GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

bamlss.vis

An R Package to Interactively Analyze and Visualize Bayesian Additive Models for Location, Scale and Shape (bamlss) Using the Shiny Framework

Stanislaus Stadlmann

7. December 2017

Georg-August University of Göttingen

Table of contents

1. Introduction

2. Motivating BAMLSS

Introduction

Introduction

Distributional Regression

- An emerging field in regression methods
- Each parameter of a response distribution beyond the mean can be modeled using a set of predictors

Distributional Regression

- An emerging field in regression methods
- Each parameter of a response distribution beyond the mean can be modeled using a set of predictors
- Notable frameworks:
    1. Generalized Additive Models for Location, Scale and Shape, coined by Rigby and Stasinopoulos (2001)
    2. Bayesian Additive Models for Location, Scale and Shape, coined by Umlauf, Klein, and Zeileis (2017)
- Differences: Estimation techniques

Introduction

Response

Let $y \sim D(\theta_1, \ldots, \theta_K)$

Problem

- Often, distribution parameters $h(\eta_l) = \theta_l$ do not directly equate to $E(y)$, $Var(y)$

Introduction

Response

Let $y \sim D(\theta_1, \ldots, \theta_K)$

Problem

- Often, distribution parameters $h(\eta_l) = \theta_l$ do not directly equate to $E(y)$, $Var(y)$
- Therefore hard to interpret effects on distribution moments because:
  1. Link function $h_l(\cdot)$ transforms effects
  2. Transformed effects are for parameters θ_l which often do not directly equate moments

Problem

- Thus: Package needed which
  1. Makes it easy to graphically display and compare predicted distributions
  2. Displays the influence of a covariate on the distributional moments
- $\Rightarrow$ `bamlss.vis` was born, solving these problems in R with a Shiny App.

Motivating BAMLSS

Additive Models (AM)

Overview

- Proposed by Friedman and Stuetzle (1981)
- Dependent variable y is related to non-parametric effects in an additive way

Additive Models (AM)

Overview

- Proposed by Friedman and Stuetzle (1981)
- Dependent variable y is related to non-parametric effects in an additive way

Model specification

$$y_i = f_1(z_{i1}) + f_2(z_{i2}) + \ldots + f_k(z_{ik}) + \epsilon_i \quad \text{(only nonparametric effects)}$$

$$y_i = \sum_{j=1}^{K} f_j(z_{ij}) + \sum_{l=1}^{Q} \beta_l x_{il} + \epsilon_i \quad \text{(with parametric effects)}$$

Additive Models (AM)

Overview

- Proposed by Friedman and Stuetzle (1981)
- Dependent variable y is related to non-parametric effects in an additive way

Model specification

$$y_i = f_1(z_{i1}) + f_2(z_{i2}) + \ldots + f_k(z_{ik}) + \epsilon_i \quad \text{(only nonparametric effects)}$$

$$y_i = \sum_{j=1}^{K} f_j(z_{ij}) + \sum_{l=1}^{Q} \beta_l x_{il} + \epsilon_i \quad \text{(with parametric effects)}$$

Why additive?

- Curse of dimensionality
- Easier to separate covariate effects

Structured Additive Regression (STAR) Models

Motivation

- AM allow for non-parametric effects, but sometimes even more flexibility is needed
- STAR (Fahrmeir et al., 2003) also support structured terms, which include:

Structured Additive Regression (STAR) Models

Motivation

- AM allow for non-parametric effects, but sometimes even more flexibility is needed
- STAR (Fahrmeir et al., 2003) also support structured terms, which include:
  1. Nonlinear effects of a single variable
  2. Spatial effects of location index s
  3. Interactions between a continuous covariate and a categorical variable
  4. Nonlinear interactions between two continuous covariates
  5. Random Effects with intercept ν_0 and slope ν_j deviations from main effects

Structured Additive Regression (STAR) Models

Motivation

- AM allow for non-parametric effects, but sometimes even more flexibility is needed
- STAR (Fahrmeir et al., 2003) also support structured terms, which include:
    1. Nonlinear effects of a single variable
    2. Spatial effects of location index s
    3. Interactions between a continuous covariate and a categorical variable
    4. Nonlinear interactions between two continuous covariates
    5. Random Effects with intercept ν_0 and slope ν_j deviations from main effects

Model specification

$$y_i = \underbrace{\kappa_i^{add}}_{\text{AM components}} + f_{struc}(\mathbf{v}_{i1}) + \epsilon_i$$

where $\mathbf{v}_1$ can be a one- or multidimensional variable.

Generalized STAR Models

Motivation

- AM and STAR assume normalty and directly model $E(y)$
- Generalized STAR models use link function $g(\cdot)$ of Generalized Linear Models
- Adds ability to model $E(y)$ of all exponential families, e.g. binomial or poisson distribution

Model specification

$$\begin{aligned} g(\mu_i) &= \eta_i \\ \eta_i &= f_1(z_{i1}) + \ldots + f_K(z_{iK}) + \beta_0 + \beta_1 x_{i1} + \ldots + \beta_Q x_{iQ} \end{aligned}$$

Structured Additive Distributional Regression

Motivation

- Often, more than just the location of a distribution is of interest
- Scale/Shape (Variance, Kurtosis) might also be dependent on covariates
- Structured Additive Distributional Regression allows modeling of all distributional parameters θ_l

Model specification

Let $y \sim D(\theta_1, \ldots, \theta_K)$. Then:

$$\begin{aligned} g_l(\boldsymbol{\theta}_l) &= \boldsymbol{\eta}_l \\ &= f_{1l}(\mathbf{X}; \boldsymbol{\beta}_{1l}) + \ldots + f_{Q_l l}(\mathbf{X}); \boldsymbol{\beta}_{Q_l l}) \end{aligned}$$

where every θ_l can be modeled with effect types of different subsets of $\mathbf{X}$.

Bayesian Models for Location, Scale and Shape

overview...

Thanks!

Thanks for your attention!

Literatur

L. Fahrmeir, T. Kneib, and S. Lang. Penalized additive regression for space-time data: a bayesian perspective, 2003. URL `http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-1687-9`.

J.H. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of the American statistical Association*, 76(376):817–823, 1981.

R.A. Rigby and D.M. Stasinopoulos. The gamlss project: a flexible approach to statistical modelling. In *New trends in statistical modelling: Proceedings of the 16th international workshop on statistical modelling*, volume 337, page 345. University of Southern Denmark, June 2001.

References ii

N. Umlauf, N. Klein, and A. Zeileis. Bamlss: Bayesian additive models for location, scale and shape (and beyond). Working papers, Working Papers in Economics and Statistics, 2017. URL `https://EconPapers.repec.org/RePEc:inn:wpaper:2017-05`.