Projects in course TMA4315 – 2013

Håvard Rue and Erik Solbu

October 22, 2013

Choose one of the projects below, and inform either Håvard Rue or Erik Solbu about your choice no later than 24th Oct 2013. The report is due 21st Nov, but the presentations will take place the 19th and the 21st Nov. You can work in groups no larger than size two. You are required to have one meeting with either HR or ES in each of the weeks: 44, 45 and 46, even though you have "nothing" to discuss.

Project 1: Write your own `glm`-package

The goal is to use what you have learnt so far in this course, to write your `my.glm`-package in `R` (as an `R`-package). It should accept Gaussian, binomial and Poisson data, and follow the formula-construct in `glm` (and give similar results as `glm`); have methods for plotting, etc. You should aim for a high enough standard that you could have used your own package to do all the previous exercises in this course. The `R-package` should be written so it pass the `R-package-check`: `R CMD check <my.glm>`

The critical part which parse the formula, can be done as follows

```
> my.glm
function(formula,  data, ...)
{
    stopifnot(is.data.frame(data))
    X = model.matrix(formula, data)

    return(X)
}
```

The "X" is the model-matrix, so that $\eta = X\beta$; look at the column-names and you see...

The report should contain a description of the background for your choices, and a demonstration of how to use the package and a comparison with the `glm`-package in `R`. (You do not have to write a `glm`-clone!)

Project 2: Models for splines and smooth curves

An easy entry to this field is the report by Eilers and Marx[1] Based on this article (and/or others), give an overview of the ideas, how to do the estimation and (possible) applications. You should implement the examples in the report by yourself, in `R`.

[1]Flexible smoothing with B-splines and penalties, Paul H. C. Eilers and Brian D. Marx, Source: Statist. Sci. Volume 11, Number 2 (1996), 89-121.

Project 3: Bayesian `glm`'s and `INLA`

This project should study the Bayesian approach to formulate a GLM-model, how to set priors and do the inference. A particular case, is when the response is Gaussian for which (with conjugate priors), we can derive analytical expression for the posterior distribution[2]. The computational issue should also be discussed for non-Gaussian response, either by concentrating on MCMC-methods or try to understand the main idea of `INLA` (see `www.r-inla.org`).

Project 4: Data-analysis of football/handball/icehockey matches

This is a data-analysis project, working with football/handball/icehockey data, with the aim of producing one-step-ahead predictions[3]. Static model (no dependence in time) and dynamic models should be discussed, and the results for last-half-of-a-season predictions should be computed and discussed.

Project 5: Cox-proportional-hazard models

This is a similar project as "Project 1" with the same general aims and requirements, but only for Cox-proportional-hazard models using the Poisson augmentation, discussed in Section 7 in the lecture notes. It is sufficient to consider the case, with at least one "event" within each bin, and strategies to introduce dependency/smoothness in the baseline hazard should be discussed.[4]

[2]See f.ex the book "Bayesian core" by Marin and Robert

[3]For a background, see f.ex Rue, H. and Salvesen, O. (2000), Prediction and Retrospective Analysis of Soccer Matches in a League. Journal of the Royal Statistical Society: Series D (The Statistician), 49: 399418. doi: 10.1111/1467-9884.0024, and the follow-up PhD-thesis *Statistical Models in Ice Hockey* by Ø. Salvesen `http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-15805`

[4]See f.ex MARTINO, S., AKERKAR, R. and RUE, H. (2011), Approximate Bayesian Inference for Survival Models. Scandinavian Journal of Statistics, 38: 514-528. doi: 10.1111/j.1467-9469.2010.00715.x