Intro to generalized additive models in R (with mgcv)

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

Material available at:

`https://github.com/mfasiolo/GAM_Workshop_Dortmund_25`

Workshop content

Today's sessions will cover:

1. Intro to standard GAMs
2. Smooth effects and penalties
3. Multi-parameter GAMs, including GAMLSS
4. Quantile GAMs and Big Data methods

Focus on **GAM modelling**, not fitting/inferential/computational aspects.

On Github you can find:

1. slides
2. html files for R demos
3. exercises and solutions

We firstly cover:

1. What is an additive model?
2. Introducing smooth effects
3. Diagnostics and model selection tools
4. GAM modelling with mgcv

What is an additive model?

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 $\text{Dist}(y|\mathbf{x})$.

Model is $\text{Dist}_m\{y|\theta_1(\mathbf{x}), \theta_2, \dots, \theta_q\}$, where $\theta_1, \dots, \theta_q$ are param.

We assume that $\theta_2, \dots, \theta_q$ do not depend on $\mathbf{x}$.

Gaussian additive model:

$$y|\mathbf{x} \sim N\{y|\mu(\mathbf{x}), \sigma^2\},$$

where

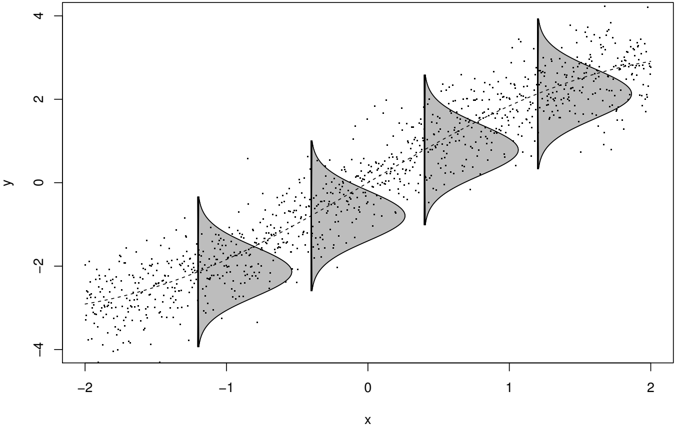
$$\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = \sum_{j=1}^{m} f_j(\mathbf{x}),$$

and

$$\sigma^2 = \mathrm{Var}(y|\mathbf{x}).$$

f_j's can be fixed, random or smooth effects.

NB: we call $\sum_{j=1}^{m} f_j(\mathbf{x})$ **linear predictor** because it is linear in $\boldsymbol{\beta}$.



Gaussian model with variable mean.
In `mgcv`: `gam(y~s(x), family=gaussian)`.

Generalized additive model (GAM) (Hastie and Tibshirani, 1990):

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

where

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

and g is a one-to-one link function.

Poisson GAM:

- $y|\mathbf{x} \sim \text{Pois}\{y|\mu(\mathbf{x})\}$
- $\mu(\mathbf{x}) = \exp\left\{\sum_{j=1}^{m} f_j(\mathbf{x})\right\}$
- $g = \log$ assures $\mu(\mathbf{x}) > 0$

Here relation between $\mathbb{E}(y|\mathbf{x})$ and $\text{Var}(y|\mathbf{x})$ is implied by model...

... or we can have extra parameters for scale and shape.

Scaled Student's t GAM:

- $y|\mathbf{x} \sim \text{ScaledStud}\{y|\mu(\mathbf{x}), \sigma, \nu\}$ (i.e. $(y - \mu)/\sigma \sim \text{Stud}(y|\nu)$)
- $\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = \sum_{j=1}^{m} f_j(\mathbf{x})$
- σ is scale parameter
- ν is shape parameter (degrees of freedom)
- $\text{Var}(y|\mathbf{x}) = \sigma^2 \frac{\nu}{\nu-2}$

Later we'll let all parameters be functions of **x**, eg:

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

These slides cover:

1. What is an additive model?
2. Introducing smooth effects
3. Diagnostics and model selection tools
4. GAM modelling with mgcv

