6: Model Checking

10/09/19

Model checking and likelihood principle

Does the average score of 55 students on a test exceed 80? Assume that the test scores are normally distributed $\mathrm{N}(\mu, \sigma^2)$.

Imagine four ways of collecting the data:

- Random sample of 55 students
- Randomly sample students in sequence and after each student cease collecting data with probability 0.02
- Randomly sample students for a fixed amount of time
- Continue to randomly sample until the sample mean is significantly different from 80 using t-test

Among all four data sampling rules, the likelihood stays the same and thus $p(\theta \mid y)$. However, $p(y^{rep} \mid y)$ is not. Model checking depends on data-collection model.

Marginal posterior checks

A (marginal) posterior predictive p-value based on the marginal posterior predictive distribution $p(y_i \mid y)$.

$$p_i = \Pr(T(y_i^{rep}) \leq T(y_i) \mid y)$$

A natural test quantity is $T(y_i) = y_i$.

For the ETS example, y_i^{rep} is generated from $P(\tilde{y}_i \mid y)$, $i = 1, \ldots, 8$ or $P(\tilde{y}_i \mid y)$, $i = 9, \ldots, 16$ depending on predictions are made for existing schools or new schools.

In the first case, all separate p_i's tend to concentrate around 0.5 while look more uniform in the second case.

Marginal posterior checks

A cross validation predictive p-value:

$$p_i = \Pr(T(y_i^{rep}) \leq T(y_i) \mid y_{-i})$$

Cross validated p-values require more computation but are supposed to be more uniform.

What patterns can reveal

When we examine the collection of marginal p-values, p_i's,

- Marginal posterior p-values concentrate near 0 or 1, data is over-dispersed compared to the model
- Marginal posterior p-values concentrate near 0.5, data is under-dispersed compared to the model

Graphical posterior predictive checks

- Direct display all the data (e.g. the speed of light data)
- Display of data summaries (e.g. the speed of light data, validating the assumption in a Bernoulli experiment) or parameter inferences
- Graphs of residuals or other measures of discrepancy between model and data