Bayesian model validation and evaluation

Tutorial at the MPIDR Summer Incubator Program

Michael Chong | 20 June 2023

Learning Objectives

How predictions are made from Bayesian models

Three tools* to assess predictions from Bayesian models (* three of many possible tools)

Implementation of these assesment tools in R

Along the way

Fit Bayesian models using R formula syntax à la `lm()` or `glm()`

Have fun with statistics!!!!

Recap

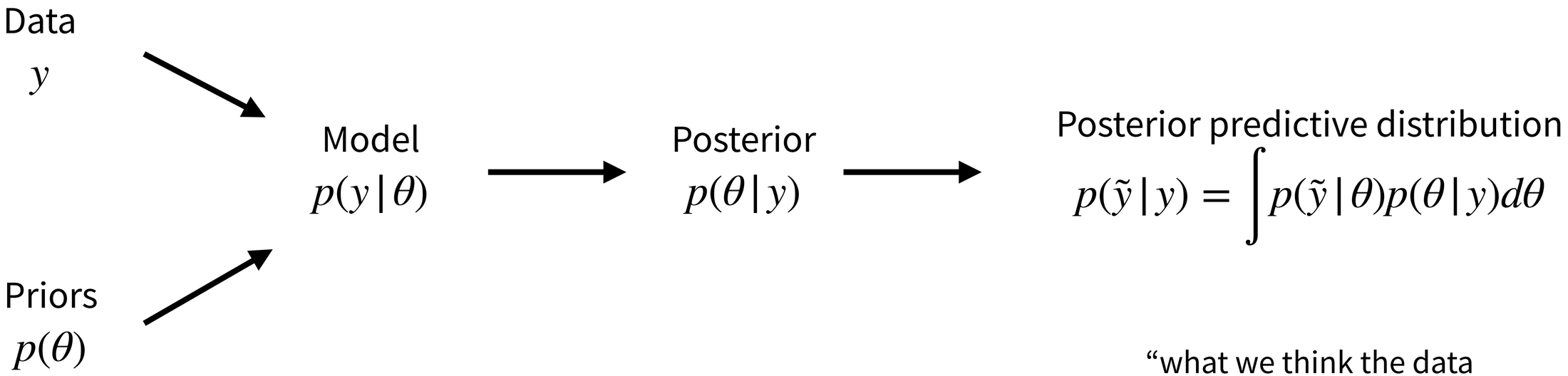
Bayesian inference

- treats parameters as random variables
- posterior distributions of parameters reflect information from prior and data

In practice

- use algorithms to approximate the posterior by drawing samples from the distribution
- various implementations available (e.g. Stan, JAGS) that make it easier to specify and fit models

Making posterior predictions



"what we think the data should look like given what we've seen"

Guiding criterion*: “good” models should produce plausible predictions

(* one of many possible criteria)

Q1

Can the model predict new observations?

Q2

Can the model replicate the observed distribution of outcomes?

Q3

Does the model make reasonable predictions for quantities of interest?

