A Conceptional Lego Toolbox for Bayesian Distributional Regression Models

Nikolaus Umlauf (Nikolaus.Umlauf@uibk.ac.at), Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Achim Zeileis

1. INTRODUCTION

Bayesian Models:
- Convenient setting for very complex models.
- MCMC simulation now possible on virtually any computer.
- Inference does not rely on asymptotics.

Existing Implementations:
- Provide infrastructures for a number of regression problems.
- From univariate to multivariate distributions.
- Highly specialized and optimized engines.
- However, almost any engine has a different interface.

Basic Ideas:
- Provide a flexible and unified modeling architecture.
- Use specialized/optimized engines to apply Bayesian structured additive distributional regression a.k.a. Bayesian additive models for location scale and shape (BAMLSS) and beyond.
- Facilitate new algorithms and extensions.
- The approach should have maximum flexibility/extendability, also concerning functional types.

2. MODEL STRUCTURE

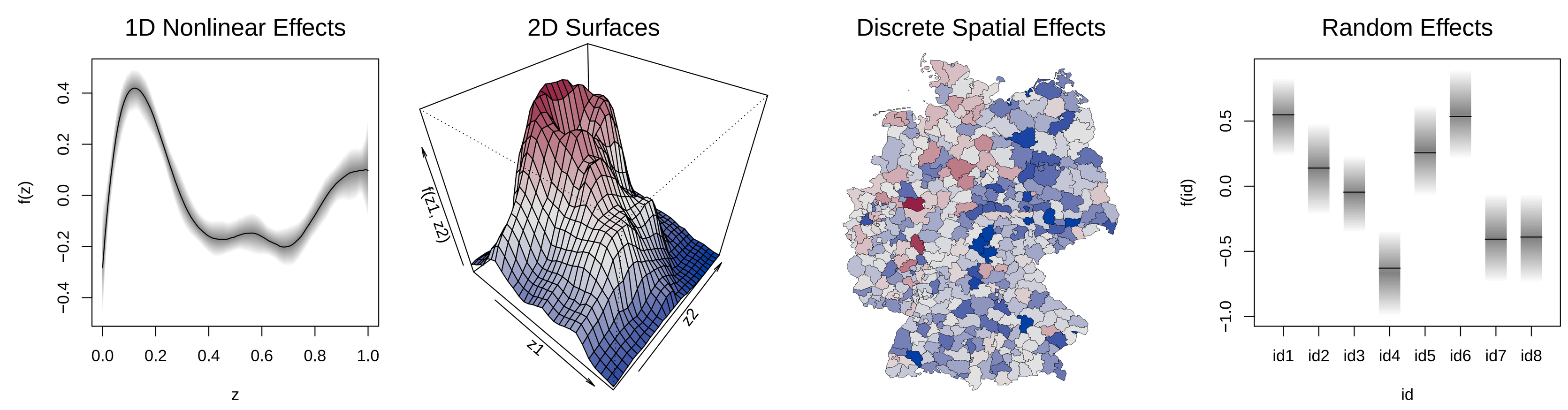
Within the GAMLSS model class all parameters of the response distribution can be modeled by explanatory variables

$$y \sim \mathcal{D}\left(h_1(\theta_1) = \eta_1,\ h_2(\theta_2) = \eta_2, \dots,\ h_K(\theta_K) = \eta_K\right),$$

where $\mathcal{D}$ denotes any distribution available for the response variable y and θ_k, $k = 1, \dots, K$, are parameters that are linked to additive predictors. The k-th additive predictor is given by

$$\eta_k = \eta_k(\mathbf{x}; \boldsymbol{\beta}_k) = f_{1k}(\mathbf{x}; \boldsymbol{\beta}_{1k}) + \ldots + f_{J_k k}(\mathbf{x}; \boldsymbol{\beta}_{J_k k}),$$

with unspecified (possibly nonlinear) functions $f_{jk}(\cdot)$ of a generic covariate vector $\mathbf{x}$, $j = 1, \dots, J_k$ and $k = 1, \dots, K$. Examples of functions $f_{jk}(\cdot)$:



3. LEGO TOOLBOX

Terms:
Each vector of function evaluations is a composition of

$$\mathbf{f}_{jk} = f_{jk}(\mathbf{X}_{jk}, \boldsymbol{\beta}_{jk}),$$

$\mathbf{X}_{jk}$ ($n \times m_{jk}$) is a design matrix, $\boldsymbol{\beta}_{jk}$ ($q_{jk} \times 1$) are regression coefficients. Note: functions are not necessarily linear in the parameters, e.g., growth curves.

