More mgcViz tools

Matteo Fasiolo

Joint work with:

Simon N. Wood (University of Bristol, UK)
Yannig Goude (EDF R&D)
Margaux Zaffran (ENSTA Paris)
Raphaël Nedellec (Talend, formerly EDF R&D)

Material available at:

https://github.com/mfasiolo/workshop_EDF19

More tools mgcViz: talk structure

These slides cover:

1. Quantile GAMs
   - What are quantile GAMs (QGAMs)
   - Using QGAMs with qgam and mgcViz
2. Loss-based checks
   - Loss-based vs goodness-of-fit checks
   - Future developments
3. Hands-on session

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.

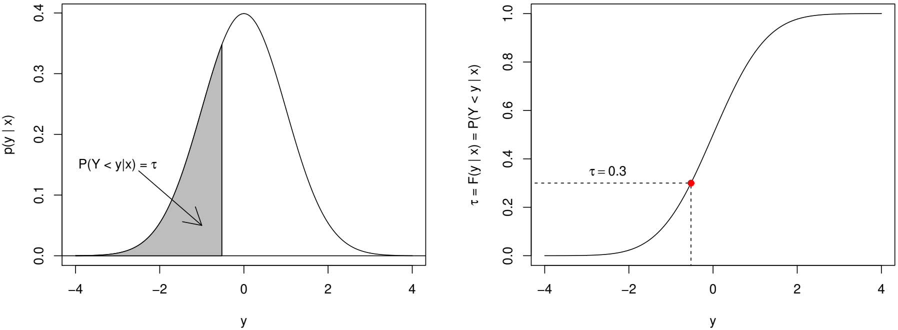
What is quantile regression

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

We consider continuous (not discrete) y.

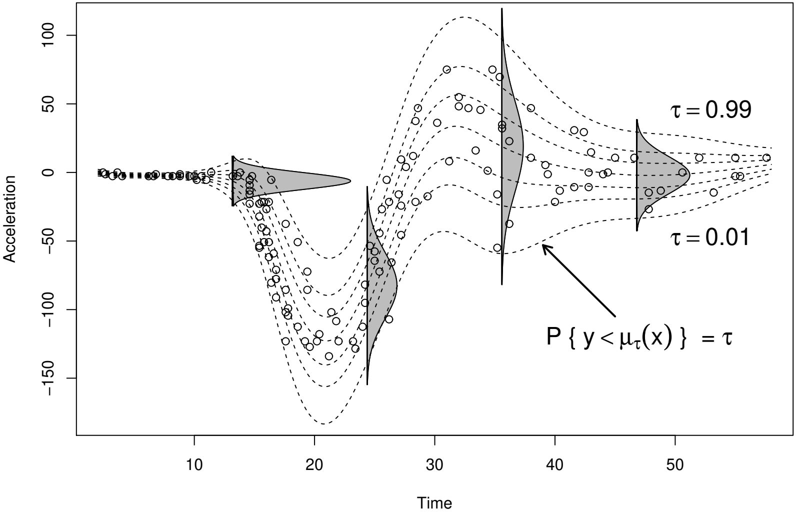
Define $F(y|\mathbf{x}) = \text{Prob}(Y \leq y|\mathbf{x})$.

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



What is quantile regression

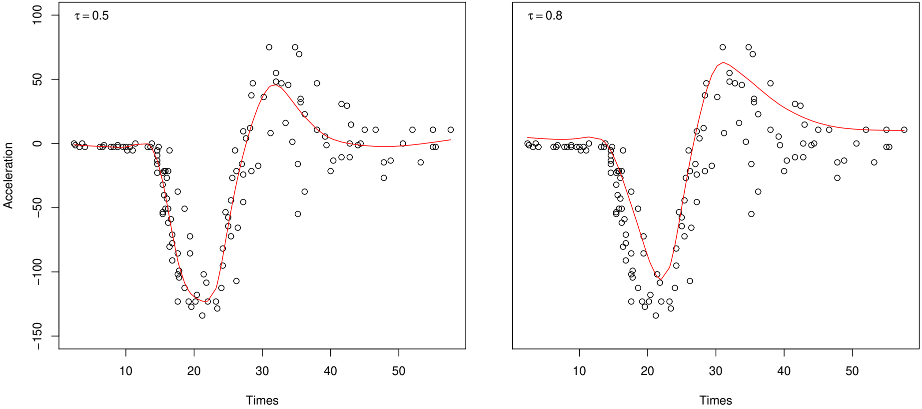
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



What is quantile regression

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

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



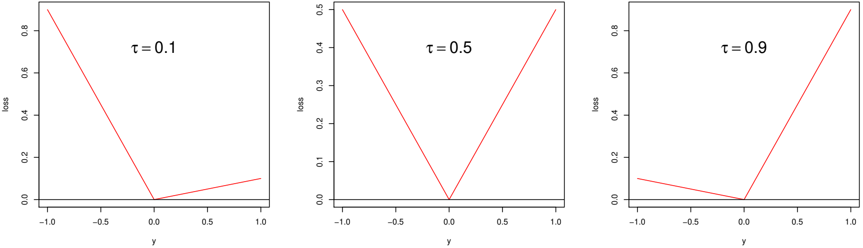
Quantile GAM estimation

In parametric GAMs $\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x})$.

