gamlss

for statistical modelling

Generalized additive models for location scale and shape

Bob Rigby

Mikis Stasinopoulos

LONDON metropolitan university

Summary of GAMLSS course

Monday May 7th 2007

10.00-10.45	Ch1 Introduction to GAMLSS modelling
11.00-11.45	Ch2 Introduction to GAMLSS package in R
11.45-12.45	Practical 1
14.00-14.45	Ch3 Discrete distributions and regressions models
14.45-15.30	Ch4 Continuous distributions and regression models
16.00-17.00	Practical 2

Summary of GAMLSS course

Tuesday May 8th 2007

10.00-10.45	Ch5 Finite mixture distributions and models
11.00-11.45	Ch6 Smoothing models
11.45-12.45	Practical 3
14.00-14.45	Ch7 Regression model selection
14.45-15.30	Ch8 Nonlinear models
	Ch9 Random effects at the observation level
16.00-17.00	Practical 4

Summary of GAMLSS course

Wednesday May 9th 2007

10.00-10.45	Ch10 Random effects at the factor level
	Ch11 General random effects models
11.00-11.45	Ch6 Centile estimation and diagnostics
11.45-12.45	Practical 3
14.00-14.45	Discussion and feedback
	Analysing your own data sets

gamlss
for statistical modelling

Ch1 Introduction to GAMLSS modelling

1.1 Introduction to statistical modelling
1.2 Motivating data examples
1.3 GAMLSS model
1.4 Population distributions for Y
1.5 Additive terms
1.6 Model estimation
1.7 Algorithms
1.8 Residuals
1.9 Conclusion

gamlss
for statistical modelling

1.1 Introduction to statistical modelling

Statistical modelling:

Model fitting,

Hypothesis testing,

Model selection,

Model diagnostic checking, plots and statistics,

Prediction.

Objective:

If hypothesis testing, use formal statistical tests.

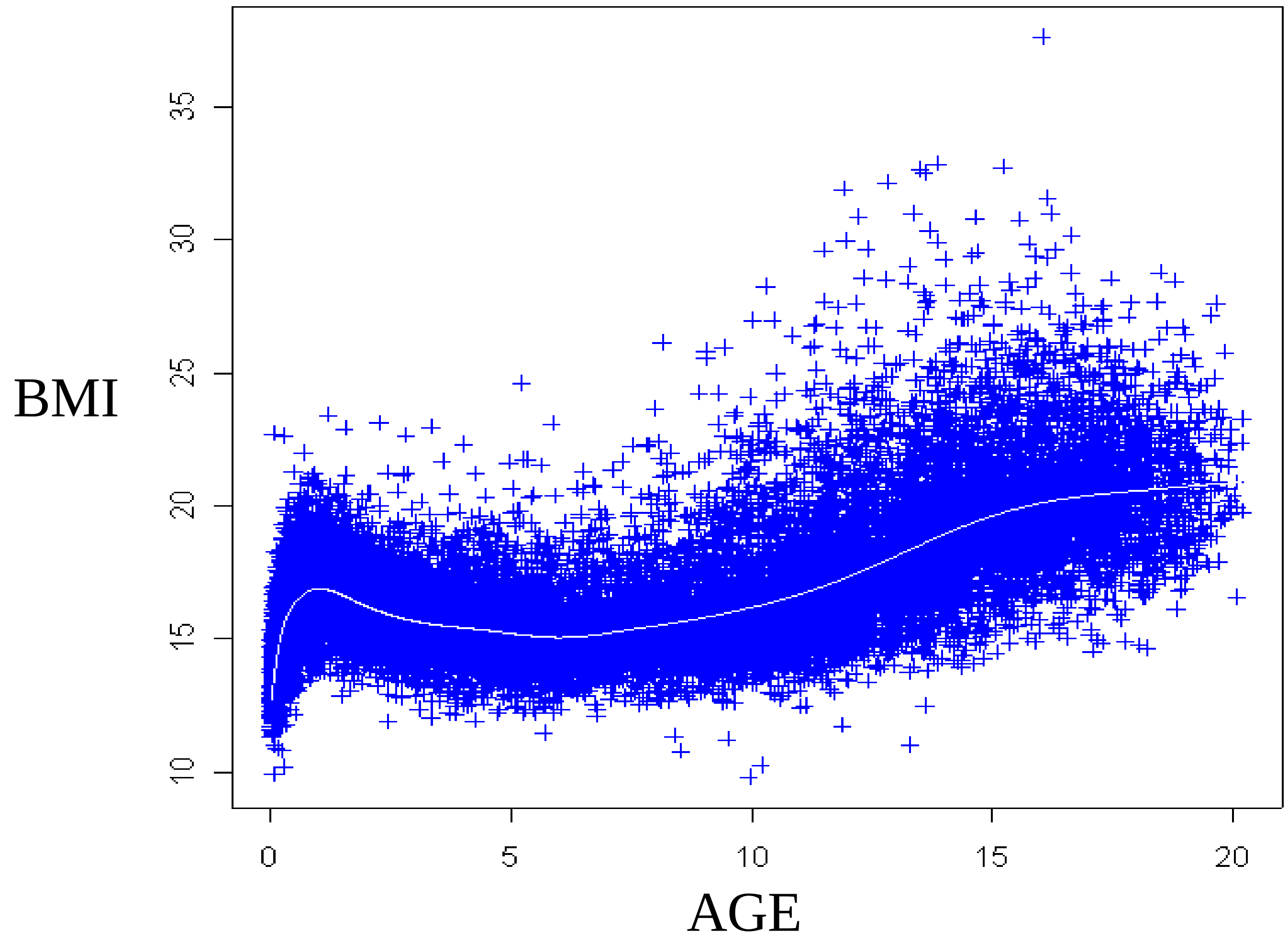
If prediction, use model selection criteria, e.g. AIC.

