bamlss: A Lego Toolbox for Flexible Bayesian Regression (and Beyond)

Nikolaus Umlauf, Nadja Klein, Thorsten Simon and Achim Zeileis

May 25, 2020

1 General comments

This is a rather impressive distributional regression package and the authors needed to be congratulated for the amount of thought and work put into it. Without damping their enthusiasm, I would have like to point out, though, that a lot of the ideas, material and models, used here, have been around for some time. Also, the Lego toolbox idea, that is, combining algorithm and method together to create new and potential more powerful models is not new either. I would suggest that there should be greater acknowledgement of this.

The authors explain the features of the package using three different examples. The first two examples are introduced in section 2, while the third is in section 5. The first example uses a binary response. This is a rather strange choice because for binary response variables the GAMLSS methodology has nothing to add compared to GLM's and GAM's. I guess, it can be seen as a gentle introduction to the package. The second example is the crash helmet data which has been extensively used over the last 35 years (including the authors in their BAMLSS 2018 paper). I am sure that there are more interesting simple data sets than can be used to demonstrate the BAMLSS package. The third example in Section 5 it more substantial and very interesting. It demonstrates the power of using boosting within BAMLSS but it misses one important aspect of a distribution regression framework. Modelling the distribution. It fits a rather inadequate distribution to the data. Section 3 described the underpinning theoretical work. Section 4 the BAMLSS package. There is no conclusion section (and maybe there should be one describing what has been achieved and what the authors would like to achieve in the future. There are also 4 Appendices.

I my opinion the following are main the advantages of the **bamlss** package:

- It integrates distribution regression techniques together.
- It makes the ideas of the GAMLSS framework attractive to Bayesian statistician.

- By combining the penalise likelihood (backfitting) approach with MCMC, it also makes MCMC more attractive to non-Bayesians.
- It shows the advance of the Bayesian approach in providing additional posterior information about the parameters of the model. [The same information is often more difficult to obtain with a more traditional likelihood based approaches].
- It demonstrates how traditional model checking and diagnostic techniques can be also applied to Bayesian fitted models.

There are also few things that I found myself wondering after I finished reading the paper that need clarification.

- The main smoothers considered are similar to the ones in package **mgcv**. It was not clear to me whether the functions are straightforward copies of the functions in **mgcv** or new functions adopting the **mgcv** notation. Also, how easy is to write a new smother function for `bamlss`.
- There are about 25 different distributions in the package. Some of them are very important, but the list falls short of the variety of univariate distributions existing in the **gamlss.dist** package. The authors claim that all the GAMLSS distributions can be fitted in BAMLSS but unfortunately they give no example of doing so. For example in example 3 the distribution used, the zero inflated negative binomial, seems to be inadequate for the training data (see the worm plot at the end of this report). So I am left with wondering how long I have to spend to implement a new distribution in BAMLSS from the ones existing in GAMLSS? one minute? one hour? one day? one week?
- Within their approach, the selection of prior distributions is rather played down, which does not distract from the problem of finding an adequate model for the data. But how sensitive are the results to changing the prior distributions used?
- In all three examples there is no distribution modelling. I can understand this in the first example, but why accept the normal distribution and zero inflated negative binomial distributions without considering alternatives?
- I do not like '(and beyond)' in the title. What is beyond "Flexible Bayesian Regression"?. I understand that there are trying to say beyond univariate regression but I think this is rather irrelevant here. "bamlss: A Lego Toolbox for Flexible Bayesian Regression" will do.

2 Specific comments

The following are comments on the text :

page 2 l 17 "gamlss family of packages (Stasinopoulos and Rigby 2007)" maybe should be replaced with "gamlss family of packages (Stasinopoulos et al. 2017 and Rigby et al. 2019)"

Stasinopoulos et al. (2017) *Flexible Regression and Smoothing: Using GAMLSS in R*, Chapman and Hall/CRC.

Rigby et al. (2019) *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*, Chapman and Hall/CRC.

page 2 l 19 "However, for complex predictor structures and response distributions beyond the exponential family, estimation may be challenging or subject to numerical instabilities."

Although this can sometimes be correct, it sounds like encouraging people to think that, "if a specific distribution is unstable to fit then go back to the exponential family". The exponential family, as we know, does have good theoretical properties but often is unsuitable for the data. There are distributions which are more difficult to fit than others because of the non orthogonality of the parameters. Also, If you are are trying to find a distribution to your data and try a lot of them there is a good change that some of those distributions are difficult to fit because they are not appropriate, that is, do not capture the features of the data.

page 5 line -11 "AICc = 1033.737" and "logLik = -508.7851" both consistent with glm and gam results but AIC of model b is 1042.228. I would like to know how the AIC is calculated.

page 5 line -5 "by the upstream backfitting algorithm". Nothing to backfit here just IWLS.

page 6 line 5 "As mentioned above, the user could also increase the number iterations and the burnin-phase, as well as adapt the thinning parameter, to make the significant bar at lag one disappear." Maybe you should show the argument here to make it easier for the user.

page 7 line 10 "heavily builds upon the R package mgcv" How?

