theta)}{q(\theta)}\right]_{q(\theta)}$$

Jensen's inequality

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

ELBO or Free Energy error (positive)

$$E[\log p(y|\theta)]_{q(\theta)} - KL[p(\theta)||q(\theta)]$$

expected log-likelihood "distance" to prior

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

Variational Laplace

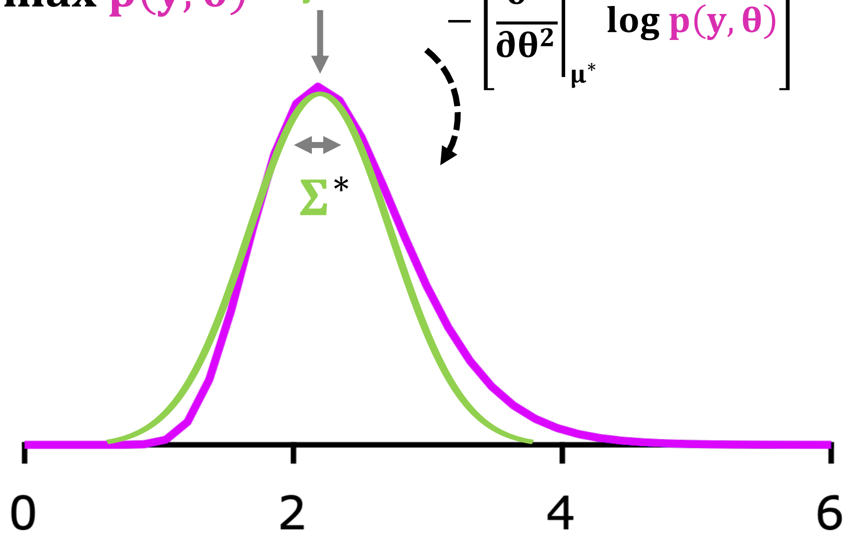
Maximize $\mathbf{F} = \mathbf{E}\left[\log \dfrac{p(y,\theta)}{q(\theta)}\right]_q$

Stochastic gradient

Analytical approximation: $\mathbf{F} \approx \mathbf{F}_{\mathbf{Laplace}}$

Find maximum: $\dfrac{d}{dq(\theta)} \mathbf{F}_{\mathbf{Laplace}} = 0$

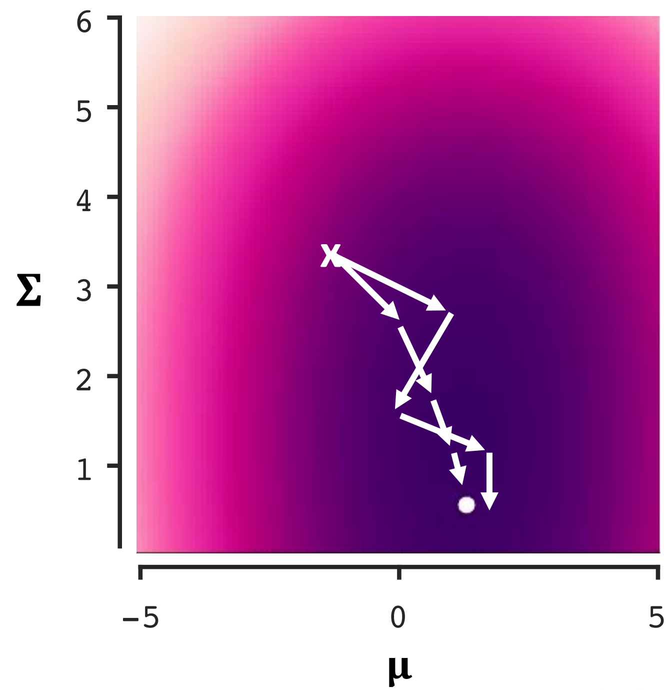
$\max p(y,\theta)$ μ^* $-\left[\dfrac{\partial^2}{\partial\theta^2}\Big|_{\mu^*} \log p(y,\theta)\right]^{-1}$



