Beyond mean modelling: GAMLSS models

Matteo Fasiolo

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/workshop_WARSAW19

These slides cover:

1. What are GAMLSS models?
2. When are they useful?
3. GAMLSS modelling in mgcv and mgcViz

Intro to GAMLSS models

Recall GAM model structure:

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

where

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

and g is the link function.

Example, Scaled Student-t distribution:

- location $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$
- scale $\theta_2 = \sigma$
- shape $\theta_3 = \nu$

Intro to GAMLSS models

In Generalized Additive Models for Location Scale and Shape (GAMLSS) (Rigby and Stasinopoulos, 2005) we let scale and shape change with the covariates $\mathbf{x}$.

GAMLSS model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu_1(\mathbf{x}), \theta_2 = \mu_2(\mathbf{x}), \ldots, \theta_p = \mu_p(\mathbf{x})\},$$

where

$$\mu_1(\mathbf{x}) = g_1^{-1}\Big\{\sum_{j=1}^{m} f_j^1(\mathbf{x})\Big\},$$

$$\ldots$$

$$\mu_p(\mathbf{x}) = g_p^{-1}\Big\{\sum_{j=1}^{m} f_j^p(\mathbf{x})\Big\},$$

and $g_1, \ldots, g_p$ are link function.

Intro to GAMLSS models

Example: **Gaussian model for location and scale**

Model is

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

where

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

$$\text{var}(y|\mathbf{x})^{1/2} = \sigma(\mathbf{x}) = \exp\left\{ \sum_{j=1}^{m} f_j^2(\mathbf{x}) \right\}$$

that is $g_2 = \log$ to guarantee $\sigma > 0$.

