Model checking

ISTA 410 / INFO 510: Bayesian Modeling and Inference

U. of Arizona School of Information

October 7, 2020

Outline

Last week:

- Metropolis-Hastings algorithm
- Gibbs sampler
- Hamiltonian Monte Carlo

Today:

- Assessing models with predictive checks

Posterior and prior predictive checks

Goals

We want to answer the following questions:

- Do the inferences from the model make sense?
- Can the model reproduce features of interest in the original data?

Tools for today:

- Posterior predictive checks
- Prior predictive checks

See Ch. 6 in BDA for more details – note that the use of p-values

Predictive checking

Recall:

- Bayesian models are generative: they give a framework for generating data given parameter values
- In the presence of a (hyper)prior, they allow generating data with or without fixed values of the parameter

Predictive checking

Recall:

- Bayesian models are generative: they give a framework for generating data given parameter values
- In the presence of a (hyper)prior, they allow generating data with or without fixed values of the parameter
- Goals of predictive checking:
    - Prior: confirm that the prior model makes possible predictions
    - Posterior: confirm that the posterior (fitted) model makes predictions that resemble existing data

Example: speed of light measurements

Example

Example (from BDA section 3.3 and 6.3): Newcomb's speed of light measurements.

Model:

$$
\begin{aligned}
y_i &\sim \mathrm{Normal}(\mu, \sigma) \\
p(\mu, \sigma) &\propto \sigma^{-1}
\end{aligned}
$$

See section 3.3 for a brief description of the model.

Assessing the model

How can we assess the accuracy of this model?

- External validation: Compare model predictions to new observations
    - Model estimates the speed of light inaccurately based on current estimates
    - But, this is more because of the experimental design and limitations of data collection

Assessing the model

How can we assess the accuracy of this model?

- External validation: Compare model predictions to new observations
  - Model estimates the speed of light inaccurately based on current estimates
  - But, this is more because of the experimental design and limitations of data collection
- Internal validation: Assess model accuracy / plausibility with the data we already have
  - Do the model predictions look right *relative to the data we have*?

Posterior predictive check for the speed-of-light

Let's do the simplest posterior predictive check for the speed-of-light model:

- Generate 66 observations from the posterior predictive distribution
- Repeat many times
- View histograms of these sets of 66 observations

Posterior predictive check for the speed-of-light

Let's do the simplest posterior predictive check for the speed-of-light model:

- Generate 66 observations from the posterior predictive distribution
- Repeat many times
- View histograms of these sets of 66 observations
- If we notice anything odd, drill down on that

Revising the model

What should we do?

- We see that the model reproduces some properties of the data, but the two outliers are not consistent with the model
- Problem: normal distribution has short tails, won't predict extreme outliers

Updating the model:

- Replace the model likelihood with something that has heavier tails:

$$
\begin{aligned}
y_i &\sim \mathrm{Normal}(\mu, \sigma) \\
y_i &\sim \mathrm{StudentT}(\nu, \mu, \sigma) \\
y_i &\sim \mathrm{Cauchy}(\alpha, \beta)
\end{aligned}
$$

Revising the model

What should we do?

- We see that the model reproduces some properties of the data, but the two outliers are not consistent with the model
- Problem: normal distribution has short tails, won't predict extreme outliers

Updating the model:

- Replace the model likelihood with something that has heavier tails:

$$y_i \sim \mathrm{Normal}(\mu, \sigma)$$

$$y_i \sim \mathrm{Cauchy}(\alpha, \beta)$$

Revising the model

What should we do?

- We see that the model reproduces some properties of the data, but the two outliers are not consistent with the model
- Problem: normal distribution has short tails, won't predict extreme outliers

Updating the model:

- Replace the model likelihood with something that has heavier tails:

$$y_i \sim \mathrm{Normal}(\mu, \sigma)$$
$$y_i \sim \mathrm{StudentT}(\nu, \mu, \sigma)$$
$$y_i \sim \mathrm{Cauchy}(\alpha, \beta)$$

Choice of test statistics

Double use of data:

- We are using the data twice: once for fitting and once for checking
- Leads to problems if the summary statistics we are looking at are directly related to model parameters
  - Example in book: use of variance as a summary statistic for the speed of light model. Variance is one of the fitted model parameters, so it is unsurprising that the variance of posterior predictive data fits the variance of observed data.

