Beyond mean modelling: GAMLSS and quantile GAMs

Matteo Fasiolo and Simon N. Wood

matteo.fasiolo@bristol.ac.uk

June 27, 2019

GAMLSS

Structure:

1. Intro to GAMs for Location Scale and Shape
2. Intro to quantile GAMs
3. GAMLSS and QGAM modelling in mgcv and qgam

Beyond mean modelling: GAMLSS models

Structure:

1. **Intro to GAMs for Location Scale and Shape**
2. Intro to quantile GAMs
3. GAMLSS and QGAM modelling in mgcv and qgam

Intro to GAMLSS models

Recall GAM model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu(\mathbf{x}), \theta_2, \dots, \theta_p\},$$

where

$$\mu(\mathbf{x}) = g^{-1}\Big\{\sum_{j=1}^{m} f_j(\mathbf{x})\Big\},$$

and g is the link function.

Example, Scaled Student-t distribution:

- location $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$
- scale $\theta_2 = \sigma$
- shape $\theta_3 = \nu$

Intro to GAMLSS models

In Generalized Additive Models for Location Scale and Shape (GAMLSS) (Rigby and Stasinopoulos, 2005) we let scale and shape change with the covariates $\mathbf{x}$.

GAMLSS model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu_1(\mathbf{x}), \theta_2 = \mu_2(\mathbf{x}), \dots, \theta_p = \mu_p(\mathbf{x})\},$$

where

$$\mu_1(\mathbf{x}) = g_1^{-1}\Big\{\sum_{j=1}^{m} f_j^1(\mathbf{x})\Big\},$$
$$\dots$$
$$\mu_p(\mathbf{x}) = g_p^{-1}\Big\{\sum_{j=1}^{m} f_j^p(\mathbf{x})\Big\},$$

and $g_1, \dots, g_p$ are link function.

Intro to GAMLSS models

Example: **Gaussian model for location and scale**

Model is

$$y|\mathbf{x} \sim N\{y|\mu(\mathbf{x}), \sigma(\mathbf{x})\}$$

where

$$\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = \sum_{j=1}^{m} f_j^1(\mathbf{x})$$

$$\text{var}(y|\mathbf{x})^{1/2} = \sigma(\mathbf{x}) = \exp\Big\{\sum_{j=1}^{m} f_j^2(\mathbf{x})\Big\}$$

that is $g_2 = \log$ to guarantee $\sigma > 0$.

Intro to GAMLSS models

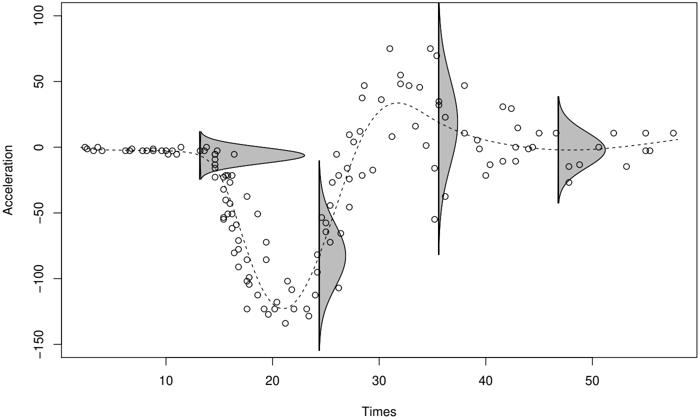
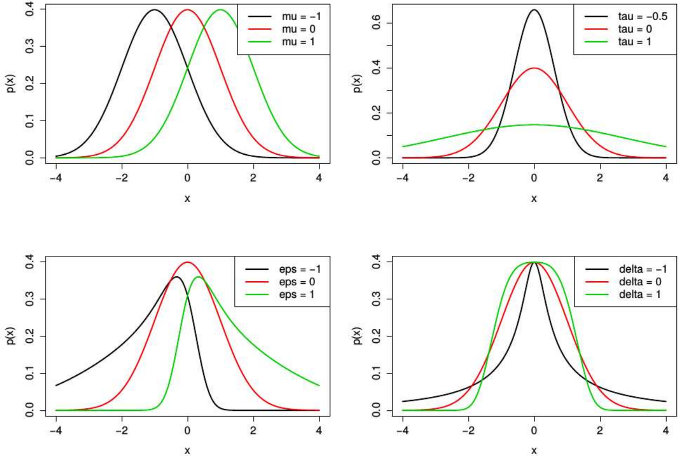


Figure: Gaussian model with variable mean and variance.
In mgcv: `gam(list(y~s(x), ~s(x)), family=gaulss)`.

