Model validation through "Posterior predictive checking" and "Leave-one-out"
Overview

The posterior predictive distribution from a fitted model
Check of fit between model and data
The “Leave-one-out” method for the 1-way ANOVA model
Example: ECG data
Summary
References
The posterior predictive distribution from a fitted model

Prior information $\pi(\theta)$

Data $x$

Posterior information $\rightarrow$ Prediction of new data

Information is represented by probability distributions on the parameter space $\Theta$
The posterior predictive distribution from a fitted model

\[ \pi(\vartheta \mid x) = \frac{\pi(\vartheta) \cdot l(\vartheta \mid x)}{\int_{\Theta} \pi(\vartheta) \cdot l(\vartheta \mid x) \, d\vartheta} \]

\[ p_\pi(\tilde{x} \mid x) = \int_{\Theta} p(\tilde{x} \mid \vartheta) \cdot \pi(\vartheta \mid x) \, d\vartheta \]

Probability model: \( p(x \mid \theta) \)

Posterior distribution

(norm. factor)

Predictive distribution for new data
Check of fit between model and data

Model selection – comparison of \(\leq 2\) models with each other

Model validation – consideration of 1 model and of its fit to the data, without reference to (an) alternative model(s)

We are now concerned with model validation only.
Data prediction as a means of model validation: Subdivide data into learning sample and validation sample, and compare the data of the validation sample with the values predicted from the learning sample (better: predicted from the posterior distribution derived from the learning sample).
Check of fit between model and data

(a) Learning sample and validation sample both of considerable size – difficult to investigate

(b) Learning sample empty – predict all data from the prior distribution ("prior predictive check")
Check of fit between model and data

(c) Validation sample empty – fit model to all data and re-check (“posterior predictive check”). Values predicted from $\pi(\theta|x)$ will the formally not be “new” data but only replicates of the observed data (all covariate values remain the same).

(d) Leave-one-out method – predict each data point $x_i$ from the posterior distribution derived from all others, $\pi(\theta|x_{-i})$. 
Check of fit between model and data

If $\pi(\theta|x)$ or $\pi(\theta|x_{-i})$ is determined by MCMC simulation, predicted values can be generated at each iteration and the distribution of these predicted values can be compared with the data point $x_i$ itself.

The aberrant position of $x_i$ relative to the distribution of the predicted values is described by the “predictive p-value” $P(x_{i^*} = x_i | x)$ or $P(x_{i^*} = x_i | x_{-i})$. 

Check of fit between model and data

Predictive p-values close to 0.5 show that the fit for that data point is good.

Calibration is a difficult problem: Which deviation from 0.5 should be considered as a relevant lack of fit? Predictive p-values are not $\mathbb{U}[0,1]$ distributed, see e.g. Hjort/Dahl/Steinbakk (2006). Remains open for artificial data (O’Hagan 2003, Sharples 1990). Therefore we turn to measured data (ECG data) where an external (medical) relevance assessment exists.

We investigate now methods (c) and (d). Method (c), based on $\pi(\theta|x)$, is simpler – but is it adequate?
The “Leave-one-out” method for the 1-way ANOVA model

\[ x_{ij} = \mu_i + \sigma \cdot \varepsilon_{ij} \quad \text{with } \varepsilon_{ij} \text{ i.i.d. } N(0,1) \]

and common \( \sigma \)

\[ \mu_i = \mu + \tau \cdot \varepsilon_i \quad \text{with } \varepsilon_i \text{ i.i.d. } N(0,1) \text{ and independent of the } \varepsilon_{ij} \]
The “Leave-one-out” method for the 1-way ANOVA model

Marshall/Spiegelhalter (2003) investigate analytically the balanced case with known $\sigma$ and $\tau$

The degree of overoptimism of the posterior predictive check depends from $I$, the number of groups, and decreases with increasing $I$

Also, they propose a “Leave-one-group-out” method
The “Leave-one-out” method for the 1-way ANOVA model

\[ x_{ij} = \mu_i + \sigma \cdot \varepsilon_{ij} \quad \text{with } \varepsilon_{ij} \text{ i.i.d. } N(0,1) \]
and common \( \sigma \)

\[ \mu_i = \mu + \tau \cdot \varepsilon_i \quad \text{with } \varepsilon_i \text{ i.i.d. } N(0,1) \text{ and independent of the } \varepsilon_{ij} \]

Prior distributions:
- \( \sigma \sim U(0,S) \text{ with } S \text{ large} \)
- \( \tau \sim U(0,T) \text{ with } T \text{ large} \)
- \( \mu \sim N(0,U) \text{ with } U \text{ large} \)
Example: ECG data

25 subjects, 3 repetitions

![Graph showing ECG data with QT interval values]
Example: ECG data

Potentially critical subjects for QTcF are those with a span of at least 12 ms between repetitions, and subject 20.

We investigate now subject 16 (repetition 1, value 409 ms) and 20 (all 3 repetitions).

See Camm (2006) for explanation of ECG intervals and correction methods.
Example: ECG data

Hierarchical random-effects model is fitted through MCMC (see Gilks/Richardson/Spiegelhalter 1996) and formulated in WinBUGS

Model
{
  for (l in 1:L) {
    y[l]~dnorm(mi[subj[l]],sigi);
  }
  for (k in 1:K) {
    mi[k]~dnorm(mu,taui);
  }

  y161~dnorm(mi[16],sigi);
  check161 <- step(409-y161);
  ... 
  mu~dnorm(390,1.0E-4);
  sigma~dunif(0,1000);
  sigi <- 1/(sigma*sigma);
  tau~dunif(0,1000);
  taui <- 1/(tau*tau);
}"
Example: ECG data

Hierarchical random-effects model is fitted through MCMC (see Gilks/Richardson/Spiegelhalter 1996) and formulated in WinBUGS

```r
# .../ecg161d.txt
# Value of subject 16,
# repetition 1 (409) left out
#
list(
  y=c(                                 subj=c(
    403,410,408,                         1,1,1,
    ...                                  ...
    393,400,                             16,16,
    ...                                  ...
    382,381,381),                        25,25,25),
  K=25, L=74)
```
Example: ECG data with model fitted to all data

25 subjects, 3 repetitions
Example: ECG data with model fitted to all data

![Example ECG data with model fitted to all data](image)
Example: ECG data with model fitted to all data

Repetition 1 of subject 16 (measured: 409 ms) is predicted as 400 +/- 5.7 ms

The predictive p-value for $x_{16,1}$ is 0.9427
Example: ECG data without data point in question

25 subjects, 3 repetitions
Example: ECG data without data point in question
Example: ECG data without data point in question

Repetition 1 of subject 16 (measured: 409 ms) is predicted as 396 +/- 5.7 ms

The predictive p-value for $x_{16,1}$ is 0.9896
Example: ECG data without data point in question

25 subjects, 3 repetitions
Example: ECG data without data point in question

Omission of any single data point of subject 20 does not change the situation as the 3 points “mask” each other
Example: ECG data without data point in question

Medically, also the data point with a “leave-one-out” predictive p-value of 0.9896 is explainable by biological variation in healthy volunteers – short peak of HR without adaptation of QT (“QT hysteresis”) is a plausible explanation

Think of an extended model with individual $\sigma_i$s
Summary

“Leave-one-out” is sensitive to masking (not new)

Comparing the data with the values predicted from the model fitted to all data (including the data point in question) is overoptimistic, predictive p-values are pulled towards 0.5
Summary

Calibration remains a problem for each model validation procedure
References


