Stellingen

behorende bij het proefschrift

Prediction Sets via Parametric and Nonparametric Bayes: with Applications in Pharmaceutical Industry

van Xiangyi (Gregory) Chen

1. Applying a suitable experimental design and statistical analysis helps to create high-performance manufacturing processes and laboratory methods, to provide rigorous evidence to the regulators during the validation phase, to monitor performance across different production sites throughout the commercial phase, and to shed light on root causes of emerging trouble.

2. If the past data is sufficiently informative, estimates of unknown parameters can simply be plugged in, and estimation uncertainty ignored. More realistically both sources of randomness should be taken into account, and it is desirable to make precise how the desired coverage $1 - \delta$ should be achieved.

3. An advantage of Bayesian tolerance set is that while their form is determined by the future variable Z, through the prediction problem Q_θ , the model for the data X enters only through the posterior distribution $\Pi(\theta \in \cdot | X)$. If in the former the dependence on the parameter θ is not too complicated, then the problem is solvable for even complicated data models.

4. The depth paradigm provides a way to form prediction sets without particular shape assumptions. It is particularly useful in the multivariate case, since ranking of more than one-dimensional data points is not straightforward otherwise.

5. A default prior need not be non-informative, but should be spread over the whole parameter space.

6. It turns out to be an ambitious task to assure low stability risk by controlling just a sample measurement at time 0. Sometimes the problem is how we formulate the problem.

7. Before investment, usually a good sanity check for an image classification problem is to see if human experts can actually accomplish the task with reasonable confidence or not.

8. In practice, setting a co-variate to be a random effect in a mixed model actually enforces a very strong restriction. Even assuming a flexible distribution (e.g. Gaussian mixture) for that random effect is not enough to alleviate the ill-posed model. This aspect is usually overlooked due to the popularity of the mixed model in the pharmaceutical industry.

9. Automation brings efficiency and harmonization, but should not encourage the beneficiaries to quit thinking. A data-driven or "predictive" strategy usually means taking a smart risk and making an informed decision in a dynamic environment, but uncertainty cannot be removed.

10. The significant problems we face cannot be solved at the same level of thinking we were at when we created them. This seems particularly relevant for today's pharmaceutical industry, when it is playing catch-up with data-science-driven innovation.