Example: **Sinh-arcsinh (shash) distribution**

Four parameter distribution where location, scale, skewness (asymmetry) and kurtosis (tail behaviour) can depend on **x** (Jones and Pewsey, 2009).



Intro to GAMLSS models

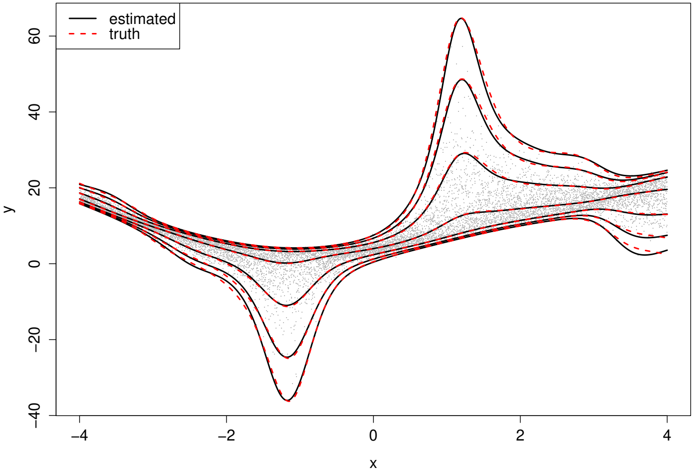
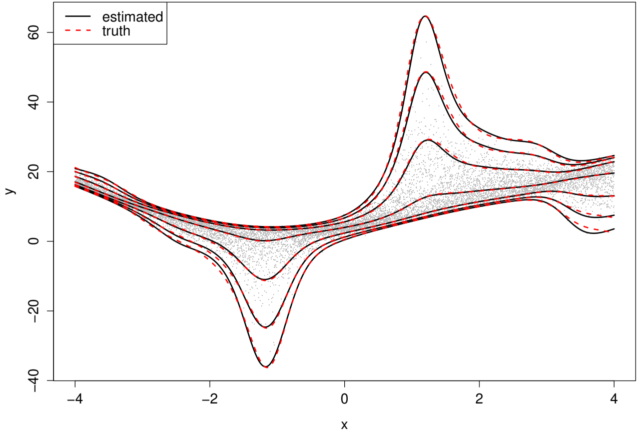


Figure: `gam(list(y~s(x), ~s(x), ~s(x), ~s(x)), family=shash)`.

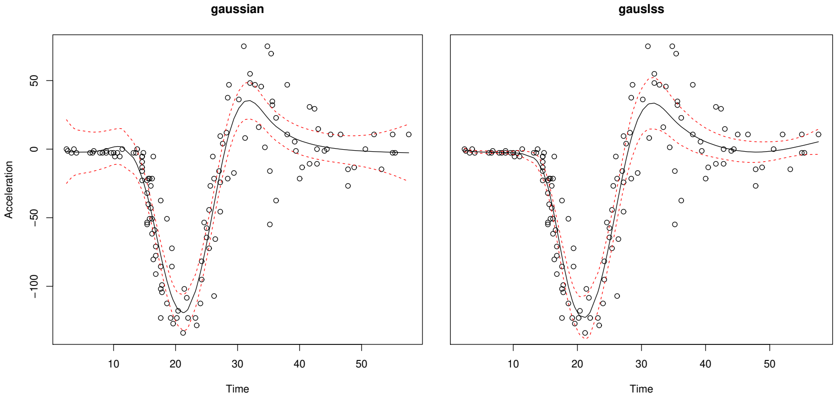
Why is this useful?