Prior predictive checking and a case study

Prior predictive checking

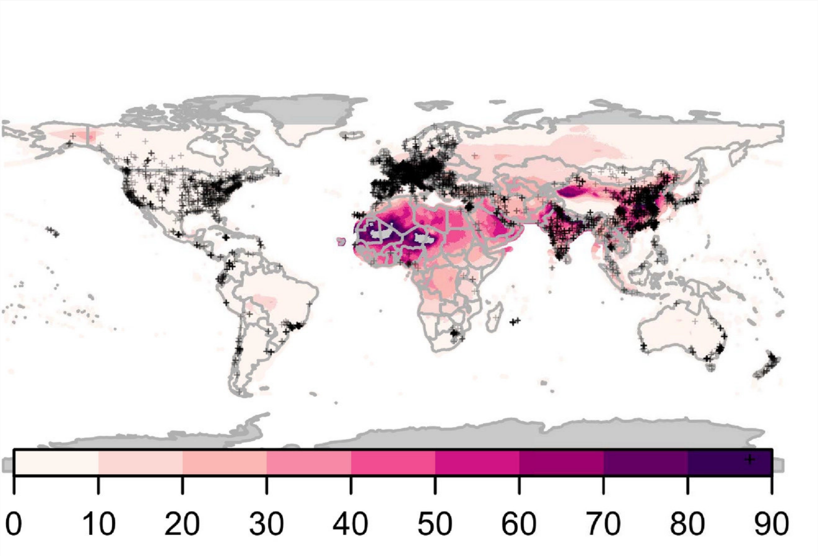
- The book doesn't mention prior predictive checks, but they are increasingly popular
- We used them a bit before: regression models and impossible slopes
- Example case study: Gabry, Simpson, Vehtari, Betancourt, Gelman, "Visualization in Bayesian workflow" (`https://doi.org/10.1111/rssa.12378`)
- In addition to prior and posterior predictive checks, features visualization of divergent HMC transitions (cf. last week)

Study overview

Problem overview:

- Particulate matter with size < 2.5 microns (PM2.5) is a public health hazard
- Want to study global health effects, but need air quality data at high geospatial resolution
  - In some areas, have ground-level monitors
  - In others, use optical data from satellites
  - Goal: fit a model predicting ground-level PM2.5 concentration from satellite predictors as a way of calibrating the satellite data
- Three models:
  - (a) a simple linear model
  - Two hierarchical models, stratified on (b) WHO region designations or (c) regions clustered by average PM2.5 concentrations as measured by ground level detectors

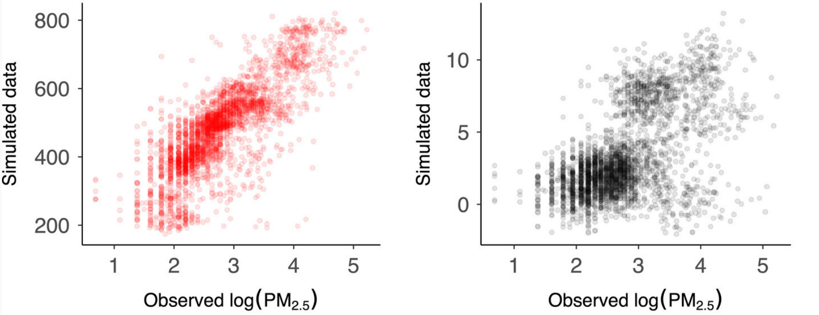
Why hierarchical models?



Prior predictive checking

vague prior

weakly informative prior



One realization from the vague prior and one from the weakly informative prior.

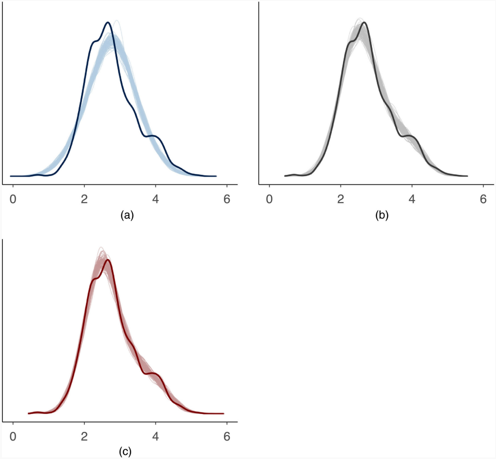
Notice the y scale! It's on a log scale; the numbers coming out of the vague prior are wildly unrealistic.

Posterior predictive checking

Two posterior predictive checks to compare the fit of the three models:

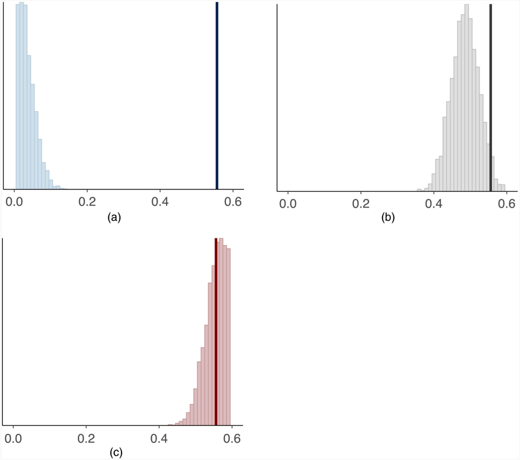
- Observed data vs. posterior predictive simulations
- Skewness as a summary statistic

Posterior check 1



Density estimate of observed data and replications of model-generated data

Posterior check 2



Skewness of replications of model-generated data

Summary

- Posterior (and prior) predictive checks can be used to evaluate fit of the model to data

Further reading:

- "Visualization and Bayesian Workflow" `https://doi.org/10.1111/rssa.12378`
- Betancourt, "Towards a Principled Bayesian Workflow" `https://github.com/betanalpha/jupyter_case_studies/blob/master/principled_bayesian_workflow/principled_bayesian_workflow.ipynb`

Next week:

- Information criteria
- Cross-validation methods