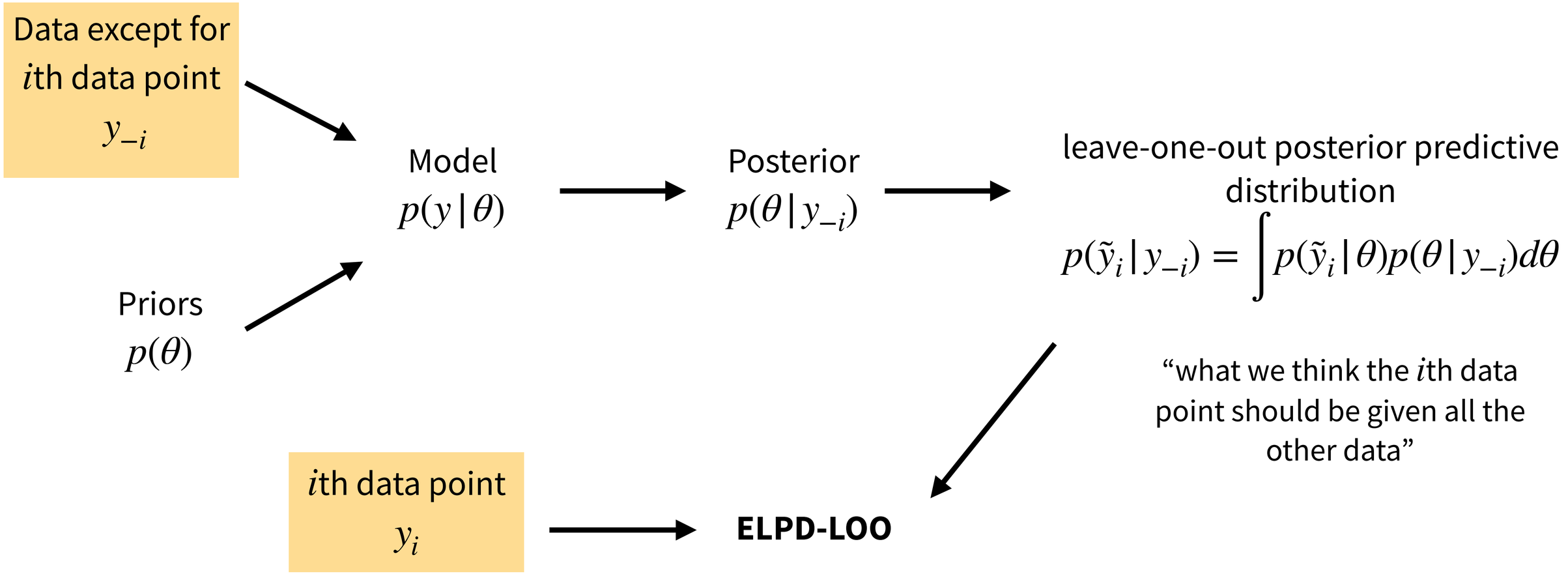
Q1: Can the model predict new observations?

Tool 1: leave-one-out expected log predictive density (or ELPD, ELPD-LOO)

- **what it is:** a numerical score of how well the model predicts left-out data
- score is based on the posterior probability that the model assigns to observed left-out values
  - “how likely does the model think the observed values are, given the other data?”
- usually approximated for computational reasons
- used to compare values (the raw value is difficult to interpret)

Q1: Can the model predict new observations?

Tool 1: leave-one-out expected log predictive density (or ELPD, ELPD-LOO)



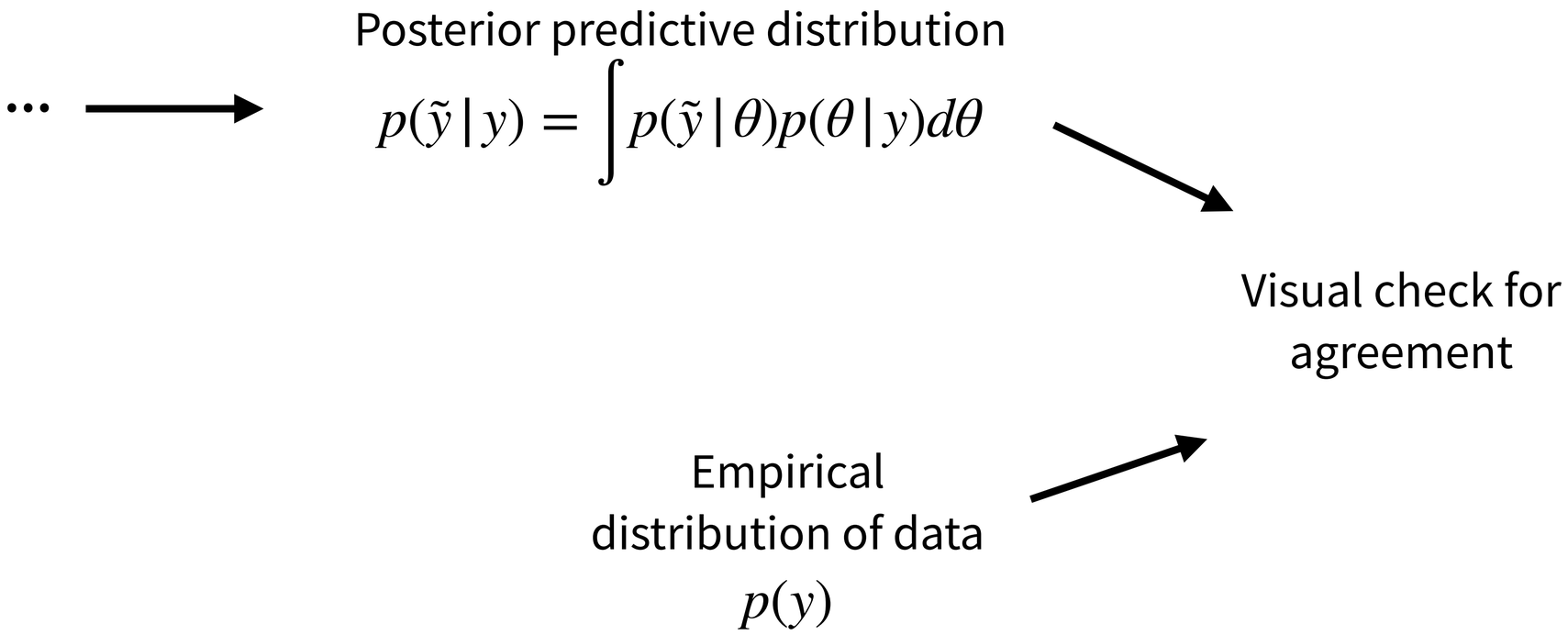
Q2: Can the model replicate the observed distribution of outcomes?