R1: you might be interested in whole distribution $y|\mathbf{x}$ not just $\mathbb{E}(y|\mathbf{x})$.



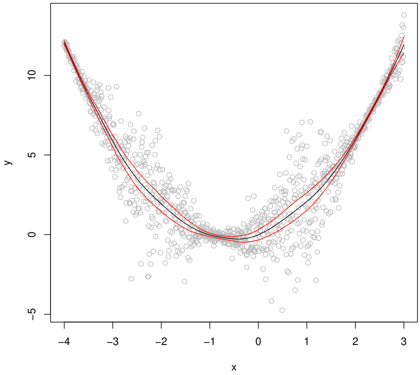
Why is this useful?

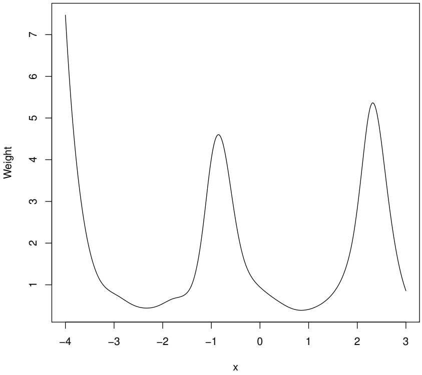
R2: standard GAM inference (e.g. p-value & confidence interval) is valid if the model for $p(y|\mathbf{x})$ is correct



Why is this useful?

R3: the accuracy of the fit is improved if the weight of each observation is inversely proportional to $\text{Var}(y|\mathbf{x})$.





Beyond mean modelling: quantile GAMs

Structure:

1. Intro to GAMs for Location Scale and Shape
2. **Intro to quantile GAMs**
3. GAMLSS and QGAM modelling in mgcv and qgam

What is quantile regression

Regression setting:

- y is our response or dependent variable
- $\mathbf{x}$ is a vector of covariates or independent variables

In **distributional regression** we want a good model for $p(y|\mathbf{x})$.

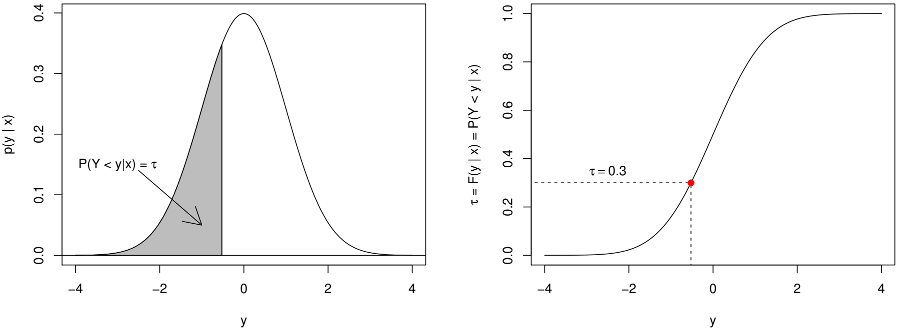
Model is $p_m\{y|\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})\}$, where $\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})$ are parameters.

Lots of options for $p_m(y|\mathbf{x})$: binomial, gamma, Poisson, Tweedie...

We consider continuous (not discrete) y.

