Advanced Bayesian Methods: Theory and Applications in R

Model Checking and Predictive Evaluation

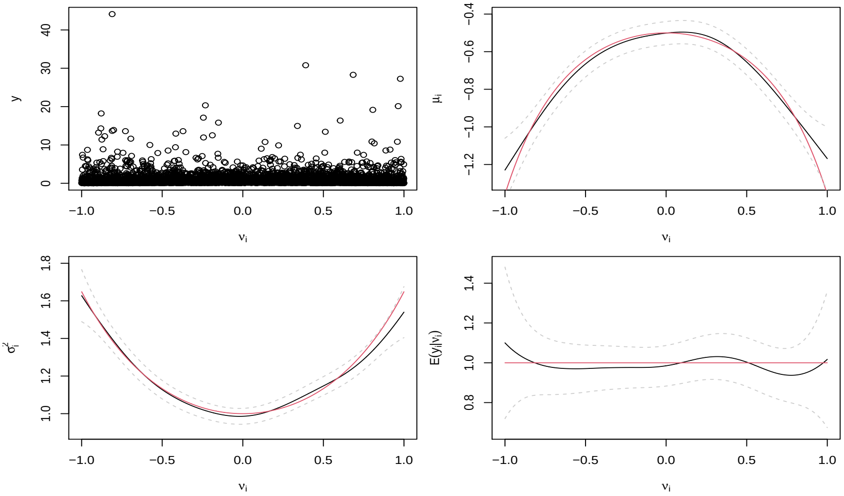
Nikolaus Umlauf

https://nikum.org/abm.html

Challenges in Distributional Regression

- Conceptually, distributional regression is very appealing and intuitive, but it comes with a number of specific challenges:
  - Interpretation of the estimated effects more difficult due to link functions and multi-parameter setup.
  - Model choice and checking to avoid model miss-specification.

Illustration for Simulated Log-Normal Data



Goals of Model Checking and Predictive Evaluation

- Finding a suitable response distribution.
- Determining the predictor (covariates and appropriate modelling variant).
- Checking model adequacy in general.

Quantile Residuals

- For a continuous random variable Y with cumulative distribution function F, the probabilitiy integral transform yields

$$F(Y) \sim \mathrm{U}(0,1)$$

  or

$$\Phi^{-1}(F(Y)) \sim \mathrm{N}(0,1).$$

- For a correctly specified distributional regression model, we should therefore have

$$u_i = \Phi^{-1}(F(y_i|\hat{\boldsymbol{\theta}}(\boldsymbol{x}_i))) \overset{a}{\sim} \mathrm{N}(0,1)$$

  and the quantile residuals u_i can, e.g., be visualized in a quantile-quantile plot.

- For discrete or multivariate data, appropriate generalisations are needed.

Information Criteria

- Information criteria such as AIC or BIC can be used in the distributional regression context, e.g.

$$\text{AIC} = -2l(\hat{\gamma}) + 2\,\text{df}(\hat{\gamma})$$

- The model fit is evaluated based on the (negative) log-likelihood $-2l(\hat{\gamma})$.
- The degrees of freedom $\text{df}(\hat{\gamma})$ of the model have to take the impact of the regularisation penalty into account.

Proper Scoring Rules

- In a distributional setting, typical predictive measures such as the mean squared error of prediction or the mean absolute error of prediction are usually not the most adequate choice.
- Proper scoring rules provide a framework for evaluating predictive distributions rather than point predictions.
- Underlying theory ensures that proper scores encourage the analyst to honestly report their uncertainty in terms of the predictive distribution.
- The cross-validated log-likelihood is the most commonly used proper score.

Scoring Rules for Real-Valued Outcomes

- Evaluate a predictive density $f(y)$ based on an observed outcome y_0.
- Spherical score
$$\mathrm{SPS}(f(y), y_0) = -\frac{f(y_0)}{\left(\int f^2(t)dt\right)^{1/2}}.$$
- Logarithmic score
$$\mathrm{LS}(f(y), y_0) = -\log(f(y_0)).$$
- Continuously ranked probability score
$$\mathrm{CRPS}(f(y), y_0) = \int \left[F(t) - 1_{[y_0,\infty)}(t)\right]^2 dt.$$
- Note: All scores are negatively oriented, i.e. smaller values indicate a better agreement between the predictive distribution and the observed values.