Introducing smooth effects

Consider additive model

$$g\{\mu(\mathbf{x})\} = f_1(\mathbf{x}) + f_2(\mathbf{x}) + f_3(\mathbf{x}),$$

where

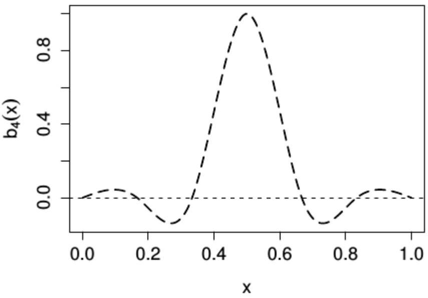
- $f_1(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$
- $f_2(\mathbf{x}) = \begin{cases} 0 & \text{if } x_2 = \text{FALSE} \\ \beta_4 & \text{if } x_2 = \text{TRUE} \end{cases}$
- $f_3(\mathbf{x}) = f_3(x_3)$ is a non-linear smooth function.

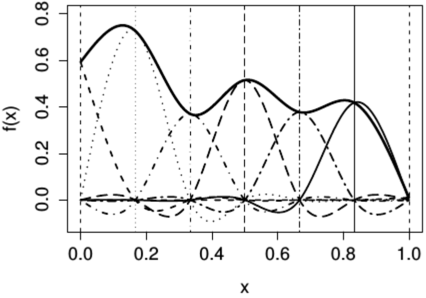
Smooth effects built using spline bases

$$f_3(x_3) = \sum_{k=1}^{r} \beta_k b_k(x_3)$$

where β_k are unknown coeff and $b_k(x_3)$ are known spline basis functions.

```
s(x, bs = "cr", k = 20)
```

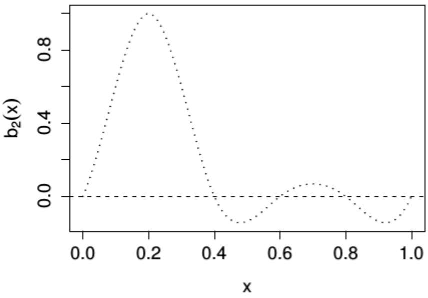


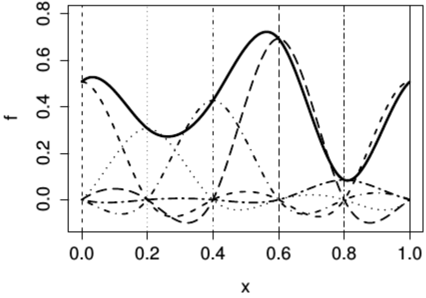


Cubic regression splines are related to the optimal solution to

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda \int f''(x)^2 \, dx.$$

`mgcv` offers many smooths (see `?smooth.terms`). `s(x, bs = "cc")`





Cyclic cubic regression splines make so that

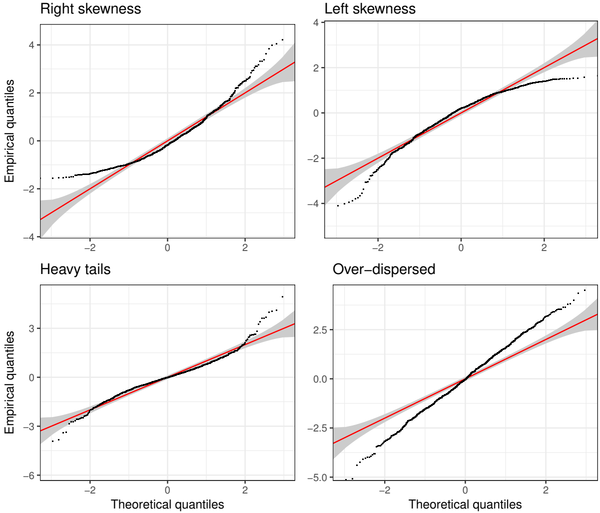
- $f(x_{min}) = f(x_{max})$
- $f'(x_{min}) = f'(x_{max})$
- $f''(x_{min}) = f''(x_{max})$

These slides cover:

1. What is an additive model?
2. Introducing smooth effects
3. Diagnostics and model selection tools
4. GAM modelling with mgcv

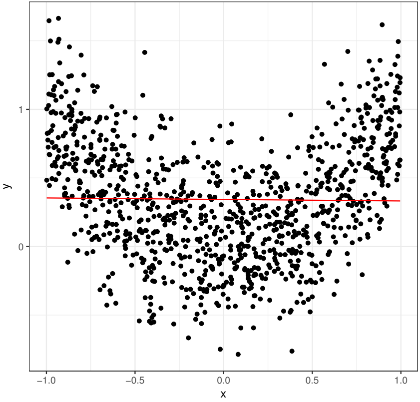
In the first hands-on session we'll use few basic diagnostics.