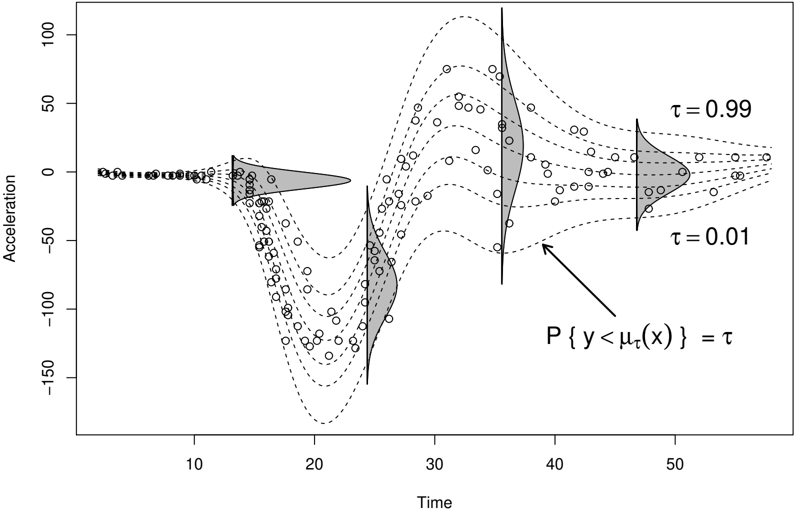
Define $F(y|\mathbf{x}) = \text{Prob}(Y \leq y|\mathbf{x})$.

The τ-th ($\tau \in (0, 1)$) quantile is $\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x})$.



What is quantile regression

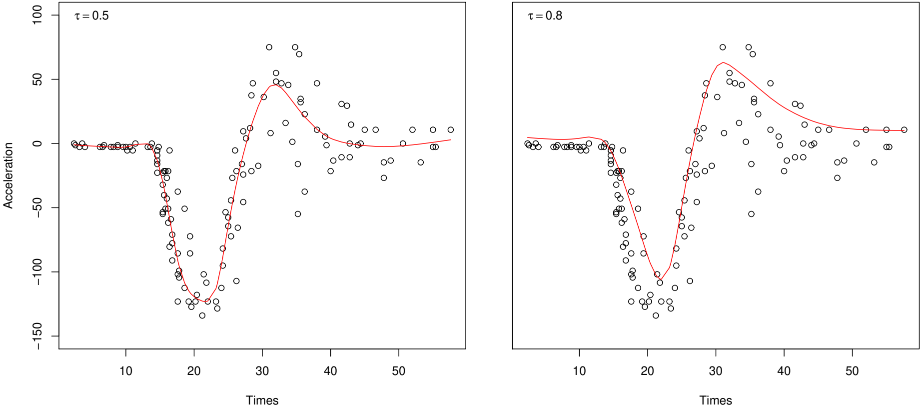
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



What is quantile regression

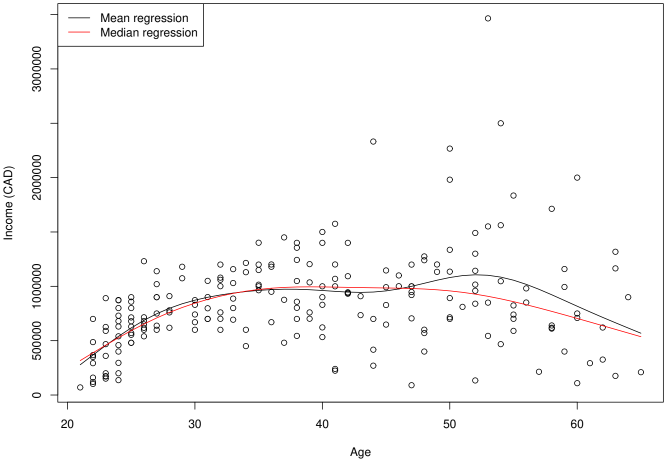
Quantile regression estimates conditional quantiles $\mu_\tau(\mathbf{x})$ directly.

No model for $p(y|\mathbf{x})$.



