Model checking

- 6.1 The place of model checking in applied Bayesian statistics
- 6.2 Do the inferences from the model make sense?
- 6.3 Posterior predictive checking
- 6.4 Graphical posterior predictive checks (can be skipped)
- 6.5 Model checking for the educational testing example

Model checking

- demo6_1.m: Posterior predictive checking - light speed
- demo6_2.m: Posterior predictive checking - sequential dependence
- demo6_3.m: Posterior predictive checking - poor test statistic
- demo6_4.m: Posterior predictive checking - marginal predictive p-value

Model checking – overview

- Sensibility with respect to additional information not used in modeling
  - e.g., if posterior would claim that hazardous chemical decreases probability of death
- External validation
  - compare predictions to completely new observations
  - cf. relativity theory predictions
- Internal validation
  - e.g. posterior predictive checking

Posterior predictive checking – example

- Newcomb's speed of light measurements
  - model $y \sim \mathrm{N}(\mu, \sigma^2)$
  - prior $(\mu, \log \sigma) \propto 1$
- demo6_1.m

Replicates vs. future observation

- Predictive $\tilde{y}$ is the next not yet observed possible observation. y^{rep} refers to replicating the whole experiment (with same values of x) and obtaining as many replicated observations as in the original data.

Posterior predictive checking

- Data y
- Parameters θ
- Replicated data y^{rep}
  - assume that the data has been generated by a process which can be well described by the model M with parameters θ
  - replicated data could be observed if the experiment were repeated
  - replace "true" data generating process by the model

$$p(y^{\text{rep}}|y, M) = \int p(y^{\text{rep}}|\theta, M) p(\theta|y, M) d\theta$$

- Test quantity (or discrepancy measure) $T(y, \theta)$
  - summary quantity used to compare the observed data and replicates from the predictive distribution

Posterior predictive checking

- Posterior predictive *p*-value

$$
\begin{aligned}
p &= \Pr(T(y^{\text{rep}}, \theta) \geq T(y, \theta) | y) \\
&= \int\int I_{T(y^{\text{rep}},\theta) \geq T(y,\theta)} p(y^{\text{rep}}|\theta) p(\theta|y) dy^{\text{rep}} d\theta
\end{aligned}
$$

  where I is an indicator function

  - having $(y^{\text{rep}\, l}, \theta^l)$ from the posterior predictive distribution, easy to compute

$$
T(y^{\text{rep}\, l}, \theta^l) \geq T(y, \theta^l), \quad l = 1, \ldots, L
$$

- Posterior predictive *p*-value (ppp-value) estimated whether difference between the model and data could arise by chance

Posterior predictive checking – example

- Independence in binomial experiment
  - model $y \sim \text{Bin}(\theta, 1))$
  - prior $\theta \sim \text{Beta}(1, 1)$
- Observations in order: 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0
- $T =$ the number switches in the series
  - observed $T(y) = 3$
  - if the observation were independent, what would be the distribution of the number of switches if the experiment were repeated?
- demo6_2.m

Posterior predictive checking – example

- How to choose test quantity
  - don't test properties which match the model parameters, because they are fitted to the data
  - test something, which is not (yet) part of the model
  - different test quantities may produce different results
- demo6_3.m

Calibration of ppp-values

- In the special case that the parameters θ are known (or estimated to a very high precision) or in which the test statistic $T(y)$ is ancillary (that is, if it depends only on observed data and if its distribution is independent of the parameters of the model) with a continuous distribution, the posterior predictive p-value $\Pr(T(y^{\mathrm{rep}}) > T(y)|y)$ has a distribution that is uniform if the model is true.
- Under these conditions, p-values less than 0.1 occur 10% of the time, p-values less than 0.05 occur 5% of the time, and so forth.

Marginal and CV predictive checking

- Consider marginal predictive distributions $p(\tilde{y}_i|y)$ and each observation separately
  - marginal posterior p-values

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

if $T(y_i) = y_i$

$$p_i = \Pr(y_i^{\text{rep}} \leq y_i|y)$$

- if $Pr(\tilde{y}_i|y)$ well calibrated, distribution of p_i would be uniform between 0 and 1
  - holds better for cross-validation predictive tests (cross-validation Ch 7)
- demo6_4.m

Calibration of marginal predictive checks

- For continuous data, cross-validation predictive p-values have a uniform distribution if the model is calibrated

Sensitivity analysis

- How much different choices in model structure and priors affect the results
  - test different models and priors
  - alternatively combine different models to one model
    - e.g. hierarchical model instead of separate and pooled
    - e.g. *t* distribution contains Gaussian as a special case
  - robust models are good for testing sensitivity to “outliers”
    - e.g. *t* instead of Gaussian
- Compare sensitivity of essential inference quantities
  - extreme quantiles are more sensitive than means and medians
  - extrapolation is more sensitive than interpolation