Tool 2: posterior predictive check of outcome

- **what it is:** set of curves that represent distributions of predicted outcomes
- used to check whether the shape of predicted values is reasonable
  - can identify issues like bias, over/underdispersion

Q2: Can the model replicate the observed distribution of outcomes?

Tool 2: posterior predictive check of outcome



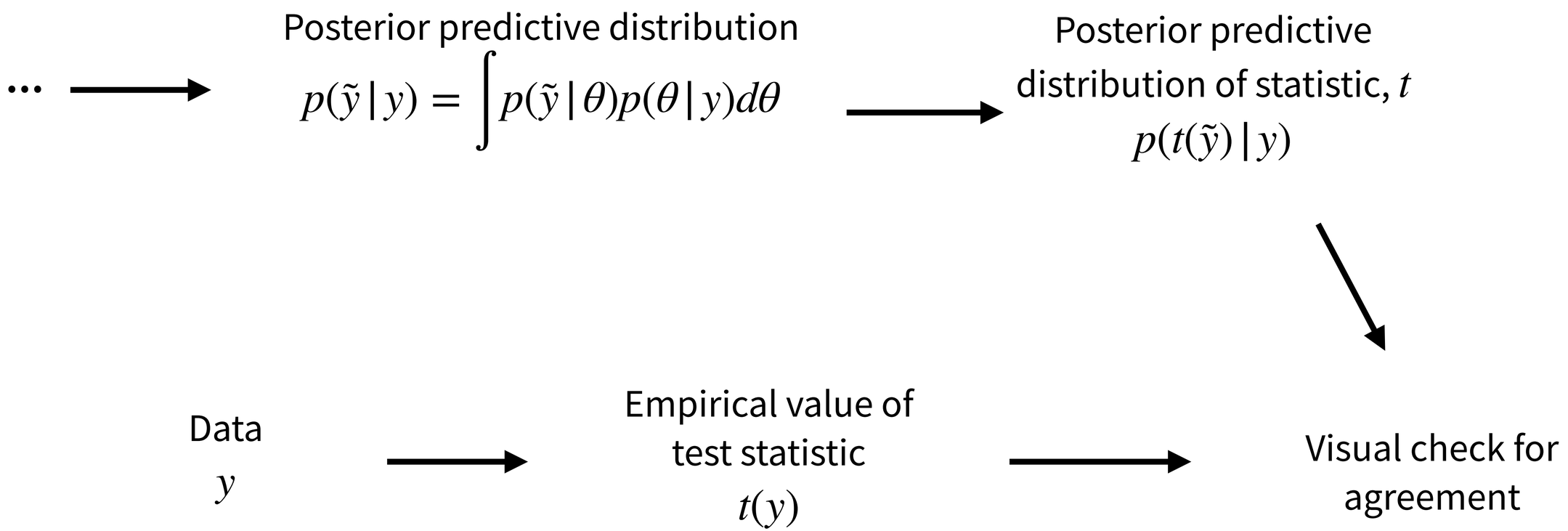
Q3: Does the model make reasonable predictions for quantities of interest?

Tool 3: posterior predictive check of test statistics

- **what it is:** histogram of predicted values for some test statistic
- can tell us if the model is good at predicting a particular *function of the data*
  - "are the predictions wrong in a consequential way?"

Q3: Does the model make reasonable predictions for quantities of interest?

Tool 3: posterior predictive check of test statistics



(Interactive examples in R)

https://github.com/michael-chong/mpidr-bayes-eval