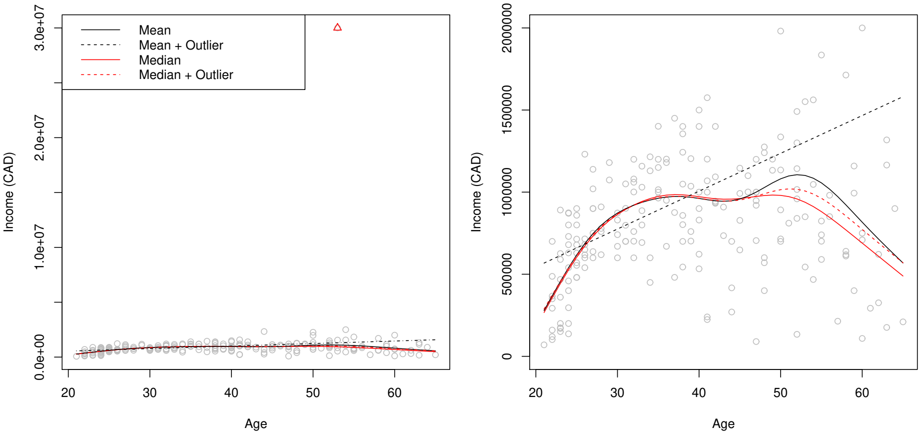
When is quantile regression useful

Median income is a better indicator of how the “average” person is doing, relative to mean income.



When is quantile regression useful

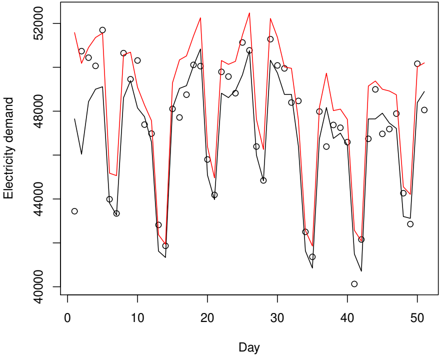
The median is also more **resistant to outliers**.

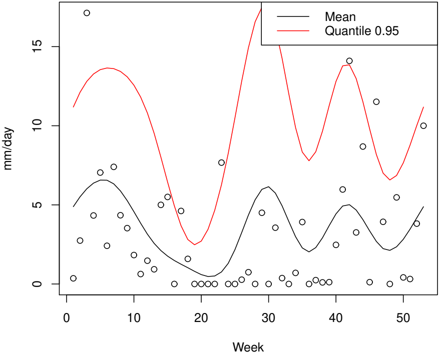


When is quantile regression useful

Some quantiles are more important than others:

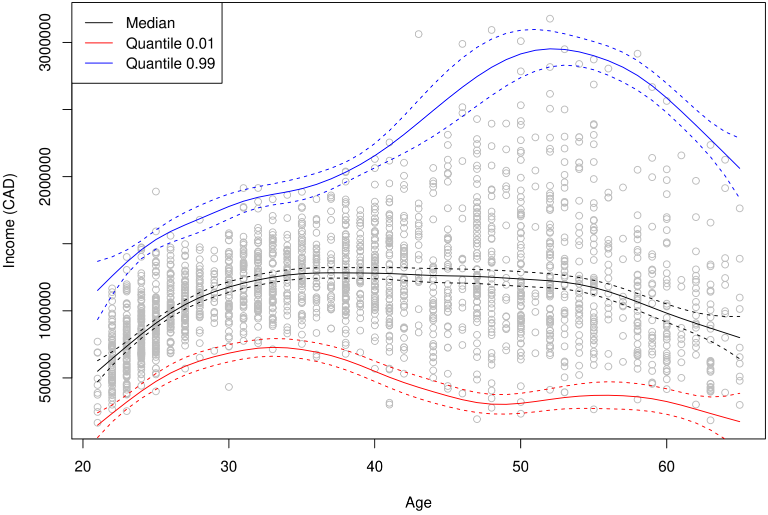
- electricity producers need to satisfy top electricity demand
- urban planners need estimates of extreme rainfall





