Generalized additive models

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

June 27, 2018

Today's plan

Morning session

1. Intro to Generalized Additive Models (GAMs)
2. Smooth effect types & Big Data methods

Afternoon session

1. Beyond mean modelling: GAMLSS models
2. Distribution-free modelling: Quantile GAMs

Intro to Generalized Additive Models (GAMs)

Structure:

1. What is an additive model?
2. Introducing smooth effects
3. Introducing random effects
4. Diagnostics and model selection tools
5. GAM modelling using `mgcv` and `mgcViz`

Structure of the talk

Structure:

1. **What is an additive model?**
2. Introducing smooth effects
3. Introducing random effects
4. Diagnostics and model selection tools
5. GAM modelling using `mgcv` and `mgcViz`

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

In a Gaussian model, the mean depends on the covariates

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

where $\mu = \mathbb{E}(y|\mathbf{x})$ and $\sigma^2 = \text{Var}(y)$.

What is an additive model

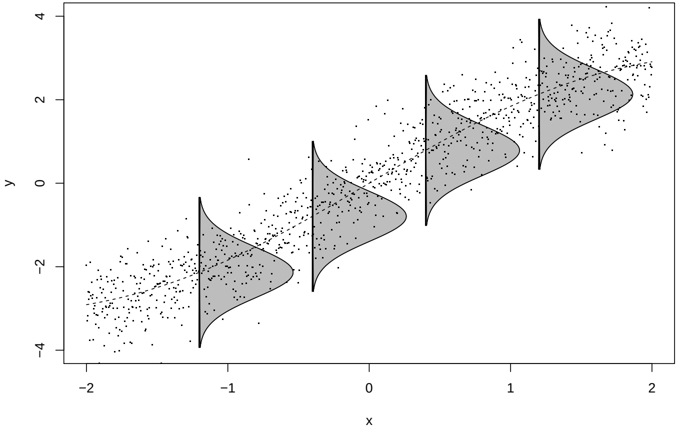


Figure : Gaussian model with variable mean.

In mgcv: `gam(y~s(x), family=gaussian)`.

What is an additive model

Gaussian additive model:

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

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

f_j's can be fixed, random or smooth effects with coefficients $\boldsymbol{\beta}$.

$\hat{\boldsymbol{\beta}}$ is the maximizer of **penalized** log-likelihood

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}} \big\{ L_y(\boldsymbol{\beta}) - \text{Pen}(\boldsymbol{\beta}) \big\}.$$

where:

- $L_y(\boldsymbol{\beta}) = \sum_i \log p(y_i|\boldsymbol{\beta})$ is log-likelihood
- $\text{Pen}(\boldsymbol{\beta})$ penalizes the complexity of the f_j's

What is an additive model

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

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

and g is the link function.

Poisson GAM:

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

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

What is an additive model

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

Scaled Student's t GAM:

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

In the afternoon will see models with multiple linear predictors, eg:

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

Structure of the talk

Structure:

1. What is an additive model?
2. **Introducing smooth effects**
3. Introducing random effects
4. Diagnostics and model selection tools
5. GAM modelling using `mgcv` and `mgcViz`

Introducing smooth effects

Consider additive model

$$\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = g^{-1}\Big\{f_1(\mathbf{x}) + f_2(\mathbf{x}) + f_3(\mathbf{x})\Big\},$$

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 spine 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.

Example 1: B-splines

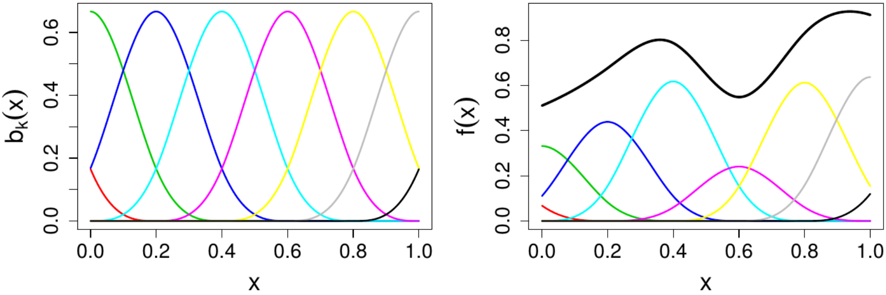


Figure : B-spline basis (left) and smooth (right).

