Distributional Modelling in R

06 - Bayesian Distributional Regression - Exercises

In this example we estimate a Bayesian distributional regression model using data on fatalities in Austria. The data is taken from the Eurostat data base (https://ec.europa.eu/eurostat/) and includes weekly fatalities in Austria from the beginning of 2000 up to week 46 in 2020. The data is provided in the **bamlss** package and can be loaded with

```
R> data("fatalities", package = "bamlss")
```

1. Subset the data set before 2020 for estimation and use the 2020 data for comparison.
2. Determine an appropriate distribution for modeling the number of fatalities using the **bamlss** package. Explore the distribution families available in the **gamlss.dist** package and compare their fit using the Deviance Information Criterion (function `DIC()`). Identify the distribution that provides the best fit to the data.
3. Subsequently, introduce a seasonal time trend based on the weeks of the year into the modeling process for the top three candidate distributions. Evaluate whether the previously selected distributional model remains the most suitable choice in light of this additional seasonal component. Analyze the resulting patterns that emerge from this addition.
4. Examine the sampling behavior of the Markov Chain Monte Carlo (MCMC) algorithm using trace plots. If required, consider adjustments such as increasing the number of iterations, extending the burn-in phase, and modifying the thinning parameter to improve convergence and sampling efficiency.
5. Additionally, evaluate the calibration of the model by examining randomized quantile residuals.
6. Calculate the 5th, 50th, and 95th percentiles for the seasonal effect, and overlay the number of fatalities observed in 2020 onto the plot. When computing the percentiles, ensure to utilize all MCMC samples within the predict method. Compute the percentiles by calculating the mean of the samples, thereby ensuring comprehensive coverage across the distribution. Analyze the resulting visualization and provide interpretation based on the observed trends and patterns.
7. Likewise, compute the seasonal probabilities of observing more than 1600 fatalities and illustrate the findings through visualization.