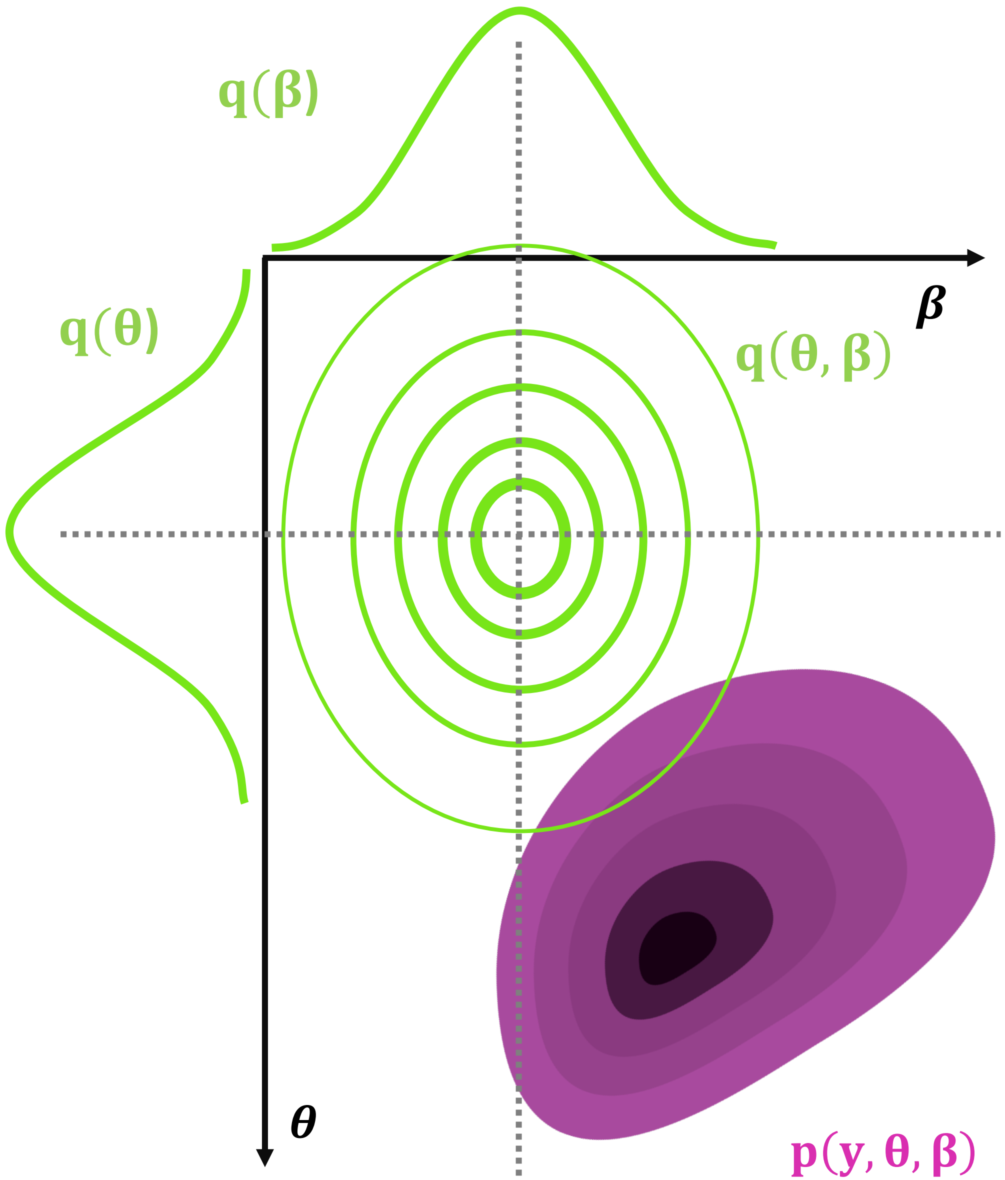
$$\log p(y) \approx \log p(y,\mu^*) + \frac{1}{2}[\log|\Sigma^*| + n_\theta \log(2\pi)]$$

Find maximum: **gradient ascent**

$$\nabla \mathbf{F} = \mathbf{E}_q[\nabla \log q(\theta)\ (\log \frac{p(y,\theta)}{q(\theta)})\ ]$$



