Posterior Predictive Checking

May 7, 2025

Prof. Gwendolyn Eadie

Why posterior samples are so useful

- Model design
  - Draw samples from the prior distribution to get a sense of what the model expects before you use the data (helpful in multidimensional problems)
- Model checking
  - Posterior predictive checks!
- Software validation
  - Simulate data from known model
  - Fit data to model and make sure you get the expected result
- Research design
  - With simulated observations, you ca test whether research design will be effective
- Forecasting
  - Simulate new predictions
  - What modifications to model may be needed in the future
- Scientific inference
  - Propagate uncertainty in future calculations

Posterior predictive distribution

- We can check our model with predictive checks
- The *Posterior predictive distribution* is given by:

$$p(y_{future}|y_{observed}) = \int p(y_{future}|\theta)p(\theta|y_{observed})d\theta$$

- Basic idea:
  - Generate mock data using the posterior distribution of the parameters
  - Compare the mock data to the real data
  - If the model is a good description of the real data, then the model should be able to generate "real" looking data.

Let's go back to our m&m's example to see how this works!

Exercise

- Perform a *posterior predictive check*
  - Draw random θ value from your posterior
  - Given θ, draw random y from the binomial
  - Repeat many times, to get a predictive distribution of y's
  - Predictive distribution of y's $\rightarrow$ $p(y_{future}|y_{oberved})$
- Plot your predictive distribution of y's, compare to your neighbour's