Practical 8. Survival analysis

Wednesday, 22 June 2022

Lecture 6 PDF version

The file `survival_data.Rdata` contains a dataset from a randomised trial of a drug which is assumed to target a particular cancer. Data are observed for $n = 367$ patients randomised to either the control ($n_1 = 189$) or the active treatment ($n_2 = 178$) and are stored in the data frame `dat`.

The data report the patients ID; the time of progression to a more severe stage of cancer; an indicator for the event of interest (mortality); an indicator for the treatment arm (coded as 0 = control and 1 = active treatment); an indicator for the patients' sex (0 = male; 1 = female); the patients' age (in years); and the Index of Multiple Deprivation (IMD) score (this is a census-based, area-level measure of socio-economic circumstances. It is coded as categorical variable taking values in the interval $[1; 5]$, where 1 indicates the least deprived and 5 indicates the most deprived areas).

1. Follow the `R` script `survival.R` to fit the Exponential and the Weibull parametric survival models under a simple MLE approach, using the following specification for the location parameter.

$$g(\mu_i) = \log(\mu_i) = \beta_0 + \beta_1 \text{arm}_i.$$

   Make sure you understand `R` and `survHE` notation to define the "`formula`" specifying the model above.
2. Follow the script to explore the output (stored in the object `m1`).
3. Based on the output provided by `R` , including the graphical representation of the survival curves, what is your preferred model to described the underlying original data? Discuss how you would justify your answer.
4. Follow the script to fit the same model under a Bayesian approach and using both INLA and HMC as the inferential engine.
5. Compare the output from the models specified using MLE and those obtained under a Bayesian modelling.
6. Follow the script to produce a combine plot with all the survival curves. Comment on similarities and differences in the models output.
7. Follow the script to perform PSA and produce suitable graphs for the survival curves distributions.