Bayesian Statistics - Homework 2

Gabry et al., *Visualization in Bayesian workflow* (2018)

Roberto Corti

May 12, 2021

Visualization in Bayesian workflow

A pipeline for our work

Visualization is an invaluable way of justifying and criticize a statistical model.
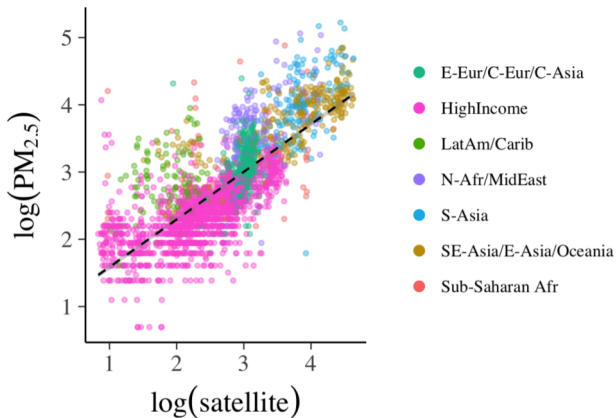
Phases of statistical workflow:

- ▶ Set up an initial model
- ▶ Model check
- ▶ Computational checks for the inference algorithm
- ▶ Posterior predictive checks
- ▶ Model comparison

Example: Estimate global $PM_{2.5}$ concentration

Exploratory data analysis

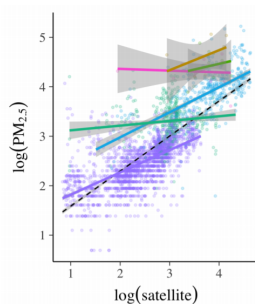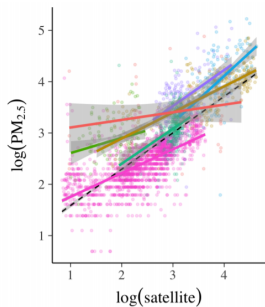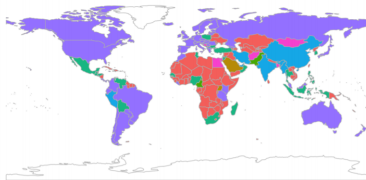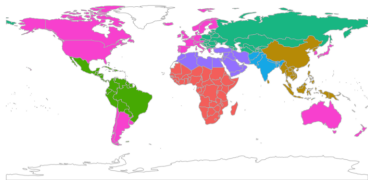## More than just plotting the data

Exploratory data analysis

More than just plotting the data

Exploratory data analysis
More than just plotting the data

- ▶ **Model 1**: simple linear regression,

- ▶ **Model 2**: multilevel model where observations are stratified by WHO super-regions,

- ▶ **Model 3**: multilevel model where observations are stratified by clustered super-region.
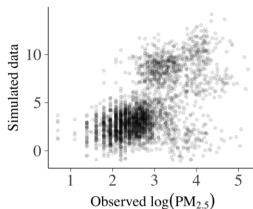
Prior predictive checking
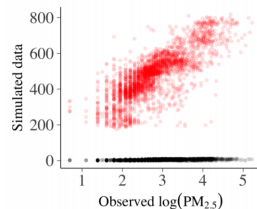
Fake data can be almost as valuable as real data

Generative model: $\theta^* \sim p(\theta) \longrightarrow y^* \sim p(y|\theta^*) \Longleftrightarrow y^* \sim p(y)$
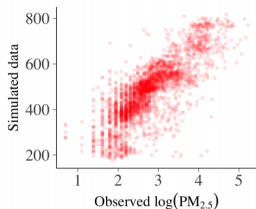
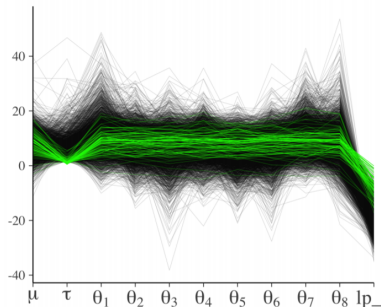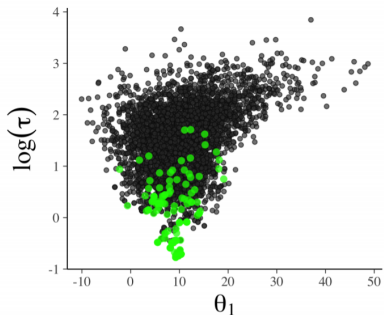

(a) Vague priors   (b) Weakly informative priors   (c) Comparison
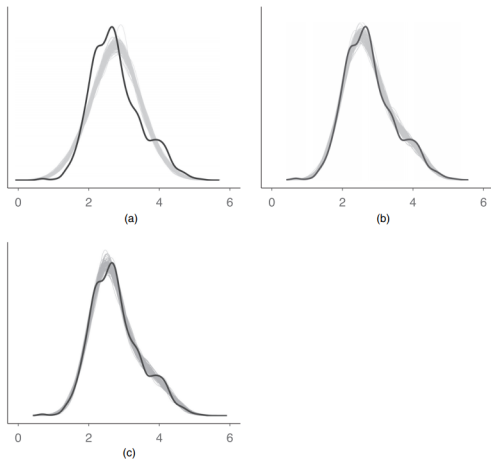
MCMC diagnostics

Moving beyond trace plots

Posterior predictive checks

How did we do?

Posterior predictive distribution: $p(\widetilde{y}|y) = \int d\theta \; p(\widetilde{y}|\theta)p(\theta|y)$
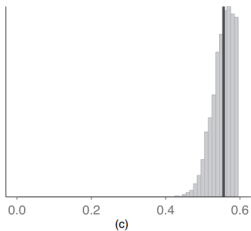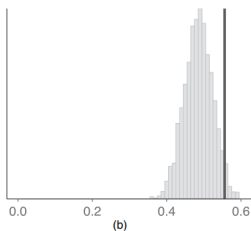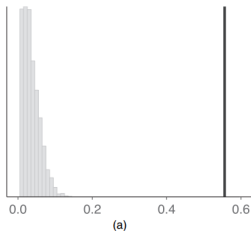
Posterior predictive checks

How did we do?

Posterior predictive distribution: $p(\widetilde{y}|y) = \int d\theta \; p(\widetilde{y}|\theta)p(\theta|y)$

Model comparison

Looking *when* and *where* a model is better than another

LOO predictive distrubution: $p(y_i|y_{-i})$

(a)

(b)