Why use gamlss?

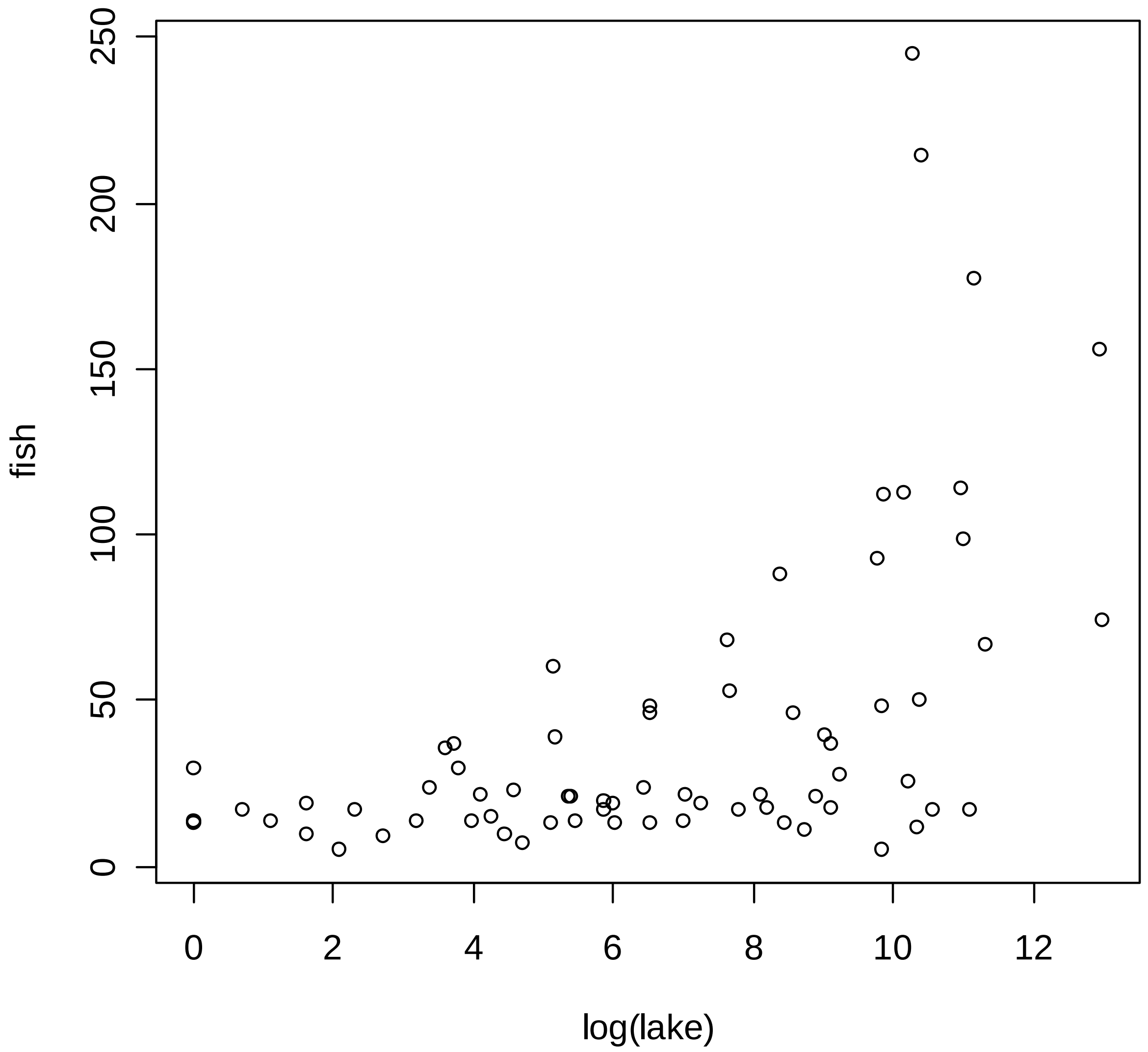
- Unified framework for univariate regression type of models
- Allows any distribution for the response variable Y
- Models all the parameters of the distribution of Y
- Allows a variety of additive terms in the models for the distribution parameters
- The fitted algorithm is modular, where different components can be added easily
- Models can be fitted easily and fast
- Explanatory tool to find appropriate set of models
- It deals with overdispersion, skewness and kurtosis