Priors:
Generic prior for linear and nonlinear effects using a basis function approach

$$p(\boldsymbol{\beta}_{jk}) \propto \left(\frac{1}{\tau_{jk}^2}\right)^{rk(\mathbf{K}_{jk})/2} \exp\left(-\frac{1}{2\tau_{jk}^2}\boldsymbol{\beta}_{jk}^\top \mathbf{K}_{jk} \boldsymbol{\beta}_{jk}\right).$$

Precision matrix $\mathbf{K}_{jk}$ corresponds to frequentist's penalty matrix, τ_{jk}^2 is equivalent to the inverse smoothing parameter, common prior $p(\tau_{jk}^2) \sim IG(a_{jk}, b_{jk})$.

Response Distribution:
Main building block pdf $f(\mathbf{y}|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ and corresponding log-likelihood

$$\ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^{n} \log\ f(y_i; \theta_{i1} = h_1^{-1}(\eta_{i1}(\mathbf{x}_i, \boldsymbol{\beta}_1)), \dots, \theta_{iK} = h_K^{-1}(\eta_{iK}(\mathbf{x}_i, \boldsymbol{\beta}_K))).$$

Log-posterior

$$\log\ p(\boldsymbol{\vartheta}; \mathbf{y}, \mathbf{X}) \propto \ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) + \sum_{k=1}^{K}\sum_{j=1}^{J_k} \left\{\log\ p_{jk}(\boldsymbol{\vartheta}_{jk})\right\},$$

e.g., $\boldsymbol{\vartheta}_{jk} = (\boldsymbol{\beta}_{jk}^\top, (\boldsymbol{\tau}_{jk}^2)^\top)^\top$ and $p_{jk}(\cdot)$ denotes combination of all priors.

Posterior Mode/Mean Estimation:
Various algorithms require

- $\frac{\partial \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\beta}_k} = \frac{\partial \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\eta}_k}\frac{\partial \boldsymbol{\eta}_k}{\partial \boldsymbol{\beta}_k} = \frac{\partial \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\theta}_k}\frac{\partial \boldsymbol{\theta}_k}{\partial \boldsymbol{\eta}_k}\frac{\partial \boldsymbol{\eta}_k}{\partial \boldsymbol{\beta}_k}$,
- $\frac{\partial^2 \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_s^\top} = \left(\frac{\partial \boldsymbol{\eta}_s}{\partial \boldsymbol{\beta}_s}\right)^\top \frac{\partial^2 \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_s^\top}\frac{\partial \boldsymbol{\eta}_k}{\partial \boldsymbol{\beta}_k} + \frac{\partial \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\eta}_k}\frac{\partial^2 \boldsymbol{\eta}_k}{\partial^2 \boldsymbol{\beta}_k}$ if $k = s$,
- $\frac{\partial^2 \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_s^\top} = \frac{\partial \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\theta}_k}\frac{\partial^2 \boldsymbol{\theta}_k}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_s^\top} + \frac{\partial^2 \ell(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}_s^\top}\frac{\partial \boldsymbol{\theta}_k}{\partial \boldsymbol{\eta}_k}\frac{\partial \boldsymbol{\theta}_s}{\partial \boldsymbol{\eta}_s}$.

Hence, implementing new distributions usually requires derivatives for $\boldsymbol{\eta}_k$, only.
Generic (blockwise) iterative updating scheme

$$\boldsymbol{\beta}_k^{(t+1)} = U_k(\boldsymbol{\beta}_k^{(t)}|\cdot),$$

e.g., for backfitting or MCMC.

