Day 3: Bayesian Distributional Regression

Exercise 1 (Motorcyle Data)

- Load the data set `mcycle` from the `R` package `MASS` and make yourself familiar with the data set using the corresponding help page given in the package (type `?mcycle`). Explore the relationship between the variable `accel` and `times` in a scatterplot.
- Now estimate the two following models:
  - A Gaussian mean regression model with `accel` as dependent variable and `times` as continuous covariate with nonlinear effect.
  - A location-scale model where you also model the log-variance of the error term as a nonlinear function of `times`.
- Compare the two models based on the DIC and visualize the estimated effects.

Exercise 2 (Munich Rent Index)

Many cities and communities in Germany establish rent indices in order to provide the tenant and the landlord with a market review for the typical rent for the area and for given characteristics of the flat. The dataset `rent.dat` provides information collected for the rental guide in Munich with the following variables.

Variable	Explanation
`rent`	net rent per month (in Euro)
`rentsqm`	net rent per month per square meter (in Euro)
`area`	living area in square meters
`yearc`	year of construction
`district`	district the flat is located in
`location`	quality of location according to an expert assessment 1 = average location 2 = good location 3 = top location

Often, it is assumed that (conditionally on covariates) the log-rents are normally distributed with homoscedastic errors. In this exercise, we investigate whether other distributional assumptions work better. To compare the different models, the dataset `rent_test.dat` contains hold-out data for out of sample comparison.

- Estimate a mean Gaussian mean regression with nonlinear effects of `area` and `yearc`, linear effects of `location` and a random effect for `district` as a baseline and evaluate the hold-out log-likelihood.
- Prediction competition: Now try out different distributional regression specifications and evaluate them also via the log-likelihood on the hold-out data. The best out of sample log-likelihood wins the competition.