1.2 Motivating Examples:

BMI against AGE for Dutch girls



Example 2: The fish species data



1.3 GAMLSS model

$Y \sim D(\mu, \sigma, \nu, \tau)$ where D is any distribution and

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1}\boldsymbol{\gamma}_{j1}$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2}\boldsymbol{\gamma}_{j2}$$

$$g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3}\boldsymbol{\gamma}_{j3}$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4}\boldsymbol{\gamma}_{j4}.$$

Known link

Predictor

Linear terms

Additive terms

Random effects

Here $\gamma_{jk} \sim N_{q_{jk}}\left(0, G_{jk}^{-1}\right)$ and $G_{jk} = G_{jk}(\lambda)$

gamlss
for statistical modelling

Semi-parametric GAMLSS model

$Y \sim D(\mu, \sigma, \nu, \tau)$ where D is any distribution and

$$g_1(\mu) = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}(\mathbf{x}_{j1})$$

$$g_2(\sigma) = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} h_{j2}(\mathbf{x}_{j2})$$

$$g_3(\nu) = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} h_{j3}(\mathbf{x}_{j3})$$

$$g_4(\tau) = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}(\mathbf{x}_{j4}).$$

Nonlinear semi-parametric GAMLSS model

$Y \sim D(\mu, \sigma, \nu, \tau)$ where D is any distribution and

$$
\begin{aligned}
g_1(\mu) &= h_1(\mathbf{X}_1, \boldsymbol{\beta}_1) + \sum_{j=1}^{J_1} h_{j1}(\mathbf{x}_{j1}) \\
g_2(\sigma) &= h_2(\mathbf{X}_2, \boldsymbol{\beta}_2) + \sum_{j=1}^{J_2} h_{j2}(\mathbf{x}_{j2}) \\
g_3(\nu) &= h_3(\mathbf{X}_3, \boldsymbol{\beta}_3) + \sum_{j=1}^{J_3} h_{j3}(\mathbf{x}_{j3}) \\
g_4(\tau) &= h_4(\mathbf{X}_4, \boldsymbol{\beta}_4) + \sum_{j=1}^{J_4} h_{j4}(\mathbf{x}_{j4}).
\end{aligned}
$$

gamlss
for statistical modelling

Nonlinear parametric GAMLSS model

$Y \sim D(\mu, \sigma, \nu, \tau)$ where D is any distribution and

$$
\begin{aligned}
g_1(\mu) &= h_1(\mathbf{X}_1, \boldsymbol{\beta}_1) \\
g_2(\sigma) &= h_2(\mathbf{X}_2, \boldsymbol{\beta}_2) \\
g_3(\nu) &= h_3(\mathbf{X}_3, \boldsymbol{\beta}_3) \\
g_4(\tau) &= h_4(\mathbf{X}_4, \boldsymbol{\beta}_4)
\end{aligned}
$$