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

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)

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

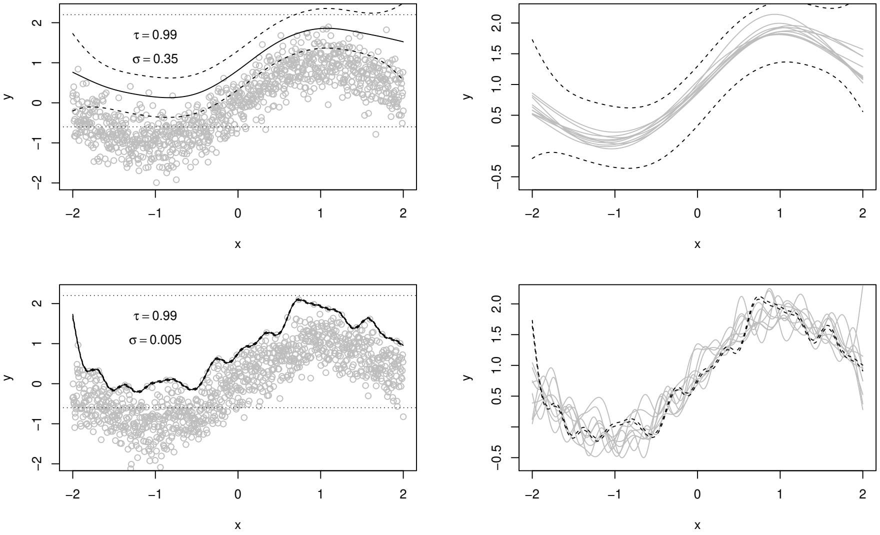
where $1/\sigma > 0$ is the "learning rate".

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

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

Technical challenges

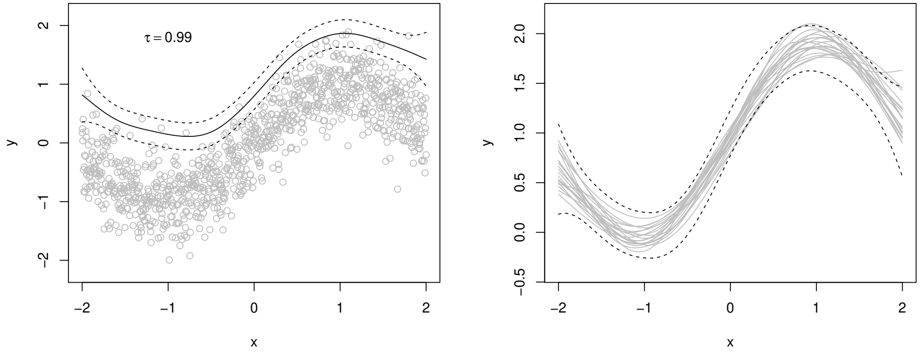
σ controls width of credible intervals:



Selecting the learning rate

We select σ so that the model-based uncertainty estimates match the sampling uncertainty, that is σ minimizes

$$\text{CalibrLoss}(\sigma) = \int \text{Dist}\{\text{var}_m(\mathbf{x}), \text{var}_{\mathbb{P}}(\mathbf{x})\} p(\mathbf{x}) d\mathbf{x},$$



Quantile GAM estimation

We use a hierarchical fitting framework:

1. Select σ to optimise coverage

$$\hat{\sigma} = \underset{\sigma}{\text{argmin}}\, \text{CalibrLoss}(\sigma).$$

2. For fixed σ, select γ to determine smoothness

$$\hat{\gamma} = \underset{\gamma}{\text{argmax}}\, \text{LAML}(\gamma)$$

where $\text{LAML}(\gamma) \approx p(y|\gamma) = \int p(y, \boldsymbol{\beta}|\gamma) = \int p(y|\boldsymbol{\beta})p(\boldsymbol{\beta}|\gamma)d\boldsymbol{\beta}$.

