Exercise 1: Additive Regression

Exercise 1 (Childhood Malnutrition in Zambia)

Childhood malnutrition is one of the most challenging problems in developing and transition countries. The data set `ZambiaNutrition.dat` contains information on the nutritional status of children (variable `stunting`, measuring chronic malnutrition, i.e. stunting) along with a set of explanatory variables

Variable	**Explanation**
mbmi	body mass index of the mother
agechild	age of the child
district	district where the mother lives
memployment	employment status of the mother
meducation	education level of the mother
gender	gender of the child

The corresponding spatial information is given in the boundary file `ZambiaBnd.bnd`.

1. Load the data and visually explore the relation of the available covariates with `stunting`.
2. Estimate a additive regression model

$$\text{stunting}_i \sim \mathcal{N}(\beta_0 + f_1(\text{mbmi}_i) + f_2(\text{agechild}_i), \sigma^2) \tag{1}$$

   for stunting with nonlinear effects of `mbmi` and `agechild` and plot the estimated effects. To use P-splines in the specification, use the function `s(x,bs="ps")`.
3. Investigate the influence of the dimension of the basis, the degree of the splines and the difference penalty, specified by the options `k` and `m`.
4. Investigate if there is a difference in the nonlinear effect of `agechild` between boys and girls by specifying a varying coefficient model.
5. Extend the regression model (1) to also include a spatial effect with a Markov random field prior of `district` and an unstructured random spatial effect using `s(dist,bs="re")`, as well as categorical effects for all remaining covariables.
6. Plot the two spatial effects and decide whether they are significant by testing if the 95% credible intervals contain zero.

Exercise 2 (Forest Health Data)

The data set `foresthealth.dat` contains the following information on the forest health status of beeches at 83 observation plots in a northern Bavarian forest district collected in yearly visual forest health inventories between 1983 and 2004:

Variable	Explanation
year	calendar time in years
x, y	locations of the observation plots
id	id for the locations
inclination	inclination of slope in percent
elevation	elevation above sea level in meters
soil	depth of soil layer in centimeters
fertilisation	binary indicator on application of fertilization
age	average age of the stand in years
canopy	density of forest canopy in percent
stand	type of stand (1 deciduous forest, -1 mixed forest)
def	binary indicator for defoliation

1. Estimate a logistic additive regression model with response `def` and nonlinear effects for `age` and `canopy`.
2. Change the link function to estimate a probit model instead.
3. Add a spatial effect based on `x`,`y` and a random effect based on `id`.
4. Can you improve the model further?