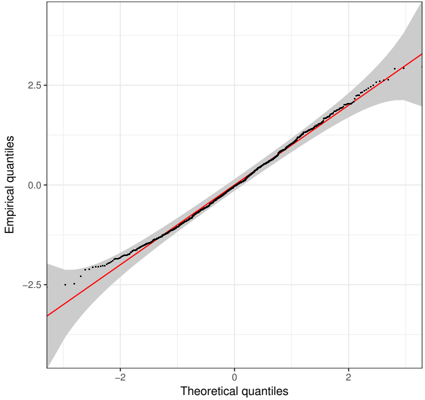
QQ-plots



Useful for choosing model $\text{Dist}_m(y|\mathbf{x})$ (e.g. Poisson vs Neg. Binom.)

Less useful for finding omitted variables and non-linearities.





Recall structure of smooth effects:

$$f(\mathbf{x}) = \sum_{j=1}^{k} \beta_j b_j(\mathbf{x}).$$

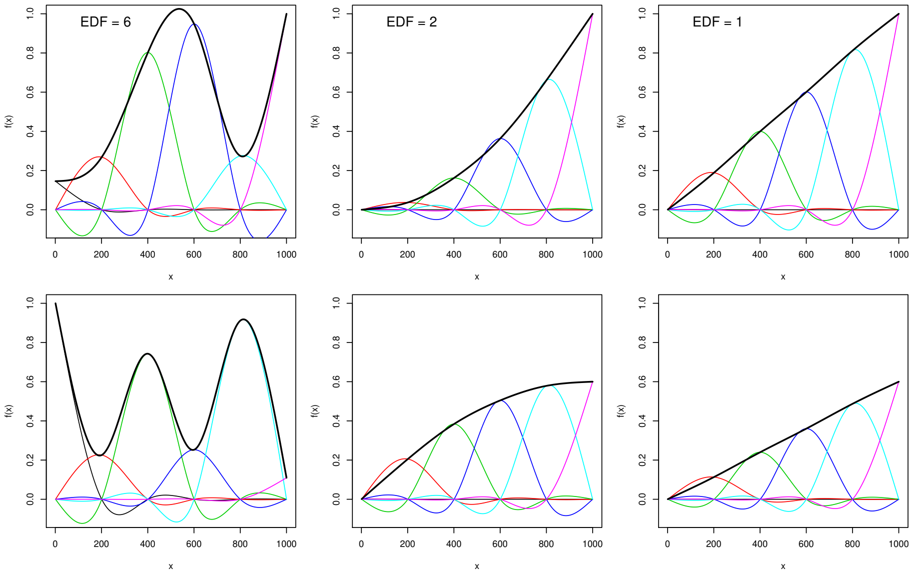
where $\boldsymbol{\beta}$ estimated by Maximum a Posterior (MAP)

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \ \log p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \Big\{ \overbrace{\log p(\mathbf{y}|\boldsymbol{\beta})}^{\text{goodness of fit}} + \underbrace{\log p(\boldsymbol{\beta}|\boldsymbol{\lambda})}_{\text{prior penalising complexity}} \Big\}$$

Exact k is unimportant, we choose it large enough and let penalty work.

Effective number of parameters we are using is $\leq k$.

Approximation is **Effective Degrees of Freedom** (EDF) $\leq k$.