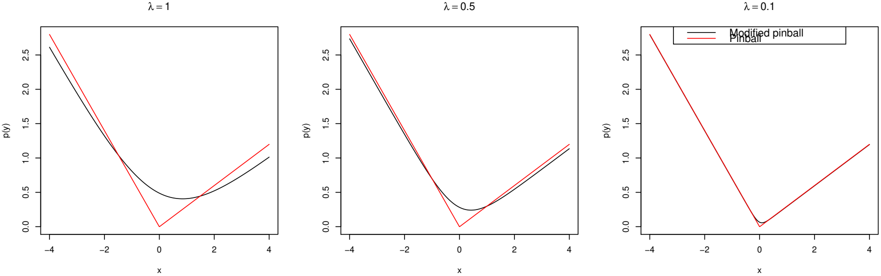
3. For fixed γ and σ, estimate $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \sum_i \rho_\tau\{y_i - \mu(\boldsymbol{\beta})\} + \text{Pen}(\boldsymbol{\beta}|\gamma).$$

Quantile GAM estimation

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

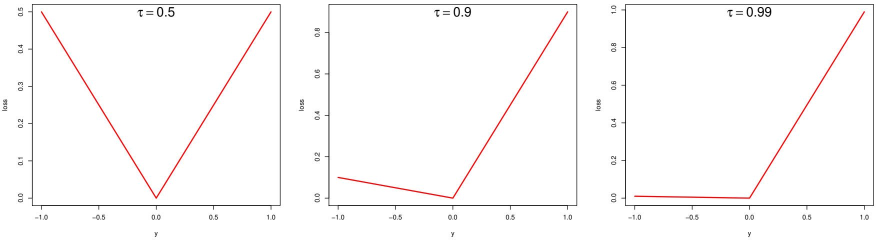
This is smooth and convex and, as $h \to 0$, we have recover pinball loss.



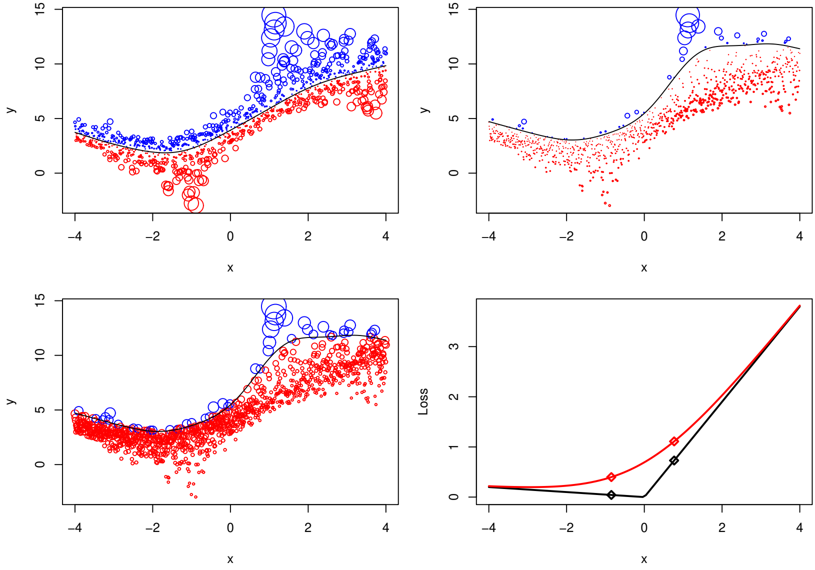
Since `qgam` 1.3.0, h (`err` parameter) is selected automatically.

Motivation for using ELF:

pinball loss becomes very asymmetric on extreme quantiles.



Smoothing the pinball loss



h (called `err` in `qgam`) selected to **balance variance and bias**.

Estimating quantile GAMs

Let h be the smoothness of loss, we derive

$$\begin{aligned}\text{AMSE}(h) =& n\mathbb{E}\{(\hat{\boldsymbol\beta} - \boldsymbol\beta_0)(\hat{\boldsymbol\beta} - \boldsymbol\beta_0)^\mathsf{T}\} \\ =& \mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{V}[\underbrace{nh^4\mathbf{B}}_{bias} - \underbrace{h\mathbf{A}}_{variance}]\mathbf{V}^\mathsf{T} + \cdots .\end{aligned}$$

We let $h = h(\mathbf{x})$ and assume location-scale model

$$y_i = \alpha(\mathbf{x}_i) + \kappa(\mathbf{x}_i) z_i,$$

where z_i are iid errors.

