Sampling is done: what now?

Sytems Biology for Scientific Computing: week four

Introduction

Recap

Last time we had a look at how to run a Bayesian analysis of data from a Monod process using Stan.

The model had

- a **non-normal measurement process**
- a **hierarchical component** to model between-strain variation
- an **ode system** to represent structural knowledge

Plan for today

Slides: what to do after you run MCMC?

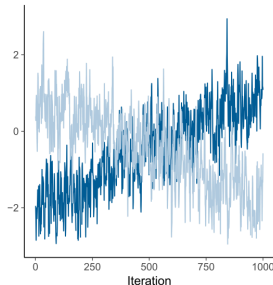**Computer**: apply these ideas to the Monod analysis.
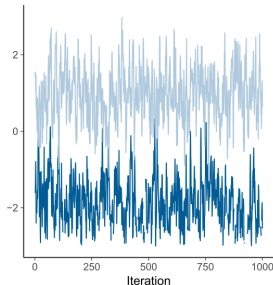
Diagnostics

$\hat{R}$: did the chains converge?

i.e.

- Do they agree?
- Are they stationary?

$\hat{R}$ should be close to 1.

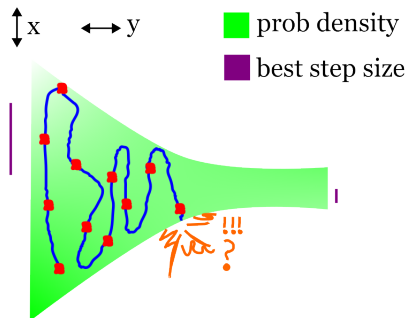Find out more: Vehtari et al.

(2021)

Divergent transitions: did HMC's numerical integration fail?

HMC is unhappy when there are different optimal step sizes in different parts of parameter space.

Divergent transitions usually mean biased sample.

This is a good feature of HMC!

Find out more: Betancourt (2017).

Finding things out

Finding things out

Thing you want to know $\approx$ Expectation of a quantity in the model

$$\approx \int f(\theta)p(\theta)d\theta$$

NB:

- Use Monte Carlo Standard Error to see if you have enough samples
- MCSE can vary for different expectations of the same quantity

Model comparison and evaluation

Very quick decision theory

Loss function: If the observation is y and the model says $p(y) = z$, how bad is that?

To choose a model, try and minimise expected loss.

To estimate expected loss, make some relevant predictions.

(relevant = probably not in-sample, not all from the same patient, etc).

Which loss function is best depends on the problem.

Find out more: Vehtari and Ojanen (2012)

Log likelihood

A good default loss function:

$$loss(y, p(y)) = -\ln p(y)$$

Out of sample log likelihood can often be estimated without MCMC.

For exact out of sample log likelihood, use cross validation.

Find out more: (Landes and Williamson 2013, sec. 2.3) Vehtari, Gelman, and Gabry (2017)

Next time

Next time

Handling projects with more than one model.

References

References I

Betancourt, Michael. 2017. "Diagnosing Biased Inference with Divergences."
  Betanalpha.github.io. https:
  //github.com/betanalpha/knitr_case_studies/tree/master/divergences_and_bias.

Landes, Jürgen, and Jon Williamson. 2013. "Objective Bayesianism and the Maximum
  Entropy Principle." *Entropy* 15 (12): 3528–91.
  https://doi.org/10.3390/e15093528.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian Model
  Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and
  Computing* 27 (5): 1413–32. https://doi.org/10.1007/s11222-016-9696-4.

References II

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. "Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion)." *Bayesian Analysis* 16 (2): 667–718. https://doi.org/10.1214/20-BA1221.

Vehtari, Aki, and Janne Ojanen. 2012. "A Survey of Bayesian Predictive Methods for Model Assessment, Selection and Comparison." *Statistics Surveys* 6 (none): 142–228. https://doi.org/10.1214/12-SS102.