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
   - Bayesian predictive checks
3. Future developments & 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.

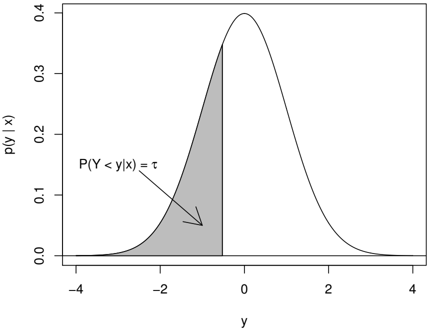
What is quantile regression

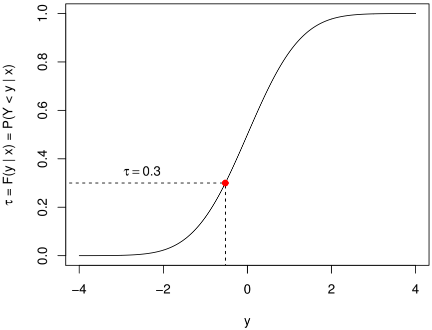
Lots of options for $p_m(y|\mathbf{x})$: binomial, gamma, Poisson, Tweedie...

We consider continuous (not discrete) y.

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

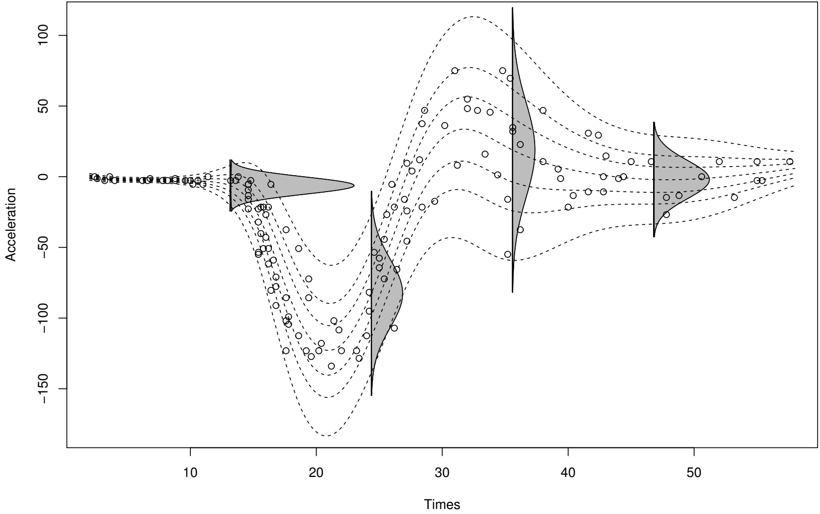
The τ-th ($\tau \in (0,1)$) quantile is $\mu_\tau(\mathbf{x}) = F^{-1}(\tau|\mathbf{x})$.





What is quantile regression

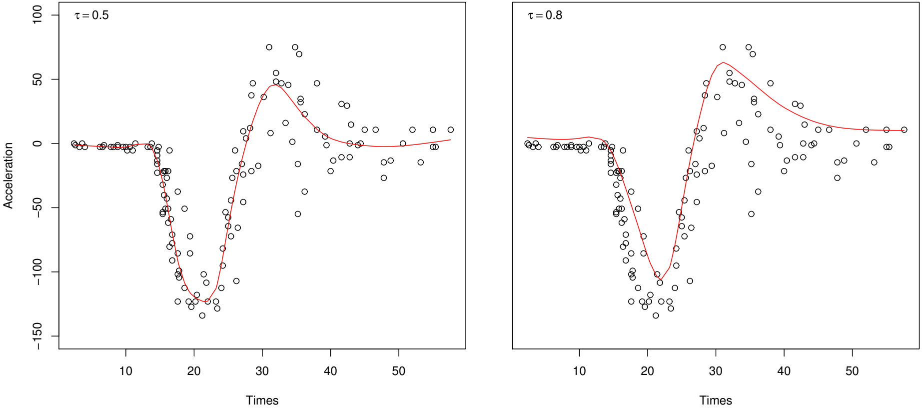
Given $p_m(y|\mathbf{x})$ we can get the conditional quantiles $\mu_\tau(\mathbf{x})$.



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

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

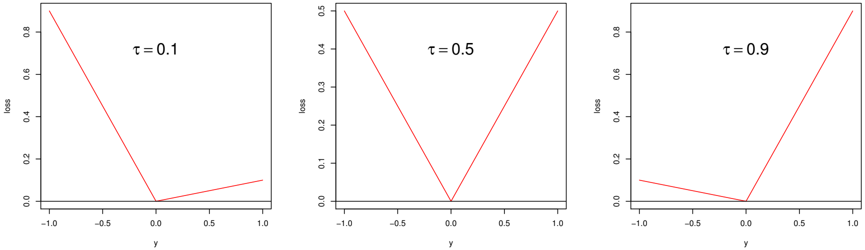
but also the minimizer of

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

where

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

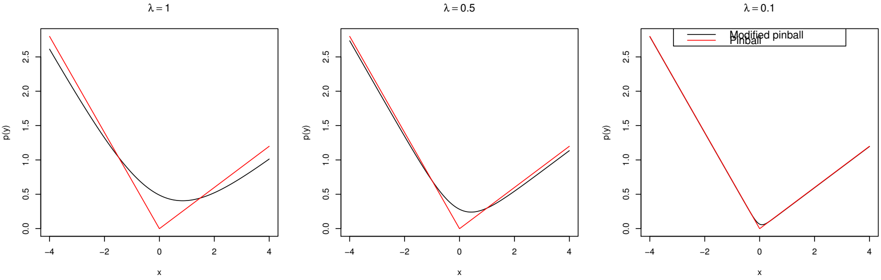
is the "pinball" loss (Koenker, 2005).



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 $\texttt{err} \in (0, 1)$ ($\downarrow \texttt{err}$ 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".

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.