Introducing smooth effects

Example 2: Thin plate regression splines (TPRS)

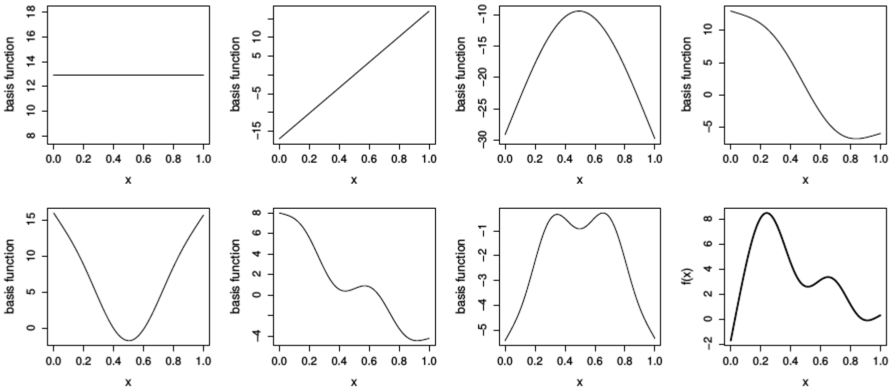


Figure : Rank 7 TPRS basis. Image from Wood (2006).

Introducing smooth effects

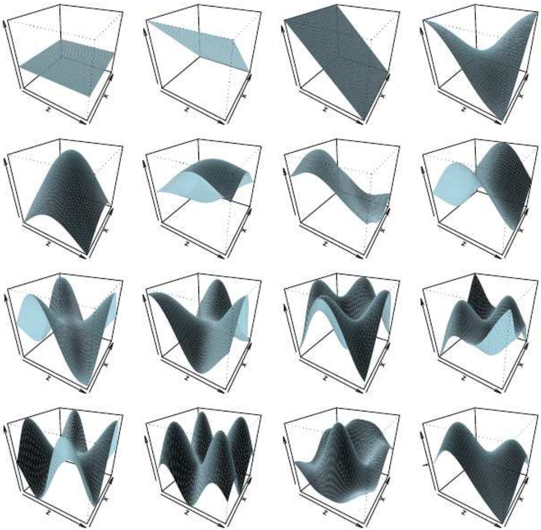


Figure : Rank 17 2D TPRS basis. Courtesy of Simon Wood.

Introducing smooth effects

In general

$$f(\mathbf{x}) = \sum_{k=1}^{r} \beta_k b_k(\mathbf{x}).$$

To determine complexity of $f(\mathbf{x})$:

- the basis rank r is large enough for sufficient flexibility
- a complexity penalty on $\boldsymbol{\beta}$ controls the wiggliness of the effects

In first morning practical we'll see only 1D effects.

In `mgcv`:

```
gam(y ~ 1 + x0 + s(x1, bs = "tp", k = 15), ...)
```

Structure of the talk

Structure:

1. What is an additive model?
2. Introducing smooth effects
3. **Introducing random effects**
4. Diagnostics and model selection tools
5. GAM modelling using `mgcv` and `mgcViz`

Introducing random effects

Suppose we have data on bone mineral density (*bmd*) as a function of *age*.

We have m subjects and n data pairs per subject

- subj 1: $\{bmd_{11}, age_{11}\}, \dots, \{bmd_{n1}, age_{n1}\}$
- subj j: $\{bmd_{1j}, age_{1j}\}, \dots, \{bmd_{nj}, age_{nj}\}$
- subj m: $\{bmd_{1m}, age_{1m}\}, \dots, \{bmd_{nm}, age_{nm}\}$