Intro to GAMLSS models

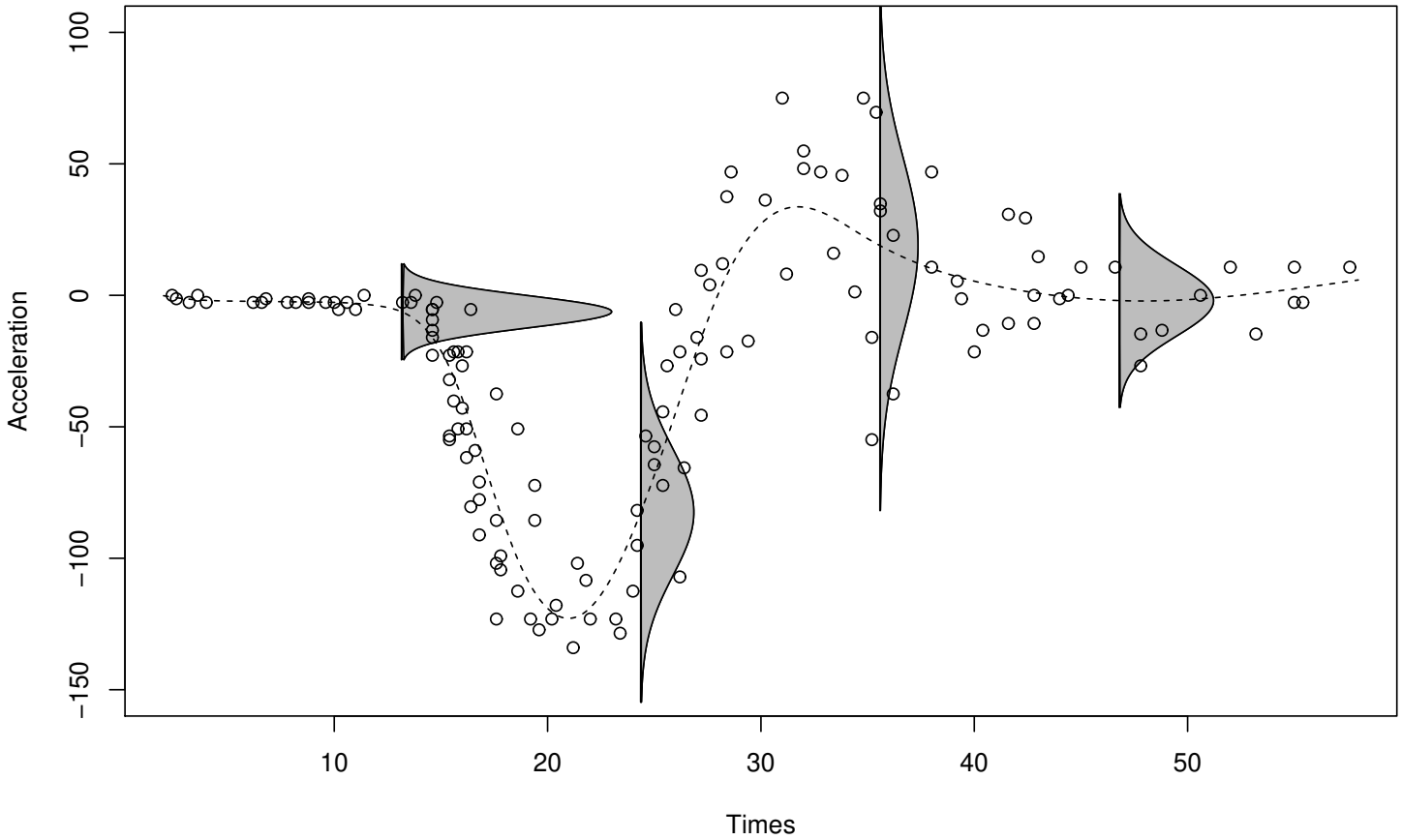
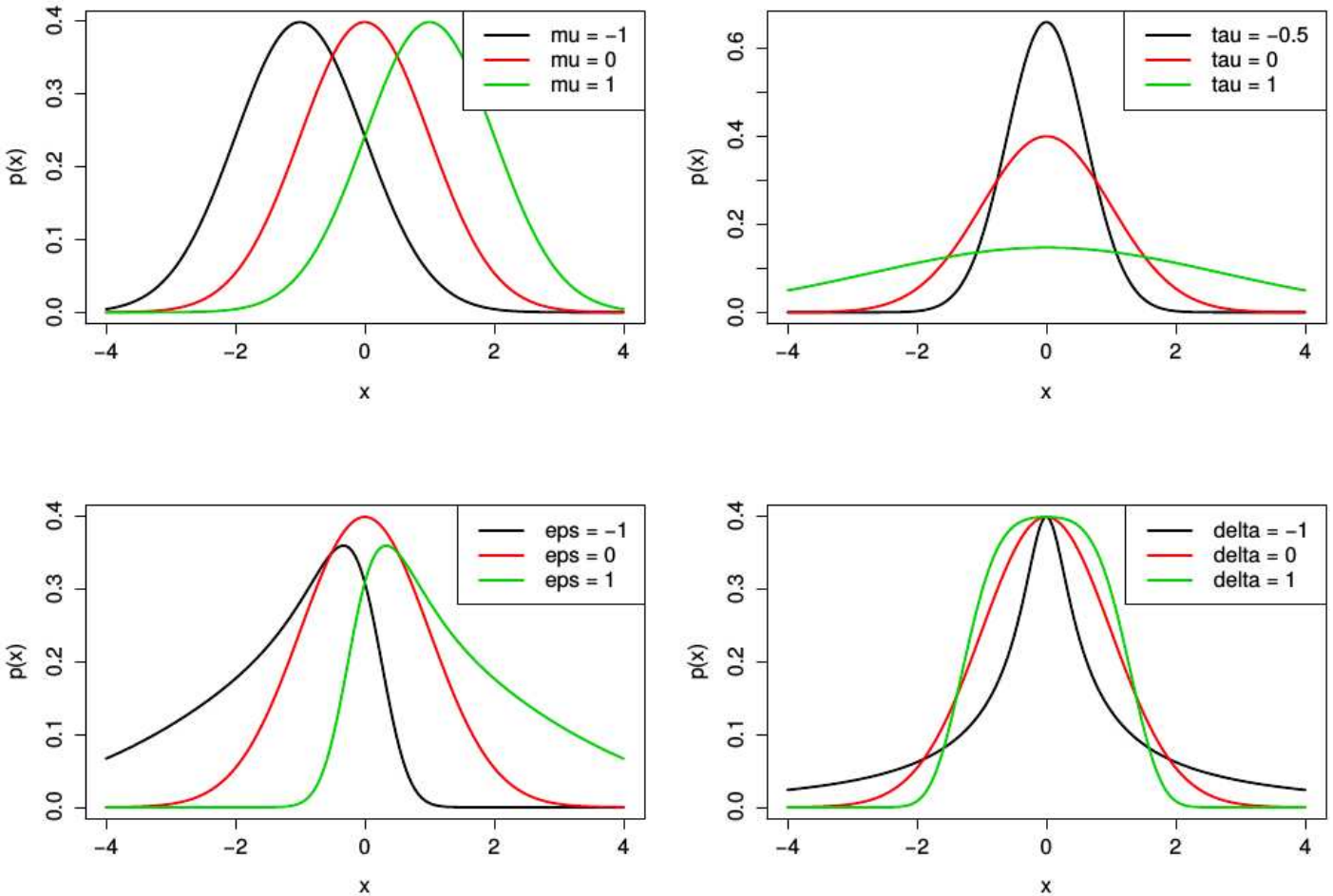