Parametric GAMLSS model

Y ~ *D*(μ, σ, ν, τ) where *D* is any distribution and

$$g_1(\mu) = \mathbf{X}_1\beta_1$$

$$g_2(\sigma) = \mathbf{X}_2\beta_2$$

$$g_3(\nu) = \mathbf{X}_3\beta_3$$

$$g_4(\tau) = \mathbf{X}_4\beta_4$$

1.4 Population distributions for *Y*

1.4.1 General comments

1) A wide range of discrete and continuous distributions implemented, including highly skew and kurtotic distributions

2) Easy implementation of new distributions

3) Different parameterisations of a distribution can be implemented

4) Truncated distributions and censored data easily implemented

gamlss
for statistical modelling

Table 1.2: Implemented GAMLSS distributions (with default link functions)

Distributions	R Name	μ	σ	ν	τ
Beta	BE()	logit	logit	-	-
Beta Binomial	BB()	logit	log	-	-
Beta Inflated (at 0)	BEOI()	logit	log	logit	-
Beta Inflated (at 1)	BEZI()	logit	log	logit	-
Beta Inflated (at 0 and 1)	BEINF()	logit	logit	log	log
Binomial	BI()	logit	-	-	-
Box-Cox Cole and Green	BCCG()	identity	log	identity	-
Box-Cox Power Exponential	BCPE()	identity	log	identity	log
Box-Cox-t	BCT()	identity	log	identity	log
Delaporte	DEL()	log	log	logit	-
Exponential	EXP()	log	-	-	-
Exponential Gaussian	exGAUS()	identity	log	log	-
Gamma	GA()	log	log	-	-
Generalized Gamma	GG()	log	log	identity	-
Generalized Inverse Gaussian	GIG()	log	log	identity	-
Gumbel	GU()	identity	log	-	-
Inverse Gaussian	IG()	log	log	-	-
Johnson's SU (μ the mean)	JSU()	identity	log	identity	log
Johnson's original SU	JSUo()	identity	log	identity	log
Logistic	LO()	identity	log	-	-
Log Normal	LOGNO()	log	log	-	-
Log Normal (Box-Cox)	LNO()	log	log	fixed	-
Multinomial	MULTIN()	log	log	log	log
Negative Binomial type I	NBI()	log	log	-	-
Negative Binomial type II	NBII()	log	log	-	-
NET	NET()	identity	log	fixed	fixed
Normal	NO()	identity	log	-	-
Normal family	NOF()	identity	log	identity	-
Poisson	PO()	log	-	-	-
Poisson inverse Gaussian	PIG()	log	log	-	-
Power Exponential	PE()	identity	log	log	-
Reverse Generalized Extreme	RGE()	identity	log	log	-
Reverse Gumbel	RG()	identity	log	-	-
Skew Power Exponential	SEP()	identity	log	identity	log
Shash	SHASH()	identity	log	log	log
Sichel	SI()	log	log	identity	-
Sichel (μ the mean)	SICHEL()	log	log	identity	-
Skew t type 3	ST3()	identity	log	identity	log
t Family	TF()	identity	log	log	-
Weibull	WEI()	log	log	-	-
Weibull (PH)	WEI2()	log	log	-	-
Weibull (μ the mean)	WEI3()	log	log	-	-
Zero inflated poisson	ZIP	log	logit	-	-
ZIP (μ the mean)	ZIP2	log	logit	-	-
Zero adjusted IG	ZAIG	log	log	logit	-

1.4.2 Discrete distributions for *Y*

One parameter distributions

BI	Binomial
PO	Poisson

Two parameter distributions

BB	Beta-Binomial
NBI	Negative Binomial type I
NBII	Negative Binomial type II
PIG	Poisson-Inverse Gaussian
ZIP	Zero inflated Poisson (also ZIP2)