Model Checking in *bamlss*

Example:

Load the Munich rent data.

```
R> library("bamlss")
R> library("gamlss.dist")
R> data("rent", package = "gamlss.data")
```

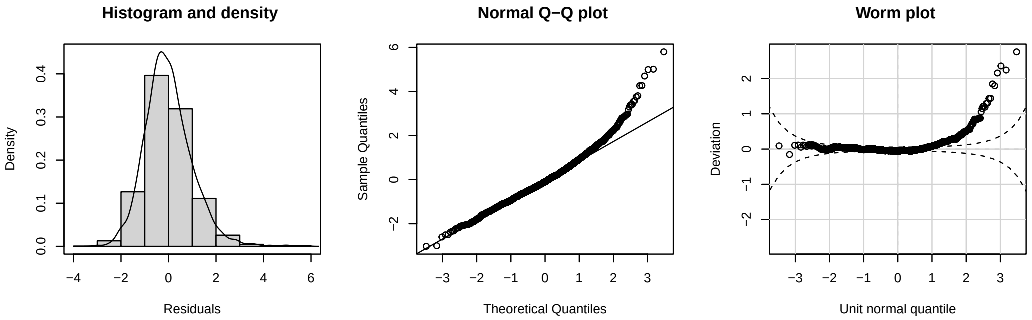
Estimate models.

```
R> f <- R ~ s(Fl) + s(A) + loc + H
R> b1 <- bamlss(f, data = rent, family = NO)
R> b2 <- bamlss(f, data = rent, family = GA)
```

Model Checking in *bamlss*

Diagnostic plots.

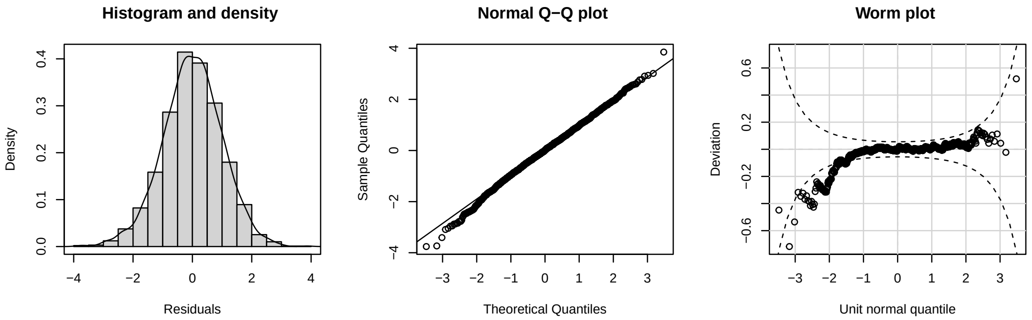
```
R> par(mfrow = c(1, 3))
R> plot(b1, which = 3:5, spar = FALSE)
```



Model Checking in *bamlss*

Diagnostic plots.

```
R> par(mfrow = c(1, 3))
R> plot(b2, which = 3:5, spar = FALSE)
```



Model Checking in *bamlss*

Log-likelihood and AIC.

```
R> logLik(b1)
'log Lik.' -14038.36 (df=11.39594)
R> logLik(b2)
'log Lik.' -13847.51 (df=11.94265)
R> AIC(b1, b2)
         df      AIC
b1 11.39594 28099.51
b2 11.94265 27718.90
R> DIC(b1, b2)
        DIC       pd
b1 28088.11 11.39594
b2 27706.96 11.94265
R> WAIC(b1, b2)
      WAIC1    WAIC2       p1       p2
b1 28089.75 28089.93 13.03652 13.12346
b2 27707.25 27707.54 12.23709 12.38185
```

Model Checking in *bamlss*

Scoring rules.

```
R> library("scoringRules")
R> p1 <- predict(b1, type = "parameter")
R> p2 <- predict(b2, type = "parameter")
R> mean(crps_norm(rent$R, mean = p1$mu, sd = p1$sigma))
```

```
[1] 165.4988
```

Using *bamlss* infrastructures.

```
R> s1 <- bamlss:::.CRPS(rent$R, as.data.frame(p1), family(b1))
R> s2 <- bamlss:::.CRPS(rent$R, as.data.frame(p2), family(b2))
R> mean(s1)
```

```
[1] 165.4988
```

```
R> mean(s2)
```

```
[1] 161.1239
```

Model Checking in *bamlss*

Quantile residuals by hand. First predict parameters.

```
R> par <- predict(b1, newdata = rent, type = "parameter")
```

Calculate probabilities using the `$p()` function.

```
R> u <- family(b1)$p(rent$R, par)
```

Compute standard normal quantiles, aka quantile residuals.

```
R> e <- qnorm(u)
```

Model Checking in *bamlss*

Q-Q plot.

```
R> qqnorm(e); qqline(e)
```