Figure: Gaussian model with variable mean and variance.
In `mgcv`: `gam(list(y~s(x), ~s(x)), family=gaulss)`.

Example: **Sinh-arcsinh (shash) distribution**

Four parameter distribution where location, scale, skewness (asymmetry) and kurtosis (tail behaviour) can depend on **x** (Jones and Pewsey, 2009).



Intro to GAMLSS models

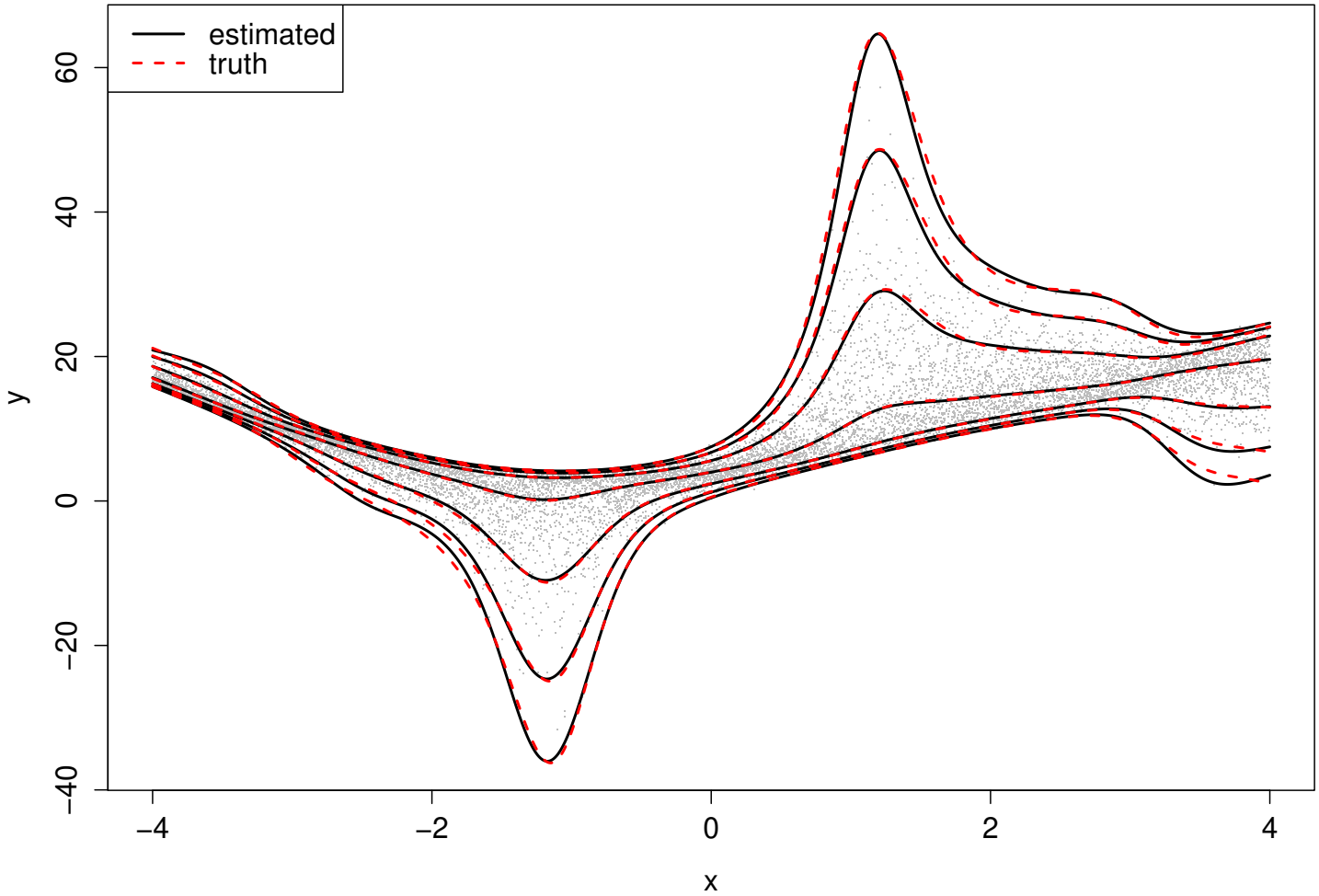


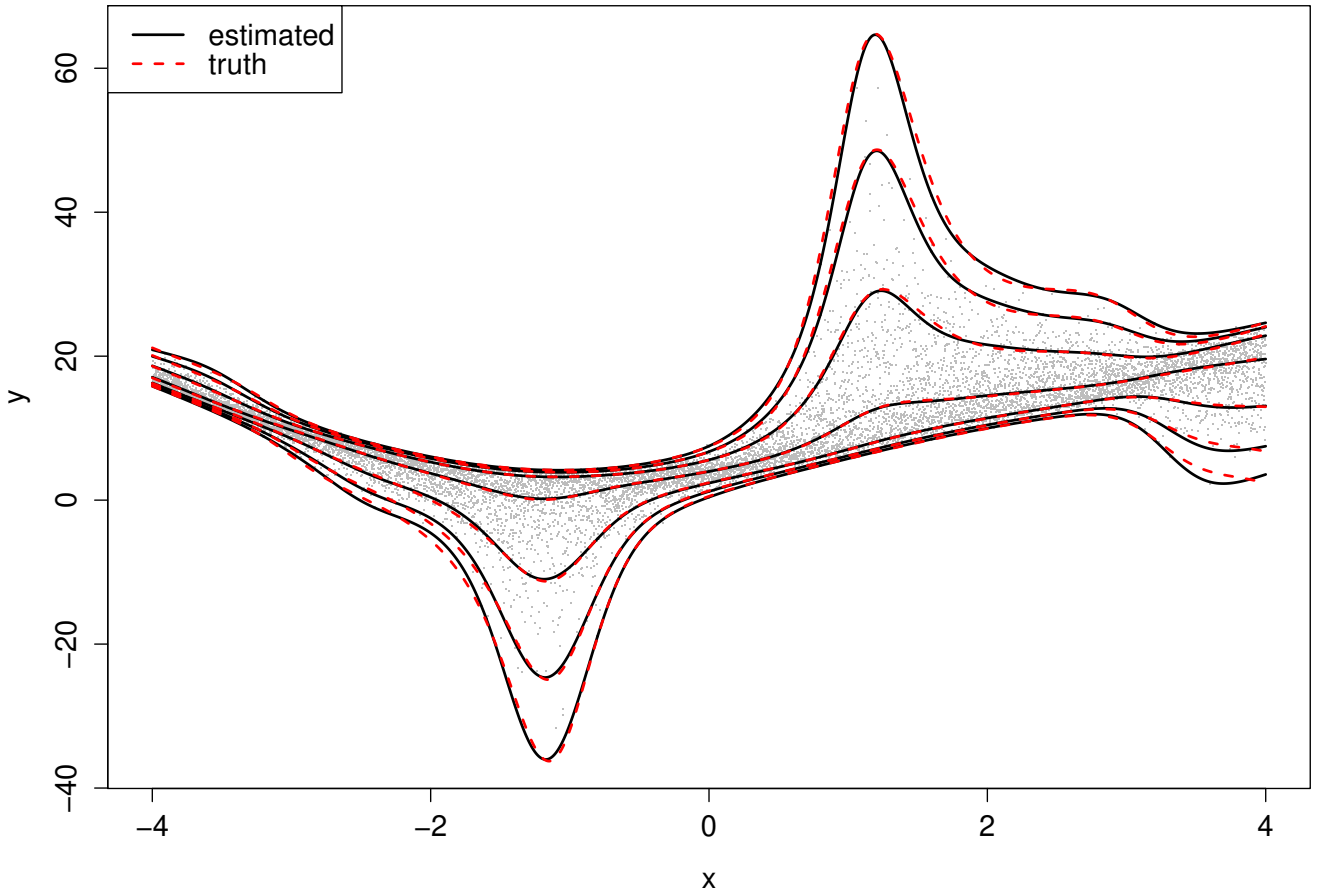
Figure: `gam(list(y~s(x), ~s(x), ~s(x), ~s(x)), family=shash)`.