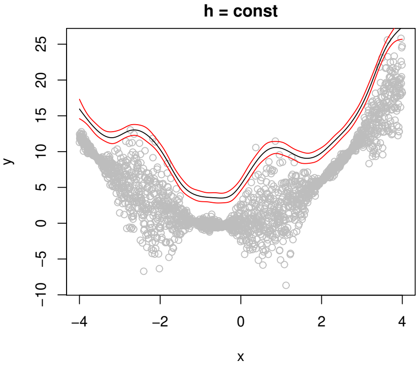
Under this model

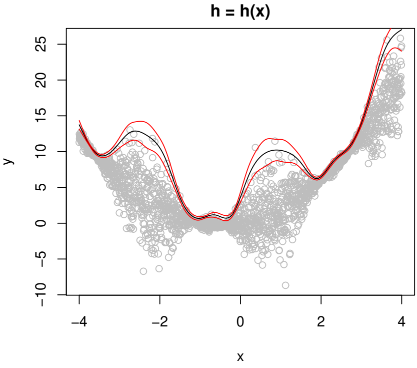
$$h^*(\mathbf{x}) = \underset{h}{\operatorname{argmin}} \,\text{Trace}(nh^4\mathbf{B} - h\mathbf{A}) = \left[\frac{d}{n}\frac{9}{\pi^4}\frac{f_z(\mu_\tau)}{f_z'(\mu_\tau)^2}\right]^{\frac{1}{3}} \kappa(\boldsymbol{x}).$$

Estimating quantile GAMs

A practical procedure:

1. estimate $\alpha(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$ and $\kappa(\mathbf{x}) = \text{var}(y|\mathbf{x})$
2. normalize $z = \{y - \alpha(\mathbf{x})\}/\kappa(\mathbf{x})$
3. estimate f_z and f_z'
4. get $h^*(\boldsymbol{x}) = \left[\frac{d}{n}\frac{9}{\pi^4}\frac{f_z(\mu_\tau)}{f_z'(\mu_\tau)^2}\right]^{\frac{1}{3}}\kappa(\boldsymbol{x})$





More tools mgcViz: talk structure

These slides cover:

1. Quantile GAMs
   - What are quantile GAMs (QGAMs)
   - Using QGAMs with qgam and mgcViz
2. Loss-based checks
   - Loss-based vs goodness-of-fit checks
   - Future developments
3. Hands-on session

Loss-based checks

Visual checks in first session were looking at goodness-of-fit.

In a forecasting context we might be interested in predictive accuracy.

Accuracy quantified using loss function (MSE, pinball, ...).

`check1D`, `check2D` can be adapted to visualize **custom losses on training or training test set**.

Loss-based checks

```
check1D(o, x, type = "auto", trans = NULL, ...)
```

Here:

- if `type = "y"` we are looking at responses y, not residuals r
- `trans` is function used to transform y

Example of loss-based check:

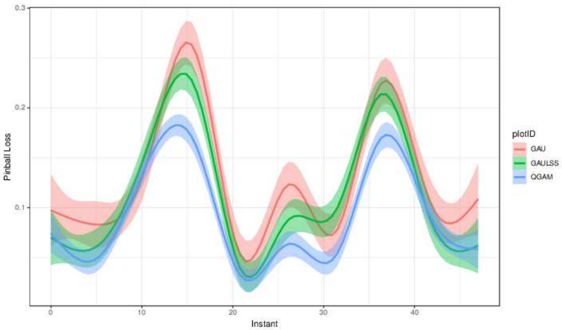
```
pl <- check1D(o = fit, x = x1, type = "y",
              trans = function(y, ...){
                      (y - fit$fitted.values)^2
                      })

pl + l_gridCheck1D(mean) + ...
```

Now we move to “qgam_demonstration.html”.

Future work

For predictive applications, it would be useful to be able to compare losses across models:

```
pl1 <- check1D(m1, function = pinball) + geom_smooth()
pl2 <- check1D(m2, function = pinball) + geom_smooth()
pl3 <- check1D(m3, function = pinball) + geom_smooth()
compare(pl1, pl2, pl3)
```



Future work

Focussing on the losses would be useful in the context of stacking

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmin}} \; \text{Loss}\Big\{ y_i - \sum_{j=1}^{m} w_j \hat{\mu}_j(\mathbf{x}_i) \Big\}$$

where

- $\hat{\mu}_j(\mathbf{x}_i)$ is prediction of j-th expert;
- $\mathbf{w} = \{w_1, \dots, w_m\} > 0$ are model weights.

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)*.

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

Fasiolo, M., R. Nedellec, Y. Goude, and S. N. Wood (2018). Scalable visualisation methods for modern generalized additive models. *arXiv preprint arXiv:1809.10632*.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*, Volume 43. CRC Press.

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.

Wood, S. N., N. Pya, and B. Säfken (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association 111*(516), 1548–1575.