EDF = 6
EDF = 2
EDF = 1
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000

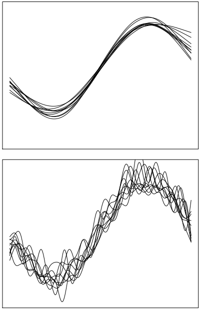
`mgcv` uses a hierarchical fitting framework:

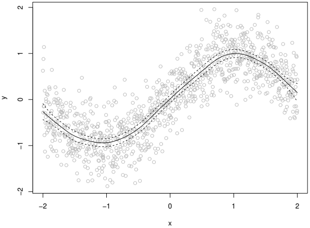
1. Select $\boldsymbol{\lambda}$ to determine smoothness

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\text{argmax}} \, \text{LAML}(\boldsymbol{\lambda}).$$

2. For fixed $\boldsymbol{\lambda}$, estimate $\boldsymbol{\beta}$ to determine actual fit

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





By default $k = 10$ but this is arbitrary.

Exact choice of k not important, but it must not be too low.

Checking whether k is too low:

1. look at conditional residuals checks
2. look at output of `gam.check(fit)`:

```
##              k'  edf k-index p-value
## s(x1)      9.00 8.60    0.91  <2e-16 ***
## s(x2)      9.00 8.13    1.02    0.76
## s(x3)      8.00 2.66    1.04    0.97
```

3. increase k and see if a **model selection criterion** improves

Popular criterion is approximate Akaike Information Criterion (AIC):

$$\text{AIC} = \underbrace{-2\log p(\mathbf{y}|\hat{\beta})}_{\text{goodness of fit}} \quad + \quad \underbrace{2\tau}_{\text{model complexity}}$$

where τ is EDF.

If $\text{AIC}_{m_1} < \text{AIC}_{m_2}$ choose model 1.

To select which effects to include we can also look at p-values:

```
summary(fit)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 267.2004    75.4197   3.543 0.000405 ***
## x1            6.2854     1.0457   6.010 2.20e-09 ***
## x2           79.8459    80.4130   0.993 0.320858
## x3          -71.2728    86.1725  -0.827 0.408284
```

The exercises will be based on the `mgcv` package for GAM modelling.

`mgcv` is a recommended R package, included in R by default.

It contains methods for:

- creating GAM models
- fitting them
- visualizing and summarizing model output

The `mgcv` ecosystem:

- `mgcViz` visualising GAMs
- `qgam` quantile GAMs
- `SCM` multivariate Gaussian GAMs
- `gamFactory` nested smooth effects in GAMs
- and many others `gamm4`, `refund`, `scam`, `vgam`, `GJRM`, `itsadug`, ...

There are alternatives to `mgcv`, such as:

- `mboost` (Hothorn et al., 2010)
- `gamlss` (Rigby and Stasinopoulos, 2005)
- `brms` (Bürkner et al., 2017)
- `BayesX` (Brezger et al., 2003)
- `INLA` (Rue et al., 2009)
- `bamlss` (Umlauf et al., 2018)

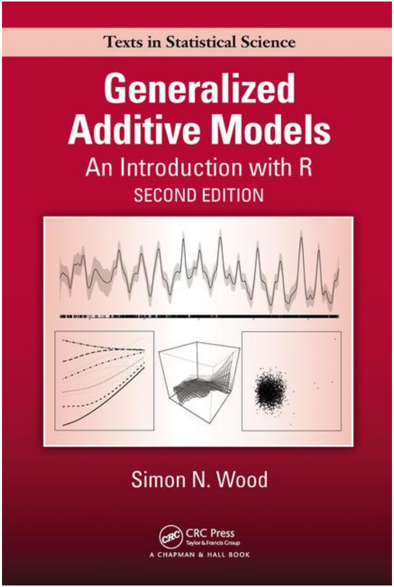
Each offers much flexibility (e.g. smooth effects types and distributions).

Strong points of `mgcv`'s methods:

1. little tuning needed (automatic smoothing parameters selection)
2. fast and stable numerical implementation

Now we'll be looking at R_demos/1_Intro_Gefcom14.html

Further reading



References I

Brezger, A., T. Kneib, and S. Lang (2003). Bayesx: Analysing bayesian structured additive regression models. Technical report, Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians ....

Bürkner, P. C. et al. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software 80*(1), 1–28.

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

Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner (2010). Model-based boosting 2.0. *The Journal of Machine Learning Research 11*, 2109–2113.

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.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology) 71*(2), 319–392.

Umlauf, N., N. Klein, and A. Zeileis (2018). BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics* **27**(3), 612–627.