Three parameter distributions

SICHEL	Sichel
DEL	Delaporte

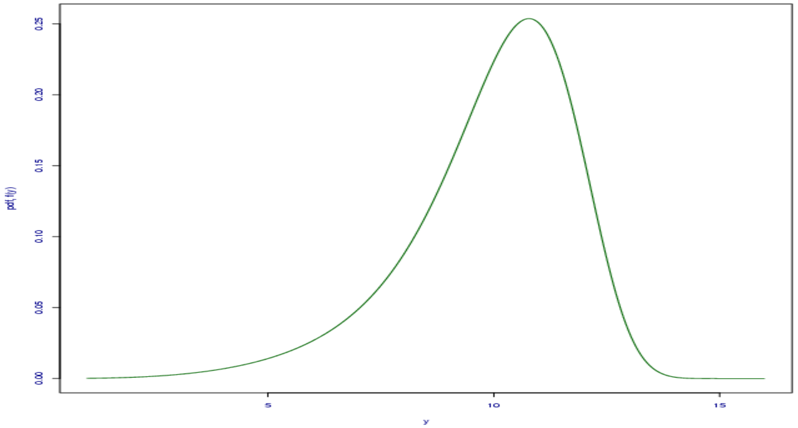
1.4.3 Continuous distributions for *Y*

Four parameters

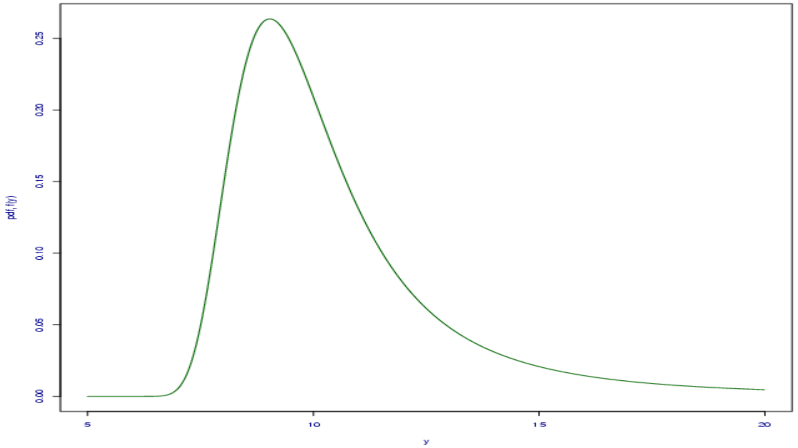
μ	location
σ	scale
ν	skewness
τ	kurtosis

Skewness and kurtosis

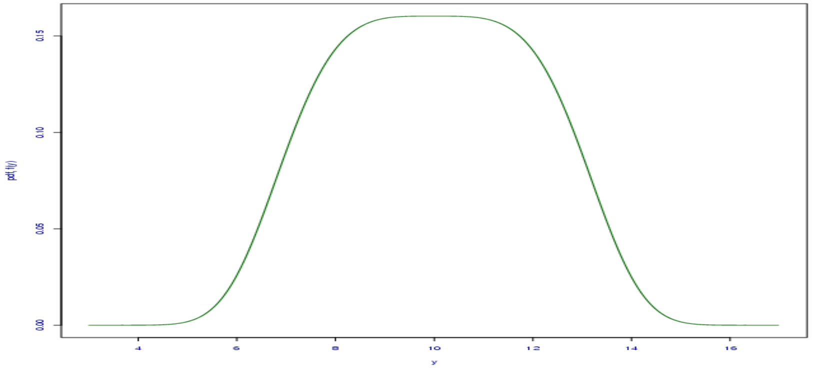
negative skewness



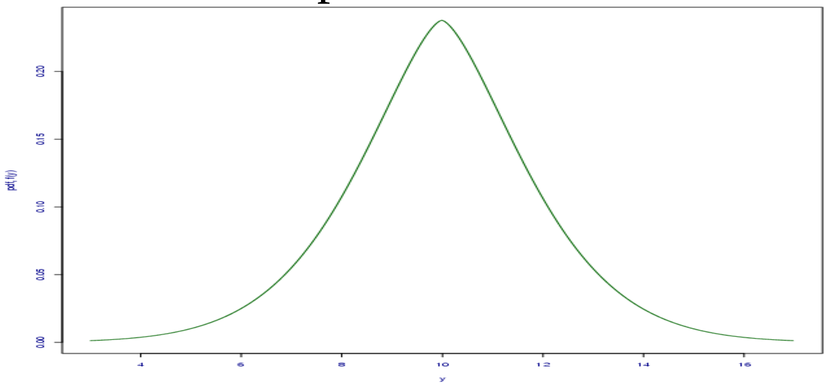
positive skewness



platykurtosis



leptokurtosis



gamlss
for statistical modelling

Continuous distributions for *Y*

Two parameter distributions

BE	Beta
GA	Gamma
GU	Gumbel
LO	Logistic
LNO	Log Normal
NO	Normal
IG	Inverse Gaussian
RG	Reverse Gumbel
WEI	Weibull (also WEI2, WEI3)