Standard linear model ignores individual differences

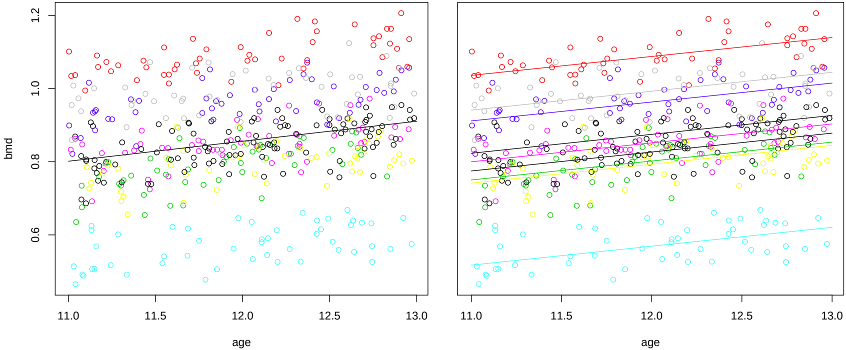
$$\mathbb{E}(bmd|age_{ij}) = \mu(age_{ij}) = \alpha + \beta age_{ij}.$$

We can include random intercept per subject

$$\mu(age_{ij}) = \alpha + \beta age_{ij} + \mathrm{a}_j,$$

where $\mathbf{a} = \{\mathrm{a}_1, \dots, \mathrm{a}_m\} \sim N(\mathbf{0}, \mathbf{\Sigma})$.

Introducing random effects



We can also include random slopes

$$\mu(age_{ij}) = \alpha + (\beta + b_j)age_{ij} + a_j,$$

where $\mathbf{a} \sim N(\mathbf{0}, \mathbf{\Sigma_a})$ and $\mathbf{b} \sim N(\mathbf{0}, \mathbf{\Sigma_b})$.

Introducing random effects

In `mgcv` random effect are specified as:

```
gam(bmd ~ 1 + s(subject, bs = "re") +
          age + s(age, subject, bs = "re"), ...)
```

In simplest case $\mathbf{\Sigma_a} = \gamma_\mathbf{a}\mathbf{I}$ and $\mathbf{\Sigma_b} = \gamma_\mathbf{b}\mathbf{I}$, that is

$$\mathbf{\Sigma_a} = \begin{bmatrix} \gamma_\mathbf{a} & 0 & 0 & \dots & 0 \\ 0 & \gamma_\mathbf{a} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \gamma_\mathbf{a} \end{bmatrix}$$

Variances $\gamma_\mathbf{a}$ and $\gamma_\mathbf{b}$ must be estimated (later I'll explain how).

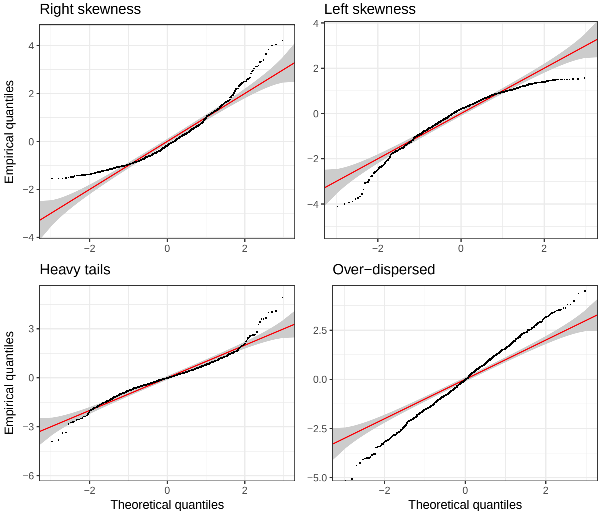
Structure of the talk

Structure:

1. What is an additive model?
2. Introducing smooth effects
3. Introducing random effects
4. **Diagnostics and model selection tools**
5. GAM modelling using `mgcv` and `mgcViz`

