More mgcViz tools

Matteo Fasiolo (University of Bristol, UK)

Joint work with:

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

matteo.fasiolo@bristol.ac.uk

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.

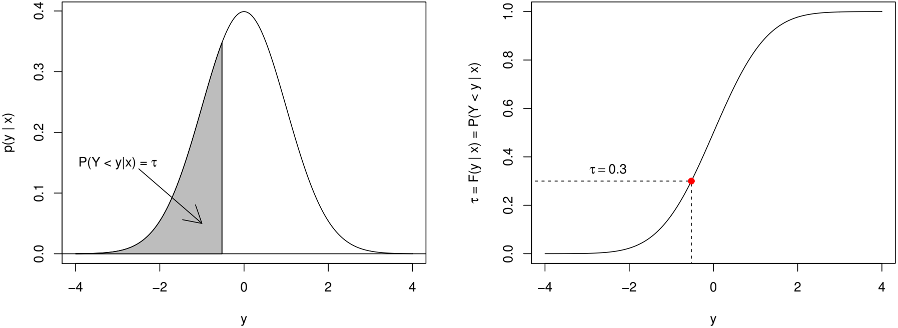
What is quantile regression

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

We consider continuous (not discrete) y.

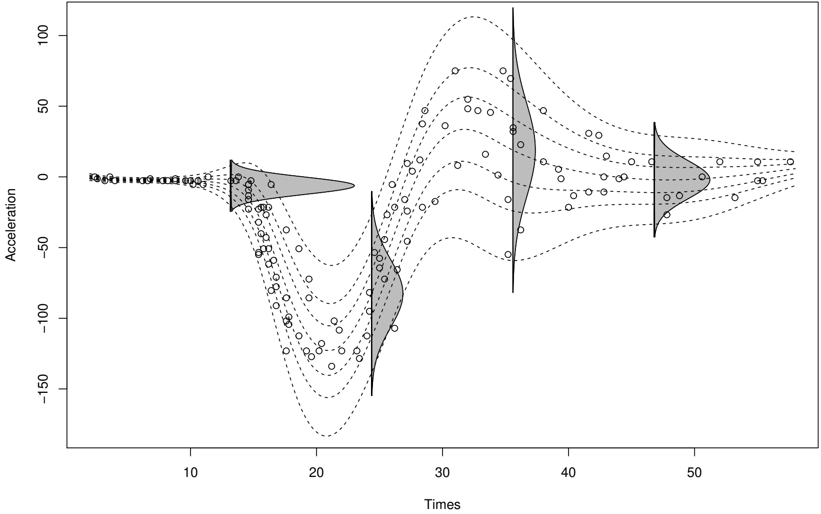
Define $F(y|\mathbf{x}) = \text{Prob}(Y \leq y|\mathbf{x})$.

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



What is quantile regression

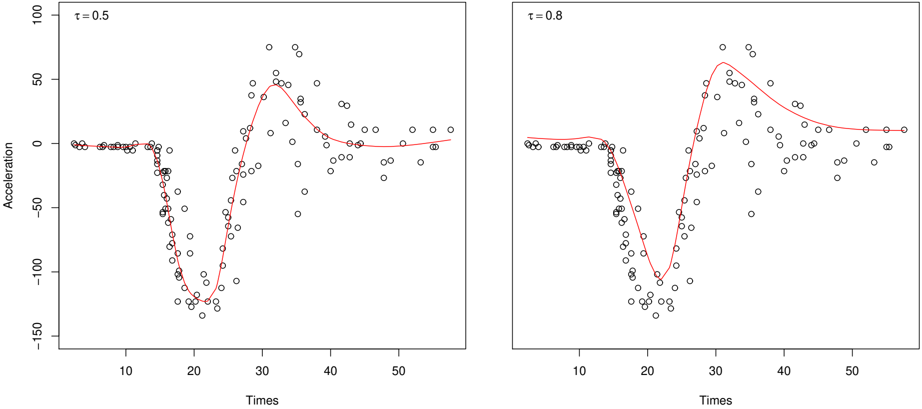
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