Continuous distributions for *Y*

Three parameter distributions

BCCG	Box-Cox Normal
PE	Power Exponential
TF	*t* family

Continuous distributions for *Y*

Four parameter distributions

BCT	Box-Cox *t*
BCPE	Box-Cox Power Exponential
JSU	Johnson Su
SHASH	Sinh Arc Shinh
SEP	Skew Exponential Power
ST3	Skew *t*

1.4.3 Mixed distributions for Y

These distributions are a mixture of discrete and continuous distributions.

Four parameter distributions

BEINF	Beta inflated
ZAIG	Zero-adjusted inverse Gaussian

gamlss
for statistical modelling

1.4.4 Mixture distributions for *Y*

These distributions are mixtures of **any** 2 or more **GAMLSS** distributions of **any** type.

They are fitted using general functions:
gamlssMX and **gamlssNP**

e.g. mixture of two Weibull distributions,
zero inflated distributions

gamlss
for statistical modelling

1.5 Additive terms

Each parameter of the distribution μ, σ, ν, τ is modelled using terms in explanatory variables x

Parametric additive terms

- Linear and interaction terms for variables and factors.
- Polynomials, inverse polynomials,
  piecewise polynomials (with fixed knots),
  fractional polynomials (Royston and Altman, 1994)
- Non-linear parametric terms

Smoothing and random effects additive terms

- Additive smoothing terms
  - ✓ loess (Cleveland *et al.*, 1993)
  - ✓ cubic splines (Green and Silverman, 1994)
  - ✓ P-splines (Eilers and Marx, 1996)
  - ✓ varying coefficient models (Hastie and Tibshirani, 1993)
- Random effects (overdispersion, simple random effects, random coefficients)
- Parameter driven Time Series (random walks)

Additive terms in GAMLSS

Each parameter of the distribution μ, σ, ν, τ is modelled using (additive) terms in explanatory variables x

Model	Formula	gamlss code
simple linear	$\beta_0 + \beta_1 x$	x
polynomial	$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$	poly(x,3)
fractional polynomial	$\beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_2}$	fp(x,2)
	for $p \in (-2, -1, 0, 0.5, 1, 2, 3)$	
e.g.	$\beta_0 + \beta_1 x^{0.5} + \beta_2 x^{-2}$	bfp(0.5,-2)
power polynomial	$\beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_2}$	pp(x,2)
loess smoother		lo(x,3)
p-spline smoother		ps(x,3)
cubic spline smoother		cs(x,3)

gamlss
for statistical modelling

1.6 Model Estimation

$\boldsymbol{y}$ response vector of length n

$X = (X_1, X_2, \ldots, X_p)$ fixed effects design matrices

$Z = (Z_{11}, Z_{21}, \ldots, Z_{J_p p})$ random effects design matrices

$\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_p^T)$ fixed effects parameters

$\boldsymbol{\gamma}^T = (\boldsymbol{\gamma}_{11}^T, \boldsymbol{\gamma}_{21}^T, \ldots, \boldsymbol{\gamma}_{J_p p}^T)$ random effects parameters

$\boldsymbol{\lambda}^T = (\boldsymbol{\lambda}_{11}^T, \boldsymbol{\lambda}_{21}^T, \ldots, \boldsymbol{\lambda}_{J_p p}^T)$ hyperparameters