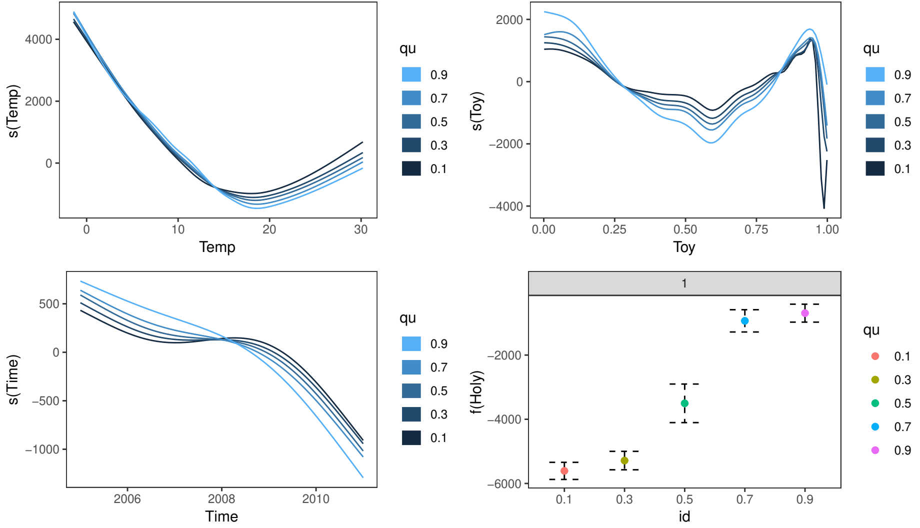
When is quantile regression useful

Effect of explanatory variables may depend on quantile



When is quantile regression useful

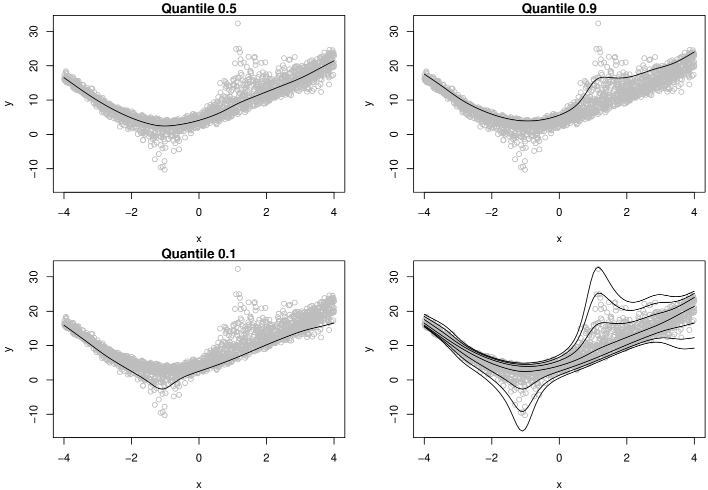
$$q_\tau(\text{Demand}) = f_1(\text{Temp}) + f_2(\text{TimeOfYear}) + f_3(\text{Trend}) + f_4(\text{Holiday}) + \cdots$$



When is quantile regression useful

No assumptions on $p(y|\mathbf{x})$:

- no need to find good model for $p(y|\mathbf{x})$;
- no need to find normalizing transformations (e.g. Box-Cox);



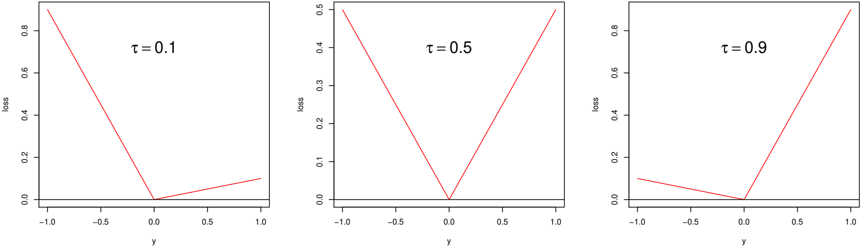
Quantile GAM estimation

Recall definition $\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x})$.

Key fact: $\mu_\tau(\mathbf{x})$ is the minimizer of

$$L(\mu|\mathbf{x}) = \mathbb{E}\{\ \rho_\tau(y-\mu)\,|\,\mathbf{x}\ \},$$

where ρ_τ is the "pinball" loss (Koenker, 2005):



In additive modelling context $\mu_\tau(\mathbf{x}) = \mu_\tau(\boldsymbol{\beta}) = \sum_{j=1}^m f_j(\mathbf{x})$.