page 10 l 3 I was wondering whether using the simple AIC would have shown the upward trend in the beginning

page 10 l 9 Not again the Silverman (1985) example! This is also used by the authors in their 2018 paper.

page 11 l -8 "seem to be less appropriate". In fact a worm plot appears to show that the residuals are OK even in those values.

page 12 l4 Maybe a small example of the use of CRPS will help.

page 12 l -10 "Umlauf et al. (2018) relax this assumption and let $f_{jk}(.)$ be an unspecified composition of covariate data and regression coefficients. For example, functions $f_{jk}(.)$ could also represent nonlinear growth curves, a regression tree, a neural network or lasso-penalized model terms". Stasinopoulos et al. (2017) *Flexible regression and smoothing: using GAMLSS in R*, Chapman and Hall/CRC, Chapter 9, were doing this before.

page 12, l -1 "Similarly, α_{jk} is the set of all prior specifications. A set of what? parameters.

page 13, l 4 I am assuming that the priors are trying to be uninformative. How uninformative are they?

page 13, l 6 "Univariate responses of any type ...". The references have a "Bayesian" bias since Rigby and Stasinopoulos have covered over the years all the distributional problems mentioned here. Their work and GAMLSS based solutions to the problems are well documented in Rigby et al. (2019).

page 13 l -13 "(Smyth 1996)". Aitkin (1987) used exactly the same zigzag algorithm to model the μ and σ^2 in the normal distribution. [Aitkin, M (1987) Modelling variance heterogeneity in normal regression using GLIM, *Applied Statistics*, pp 332-33, Vol 36.]

page 13 l -4 It was Rigby and Stasinopoulos (2005) Appendix C that showed that the zigzaging equation (4) applied over the parameters leads to the maximisation of the posterior mode. This is a rather crucial point to omit.

page 14 l -16 Please define EDF. Is it, equivalent or effective, degrees of freedom?

page 15 l 6 "flexible Bayesian regression models (and beyond)." What is beyond Bayesian regression models?

page 18 l -7 "In addition, all families of the gamlss (Stasinopoulos and Rigby 2019a) and gamlss.dist (Stasinopoulos and Rigby 2019b) package are supported." How this can be done?

page 20 l 18 "However, commonly used distributions are already implemented in bamlss; and the ones from the gamlss package can also be accessed through the bamlss package." Again as above how?

page 21 table 3 Surely you need also the "p" function for the quantile residuals.

page 21 l 12 What happens for distributions for which the expectation of the second derivative is intractable or does not exist ?

page 24 l 6 "In order to capture the truncation of the data and its overdispersion we employ a zero-truncated negative binomial distribution (Cameron and Trivedi 2013), which is specified by two parameters $\mu > 0$ and $\theta > 0$." I understand the zero truncation part and that negative binomial as a starting point but there is no statistical modelling here. For example, the truncated Sichel distribution fits lot better to the data than the truncated negative binomial as it demonstrated by the worm plots displaying both models in Figure 1.

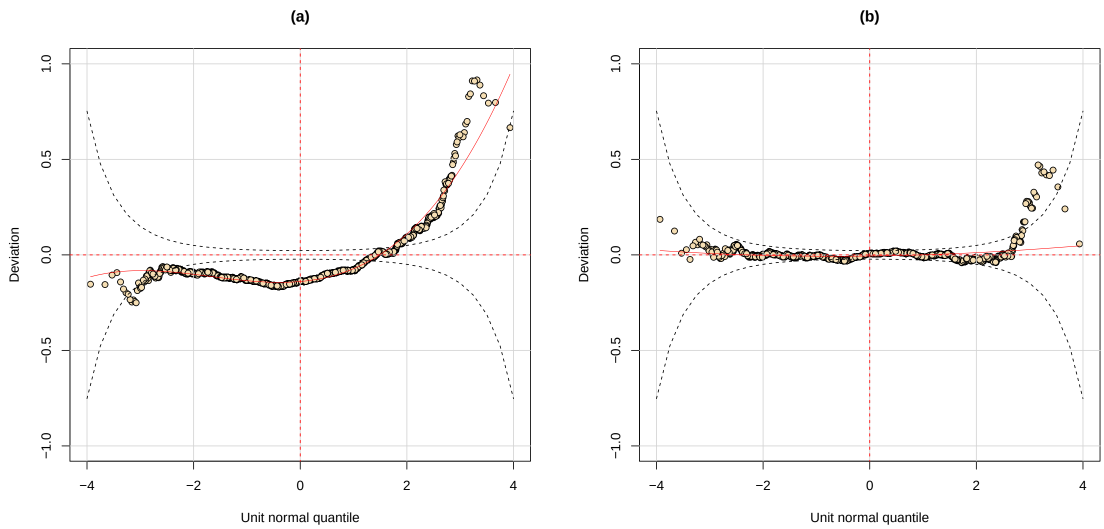

Figure 1: Worm plots (a) a fitted zero truncated negative binomial ; (b) a fitted zero truncated Sichel distribution

page 27 l -9 It appears that something is wrong here '`max(fit$mu)` [1] 2.231437e+14'. Please check the fitted values for both μ and θ.

page 30 l 14 I think you need conclusions here.