Model averaging

When several models are under consideration, the standard approach is to select a single model based on some principle. Classical principles are likelihood testing (which is the same as backwards stepwise); AIC or alternatively forward stepwise. However, there may be some value in using several models in parallel and averaging the results (like the estimates of regression coefficients). Weights may depend on fit or be chosen beforehand. This may give a better prediction. **Example 1:** Regression with several potential covariates. Suppose we want to predict the body weight of a child and have information on sex, age and height. Using all three covariates should in theory give the best prediction, but could have extra uncertainty due to the high correlation between age and height. So one could fit three models; sex and age; sex and height; sex, age and height and then average the three predictions. **Example 2:** Dose finding with alternative dose-response models. In this case, we may have alternative models, say as means f(x) at dose x, the model could be of the Emax type (f(x)=α + β x /(x+γ)) or the exponential type (f(x)= α + β (1- exp(-x γ)). Suppose conventional fit comparisons do not give a clear conclusion as to which is best. Averaging the two predictions could be better than either one. This research area has been pioneered by Claeskens and Hjort.

The project:

The project consists of studying the statistical properties of the predictions. This includes how to weight two or more models to give the most meaningful prediction. A second question is to which extent this approach can be used for testing (that is, whether it satisfies the intended significance level).

Possible project elements:
* Working with data from H. Lundbeck
* Reading about model averaging and model selection
* Implementing methods in R
* Comparing different approaches on real data
* Comparing different approaches on simulated data to study the power

A few references
Claeskens, G. (2016)**"Statistical model choice".** *Annual Review of Statistics and Its Application*, 3, to appear.

Homepage of Gerda Claeskens: https://feb.kuleuven.be/public/ndbaf45/publicationsGC.html

Contact: Elisabeth Wreford Andersen (ewan@dtu.dk)

In collaboration with: Philip Hougaard, Vice President at H. Lundbeck A/S