Quantile GAM estimation

Problem: how to perform Bayesian update $p(\boldsymbol{\beta}|y) \propto p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})$?

A coherent approach is the Gibbs posterior (Bissiri et al., 2016)

$$p(\boldsymbol{\beta}|y) \propto \underbrace{e^{-\frac{1}{\sigma}\rho_\tau\{y-\mu(\boldsymbol{\beta})\}}}_{\text{pseudo } p(y|\boldsymbol{\beta})} p(\boldsymbol{\beta}),$$

where $1/\sigma > 0$ is the "learning rate".

Recall that $p(\boldsymbol{\beta}) = p(\boldsymbol{\beta}|\boldsymbol{\lambda})$, hence we need to:

- select learning rate $1/\sigma$
- select smoothing parameters $\boldsymbol{\lambda}$
- estimate regression coefficients $\boldsymbol{\beta}$

Quantile GAM estimation

We use a hierarchical fitting framework:

1. Select $\boldsymbol{\sigma}$ to optimise coverage of credible intervals

$$\hat{\boldsymbol{\sigma}} = \underset{\boldsymbol{\sigma}}{\operatorname{argmax}} \, \mathrm{IKL}(\sigma).$$

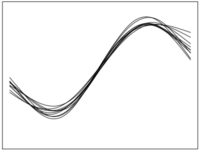
2. Select $\boldsymbol{\lambda}$ determine smoothness

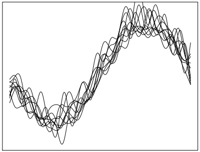
$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} \, \mathrm{LAML}(\boldsymbol{\lambda})$$

   where $\mathrm{LAML}(\boldsymbol{\lambda}) \approx p(y|\boldsymbol{\lambda}) = \int p(y\boldsymbol{\beta}|\boldsymbol{\lambda})d\boldsymbol{\beta}$.

3. For fixed $\boldsymbol{\lambda}$ and σ, estimate $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \, \log p(\boldsymbol{\beta}|y).$$

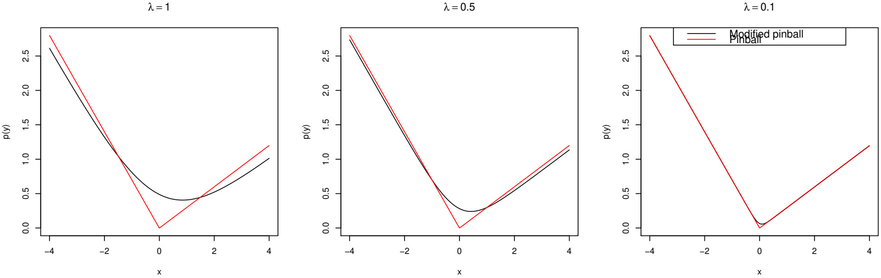




Quantile GAM estimation

`qgam` uses a modified loss which we call Extended log-F (ELF) loss.

This is smooth and convex and, as $\lambda \to 0$, we have recover pinball loss.



Since `qgam` 1.3.0, λ (`err` parameter) is selected automatically.

Learning rate can depend on covariates $\sigma = \sigma(\mathbf{x})$ (see examples).

Structure:

1. Intro to GAMs for Location Scale and Shape
2. Intro to quantile GAMs
3. **GAMLSS and QGAM modelling in mgcv and qgam**

Demonstration in R

For more details on methodology, see:

Fasiolo, M., Goude, Y., Nedellec, R. and Wood, S.N., 2017. Fast calibrated additive quantile regression. arXiv preprint arXiv:1707.03307.

and the file "intro_to_qgam.pdf".

For more software training material see

`http://mfasiolo.github.io/qgam/articles/qgam.html`

`https://mfasiolo.github.io/mgcViz/articles/qgam_mgcViz.html`

Now we move to "gamlss_qgam.html"

References I

Bissiri, P. G., C. Holmes, and S. G. Walker (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Jones, M. and A. Pewsey (2009). Sinh-arcsinh distributions. *Biometrika 96*(4), 761–780.

Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 54*(3), 507–554.