These slides cover:

1. What are GAMLSS models?
2. When are they useful?
3. GAMLSS modelling in mgcv and mgcViz

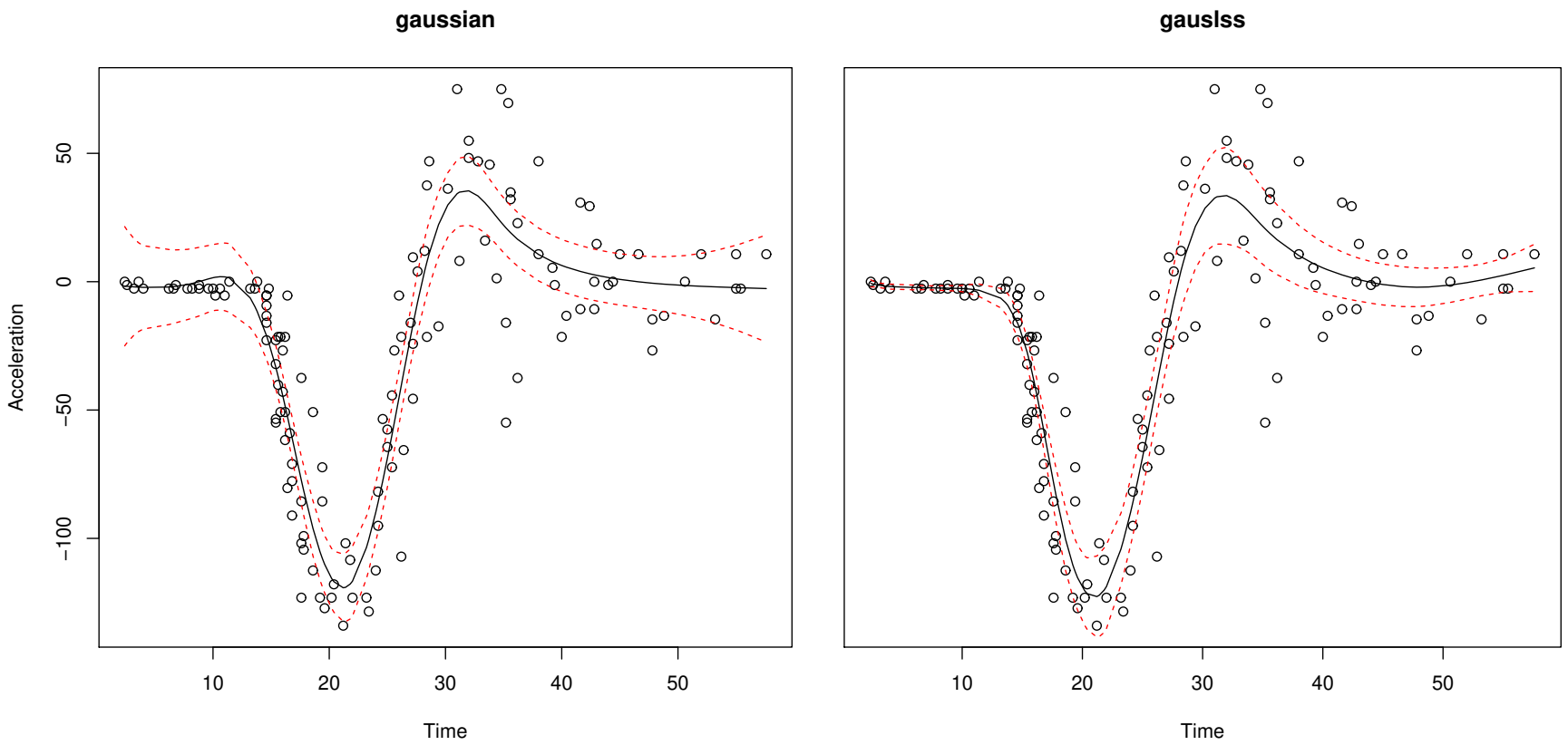
When are they useful?

