Model evaluation in nonlinear mixed effect models
Emmanuelle Comets 1, *, Karl Brendel 2 and France Mentré 1

1 INSERM UMR738, Paris, France; Université Paris Diderot, Paris, France 2 IRIS, Servier, Courbevoie, France

Objective: Model evaluation is an important part of model building, and has been the subject of regulatory guidelines. We illustrate the use of some recently proposed metrics on several simulated datasets.

Introduction

Several simulation-based metrics developed over the last decade:
- Visual Predictive Checks (VPC) [1]:
  - prediction discrepancies (pd) [2]
  - normalised prediction distribution errors (npde) [3]

Assumptions:
- model M is being built using a building dataset B
- null hypothesis: this model can be used to describe the data collected in a validation dataset V (in internal evaluation)
- General class of Posterior Predictive Check (PPC), born in the Bayesian world
- model M is used to simulate data according to the design of V
- compare a statistic computed on the real data in V to the distribution of the statistic obtained through the simulations
  - here plug-in approach (ignoring uncertainty)

Model and data

Statistical models
Model for observation \( y_{ij} \):
\[
N Holford. The Visual Predictive Check: superiority to standard diagnostic (Rorschach) plots.

Results

Table 1: parameters estimated in original dataset

![Table 1](image)

![Figure 2: 95% VPC with prediction bands, for datasets Vtrue (upper left), Vbiav (upper right), Vnpde (lower left), Vnpde2 (lower right).](image)

![Figure 3: Plot of pd versus time with prediction bands, for datasets Vtrue (upper left), Vbiav (upper right), Vnpde (lower left), Vnpde2 (lower right).](image)

![Figure 1: VPC plots for Vtrue, with several representations. Top: 2.5 and 97.5% percentiles of the simulated data; thick dashed lines: 50th percentile; dotted observations. Bottom: 95% prediction intervals around 2.5, 50, and 97.5% percentiles (coloured areas); dotted/dashed lines: 2.5, 50, and 97.5% percentiles of observed data (thick line: median). Figures 2 and 3 show plots of VPC and pd versus time with prediction bands for the 4 simulated datasets.](image)

Conclusions

- Array of complementary tools to be used by modellers:
  - pd and VPC allow to visualise patterns with time
  - npde and PLNPC provide a test
- Simulation-based metrics
  - require simulations under the model, which can be difficult to obtain, eg in the presence of drop-outs or censored data [5]
- Prediction bands obtained through repeated simulations
  - computer-intensive: final models only
  - enhance the detection model misspecifications by providing clear visual comparison of model expected behaviour versus observed data
- Tests
  - only npde provide adequate type I error thanks to decorrelation [4]
- in real data, tests may be sensitive to large datasets or outliers
- global tests: may be difficult to pinpoint exactly which aspects of the model to change
  - best used as a signal to guide further model improvement

References