Multivariate posterior



Mean field approximation

$$p(\theta, \beta | y) \approx p(\theta | y) p(\beta | y)$$

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

Maximise Variational energy

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

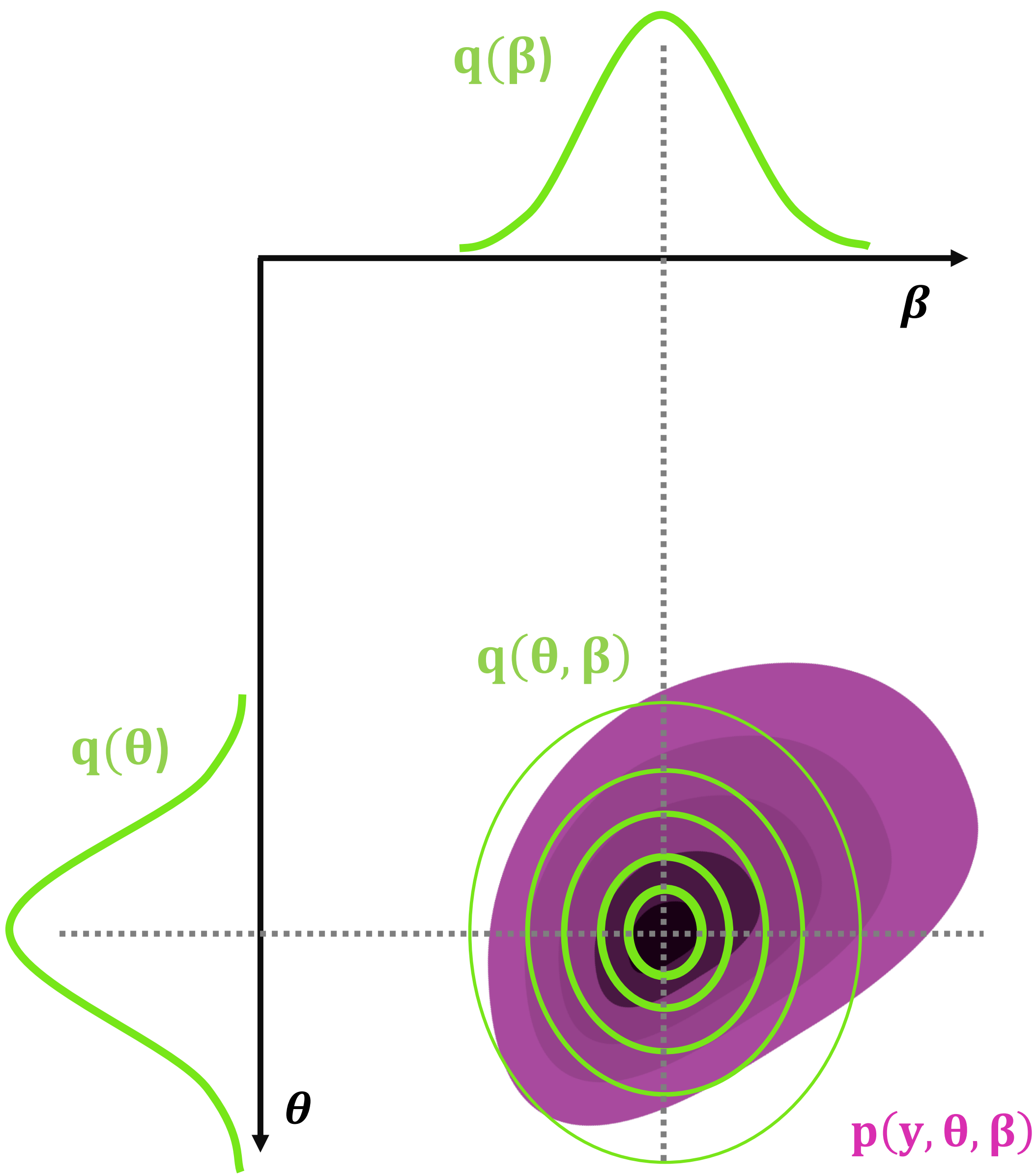
$$\approx \log p(y, \theta, \mu_\beta) + \ ...$$