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

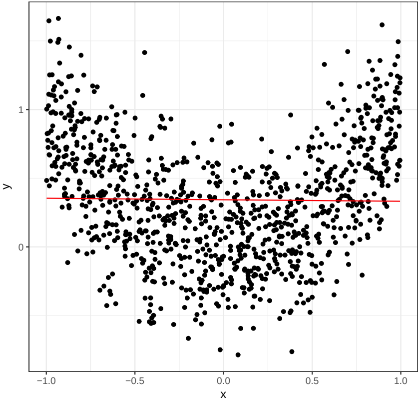
QQ-plots

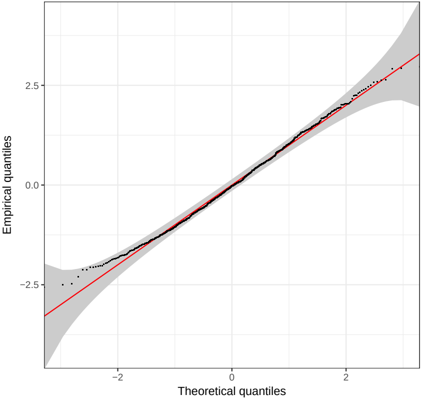


Diagnostics and model selection tools

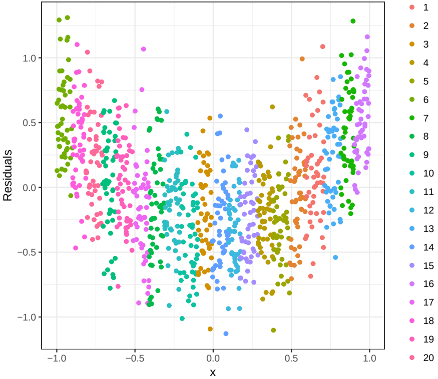
Useful for choosing model $\text{Dist}_m(y|\mathbf{x})$ (e.g. Poisson vs Neg. Binom.)

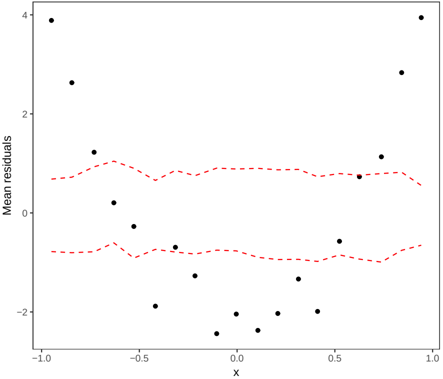
Less useful for finding omitted variables and non-linearities.





Conditional residuals checks are more helpful here.





Diagnostics and model selection tools

Recall structure of smooth effects:

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

where $\boldsymbol{\beta}$ shrunk toward zero by smoothness penalty.

Effective number of parameters we are using is $< k$.

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

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

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

Diagnostics and model selection tools

Checking whether *k* is too low:

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

```
##                k'  edf k-index p-value
## s(wM)        9.00 8.60    0.91  <2e-16 ***
## s(wM_s95) 9.00 8.13    1.02    0.76
## s(Posan)  8.00 2.66    1.04    0.97
```

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

Model selection

General criterion is approximate Akaike Information Criterion (AIC):

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

where τ is EDF.

If $\text{AIC}_{m1} < \text{AIC}_{m2}$ 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 ***
## Fl            6.2854     1.0457   6.010 2.20e-09 ***
## loc2         79.8459    80.4130   0.993 0.320858
## loc3        -71.2728    86.1725  -0.827 0.408284
```

Structure of the talk

Structure:

1. What is an additive model?
2. Introducing smooth effects
3. Introducing random effects
4. Diagnostics and model selection tools
5. **GAM modelling using mgcv and mgcViz**

GAMs in mgcv and mgcViz

`mgcv` is the recommended R package for fitting GAMs.

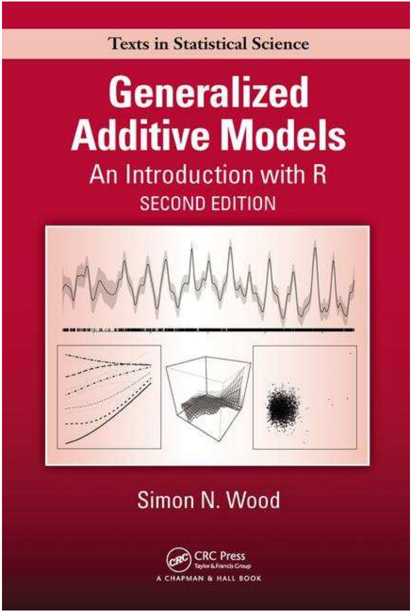
Today we'll work with `mgcViz`'s interface.

`mgcViz` extends `mgcv`'s tools for:

- plotting the estimated effects
- doing visual model checking

But most of the computation is done by `mgcv` under the hood.

Further reading



References I

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

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