What is quantile regression

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

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



What is quantile regression

The τ-th quantile is

$$\mu = F^{-1}(\tau|\mathbf{x}),$$

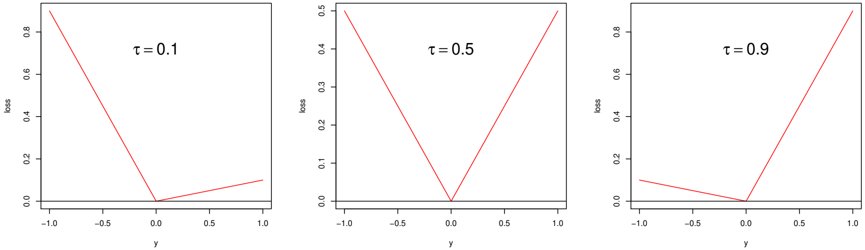
but also the minimizer of

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

where

$$\rho_\tau(z) = (\tau-1)z\mathbb{1}(z<0) + \tau z\mathbb{1}(z \geq 0),$$

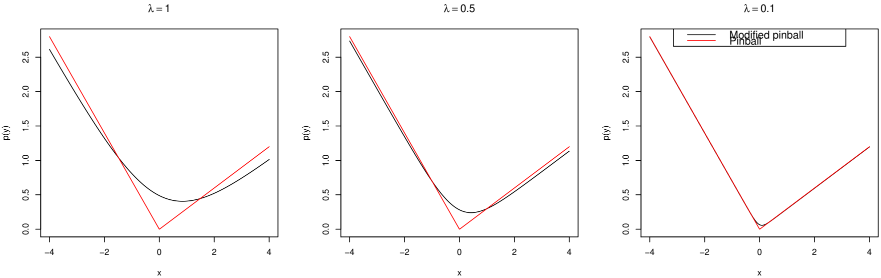
is the "pinball" loss (Koenker, 2005).



Smoothing the pinball loss

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.



NB in qgam, λ reparametrized as err $\in (0, 1)$ ($\downarrow$ err implies $\downarrow \lambda$).

Smoothing the pinball loss

Increasing `err` leads to:

- faster and more stable computation
- more bias

By default:

```
qgam(..., err = 0.05, ...)
```

which is a compromise between bias and speed.

`mgcViz` provides specific visualisations for QGAMs.

Now we move to "qgam_demonstration.html".

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:

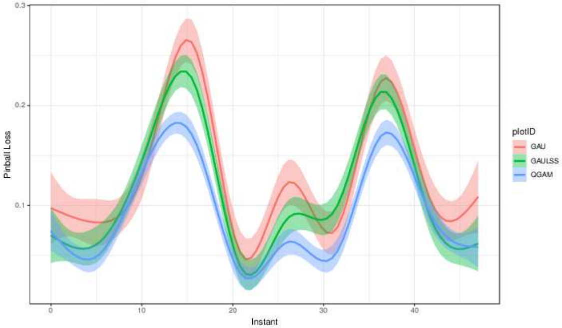
```
pl <- check1D(o = fit, x = x1, type = "y",
              trans = function(y, ...){
                      (y - fit$fitted.values)^2
                      })


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

It would be useful to have method to transform smooth effects:

```
d1 <- sm(fit, 1) %>% diff(1) # 1st derivative
d2 <- sm(fit, 1) %>% diff(2) # 2nd

plot(d1) + l_fitLine() + l_ciLine()
```

How to get confidence intervals on derivative?

Possibly useful: `tidyfun` R package.

References I

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*.

Jones, M. (2008). On a class of distributions with simple exponential tails. *Statistica Sinica 18*(3), 1101–1110.

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. (2006). *Generalized additive models: an introduction with R*. CRC press.