Iterative optimization

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

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

Multivariate posterior



Mean field approximation

$$p(\theta, \beta|y) \approx p(\theta|y)p(\beta|y)$$

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

Maximise Variational energy

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

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

Iterative optimization

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

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

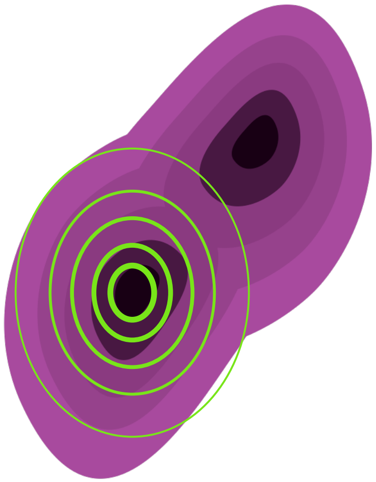
Variational inference

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.

Software

- Variational
  - VBA-toolbox
  - TAPAS
  - SPM
- Sampling
  - STAN
  - BUGS
  - JAGS
  - hBayesDM
  - hddm

VBA Toolbox

> 200 published papers

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

Online wiki + Q&A

Simulation

Inversion (single subject, hierarchical)

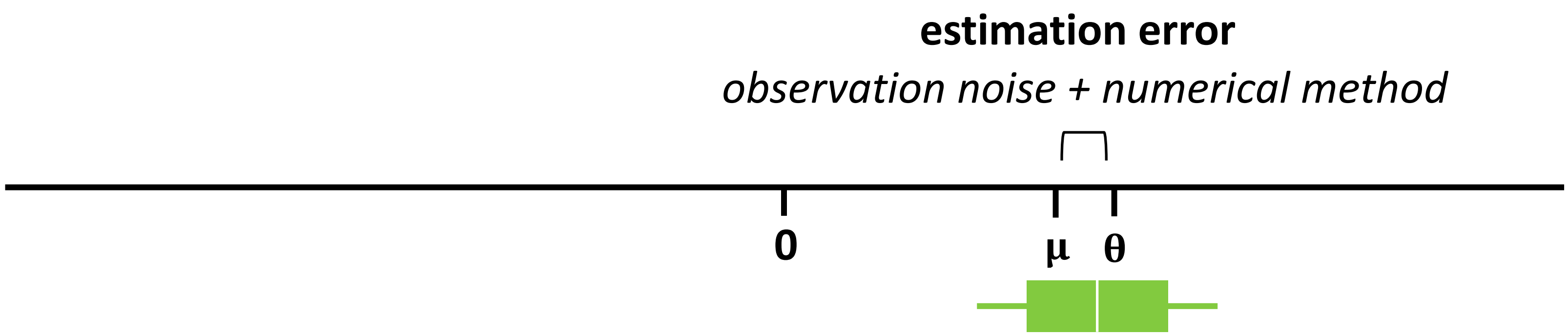
Model selection (families, btw groups, btw conditions)

Visual diagnostics

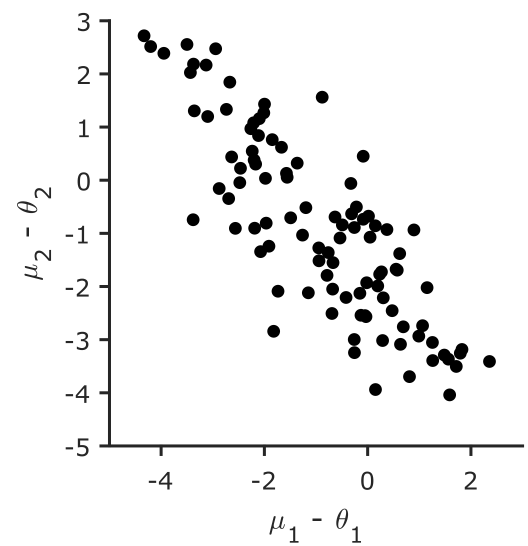
Design optimization, multisession, multimodal observations, ...

Need only the model description!

Validating your pipeline: parameter identifiability



- simulate data using your design with a realistic θ
- do check if model predictions do emerge
- invert your model (find μ)
- compute estimation error $(\mu - \theta)$
    - 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

VBA-Toolbox

mbb-team.github.io/VBA-toolbox

Easy writing workflow

pandemics.gitlab.io