4. IMPLEMENTATION

An implementation is provided in the R package **bamlss** available at

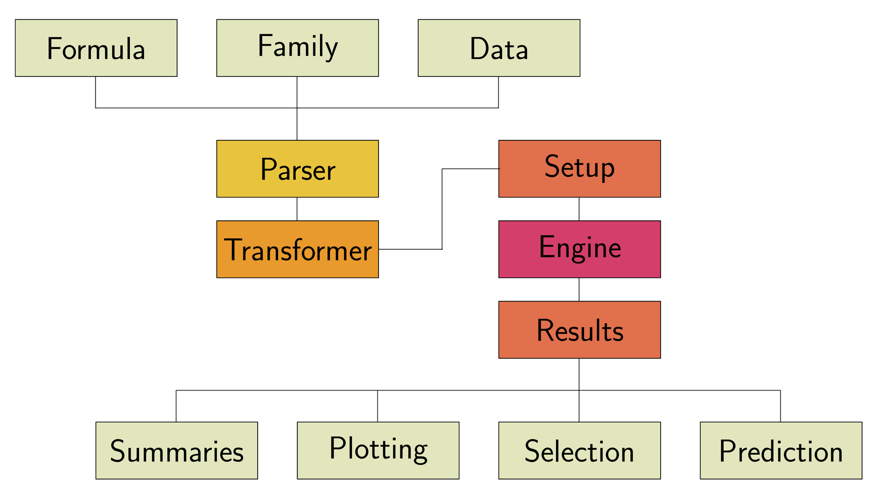
`https://R-Forge.R-project.org/projects/BayesR/`

In R, simply type

```
> install.packages("bamlss", repos = "http://R-Forge.R-project.org")
```

Generic architecture, the setup does not restrict to any specific type of engine (Bayesian or frequentist).

Various algorithms implemented, in addition support for **BayesX**, **JAGS**, **Stan**.



5. EXAMPLE

Modeling daily precipitation data with a censored normal model

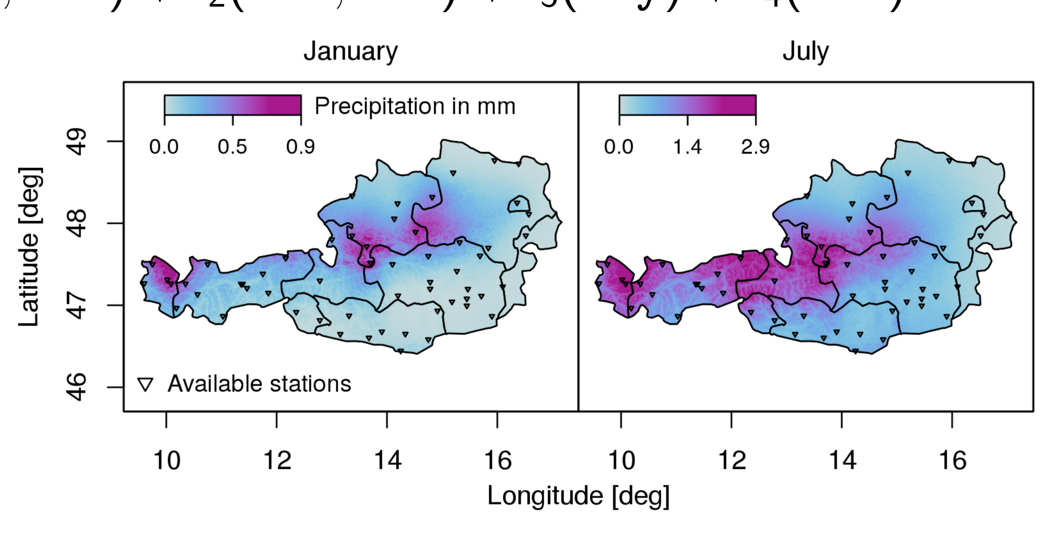
$$\mathbf{y}^\star \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2), \quad \boldsymbol{\mu} = \boldsymbol{\eta}_\mu, \quad \log(\boldsymbol{\sigma}) = \boldsymbol{\eta}_\sigma, \quad \mathbf{y} = max(\mathbf{0}, \mathbf{y}^\star).$$

For both $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, we use the following additive predictor:

$$\boldsymbol{\eta} = \beta_0 + f_1(\texttt{day}, \texttt{lon}, \texttt{lat}) + f_2(\texttt{lon}, \texttt{lat}) + f_3(\texttt{day}) + f_4(\texttt{alt}).$$

Analysis based on the HOMSTART data of the ZAMG.

Predicted precipitation for 10th of January/July.



References:

Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013). *Regression - Models, Methods and Applications*. Springer-Verlag, Berlin. ISBN 978-3-642-34332-2.

Rigby, R.A. and Stasinopoulos, D.M. (2005) Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society C*, 54(3), 507–554.

Umlauf, N., Mayr, G., Messner, J., and Zeileis, A. (2012). Why does it always rain on me? A spatio-temporal analysis of precipitation in Austria. *Austrian Journal of Statistics*, 41(1), 81–92.