R1: you might be interested in whole distribution $y|\mathbf{x}$ not just $\mathbb{E}(y|\mathbf{x})$.



When are they useful?

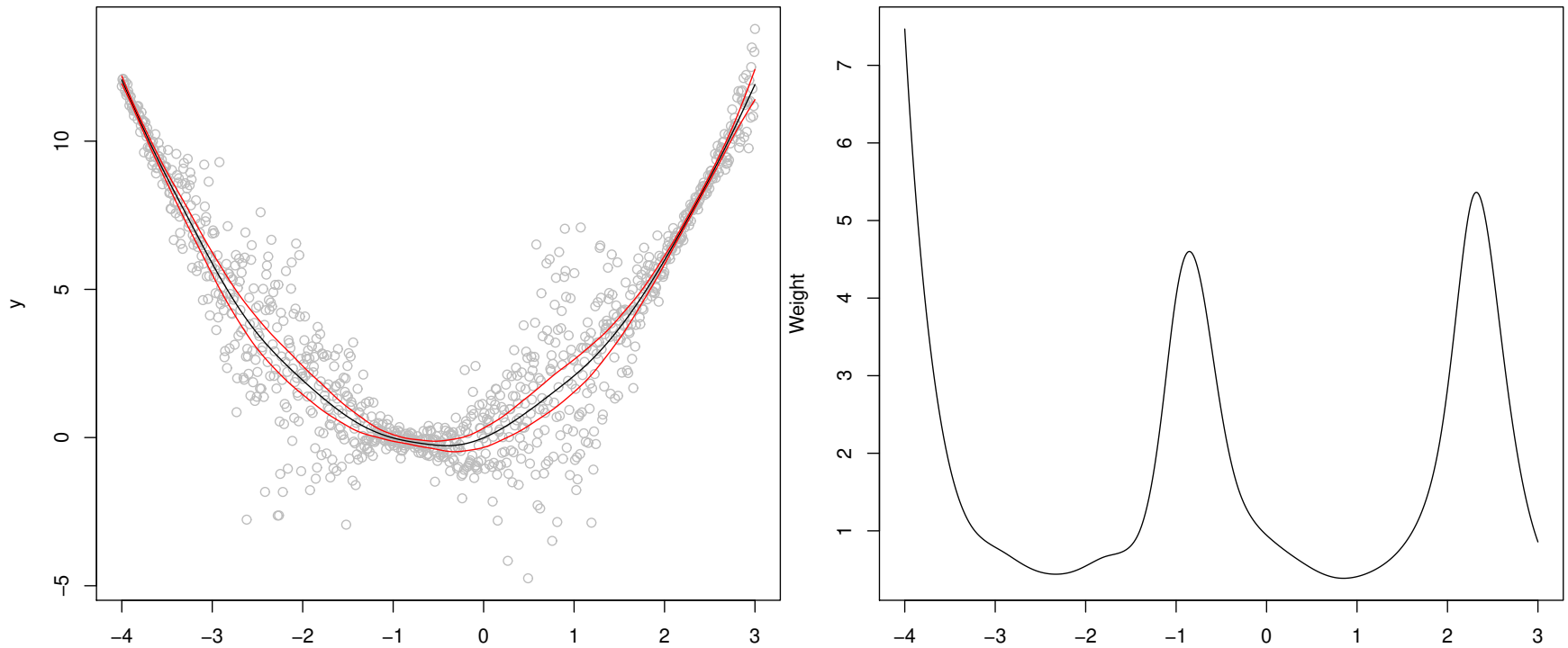
R2: standard GAM inference (e.g. p-value & confidence interval) is valid if the model for $p(y|\mathbf{x})$ is correct



When are they useful?

R3: the accuracy of the fit is improved if the weight of each observation is inversely proportional to $\mathrm{Var}(y|\mathbf{x})$, e.g. consider

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{(y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta})^2}{\sigma(\mathbf{x}_i)^2}.$$



These slides cover:

1. What are GAMLSS models?
2. When are they useful?
3. GAMLSS modelling in mgcv and mgcViz

References I

Jones, M. and A. Pewsey (2009). Sinh-arcsinh distributions. *Biometrika 96*(4), 761–780.

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.