Model estimation in GAMLSS

1) (Linear or nonlinear) parametric GAMLSS model

i.e. no random effects or smoothing terms

i.e. fixed effects but no random effects parameters

i.e. β's but no γ's or λ's

GAMLSS estimates β's by maximum likelihood estimation

2) Random effects or semi-parametric GAMLSS model

i.e. random effects or smoothing terms

i.e. both fixed effects and random effects parameters

i.e. β's, γ's and λ's

(i) GAMLSS estimates (β, γ) by posterior mode estimation

(ii) GAMLSS estimates hyperparameters λ by

(a) maximising a profile GAIC over λ

(GAIC = generalized Akaike information criterion)

or (b) maximising the marginal likelihood of λ

MAP estimation of (β, γ) given λ

MAP = maximum *a posteriori* = posterior mode

$$f(\boldsymbol{\beta}, \gamma | \mathbf{y}, \lambda) \propto f(\mathbf{y} | \boldsymbol{\beta}, \gamma) f(\gamma | \lambda)$$

assuming a uniform prior for $\boldsymbol{\beta}$.

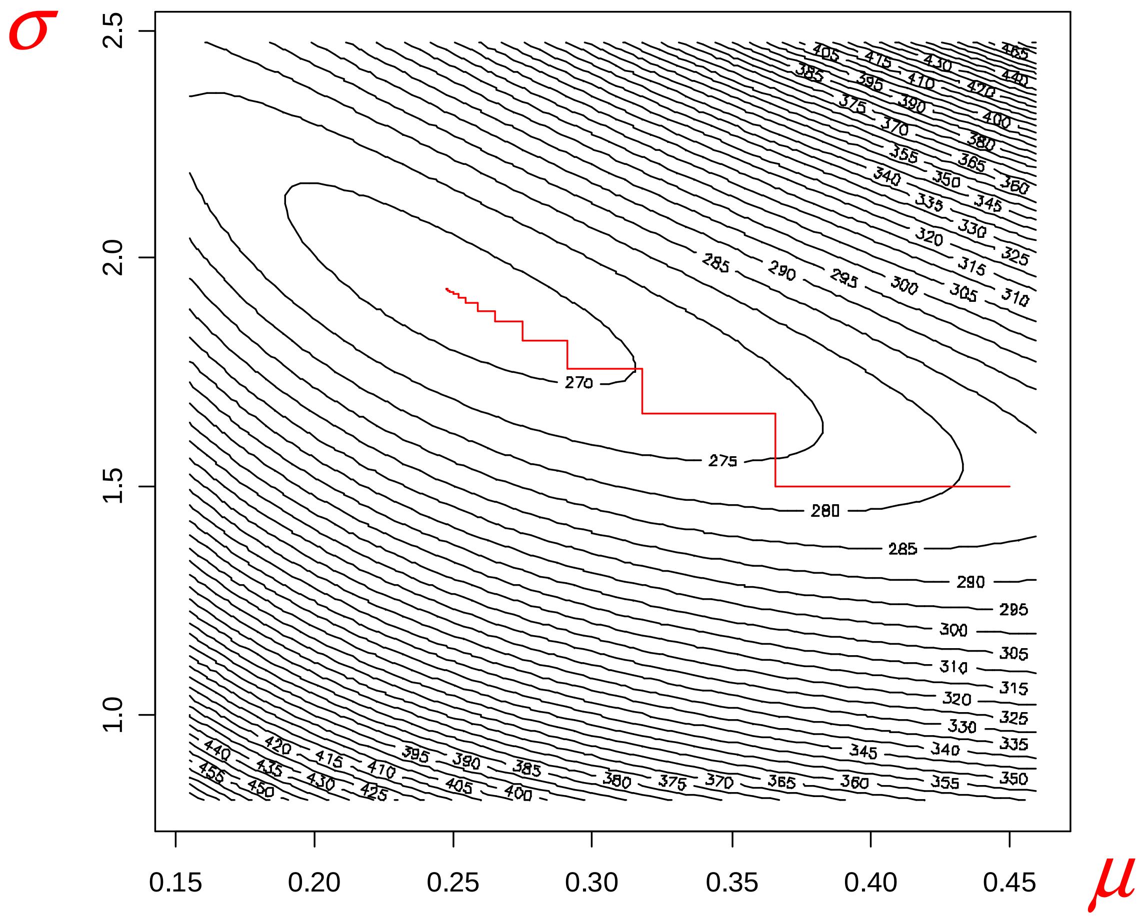
1.7 Algorithms for estimation of (β, γ) given λ

1. RS: a generalization of the MADAM algorithm, Rigby and Stasinopoulos (1996a)

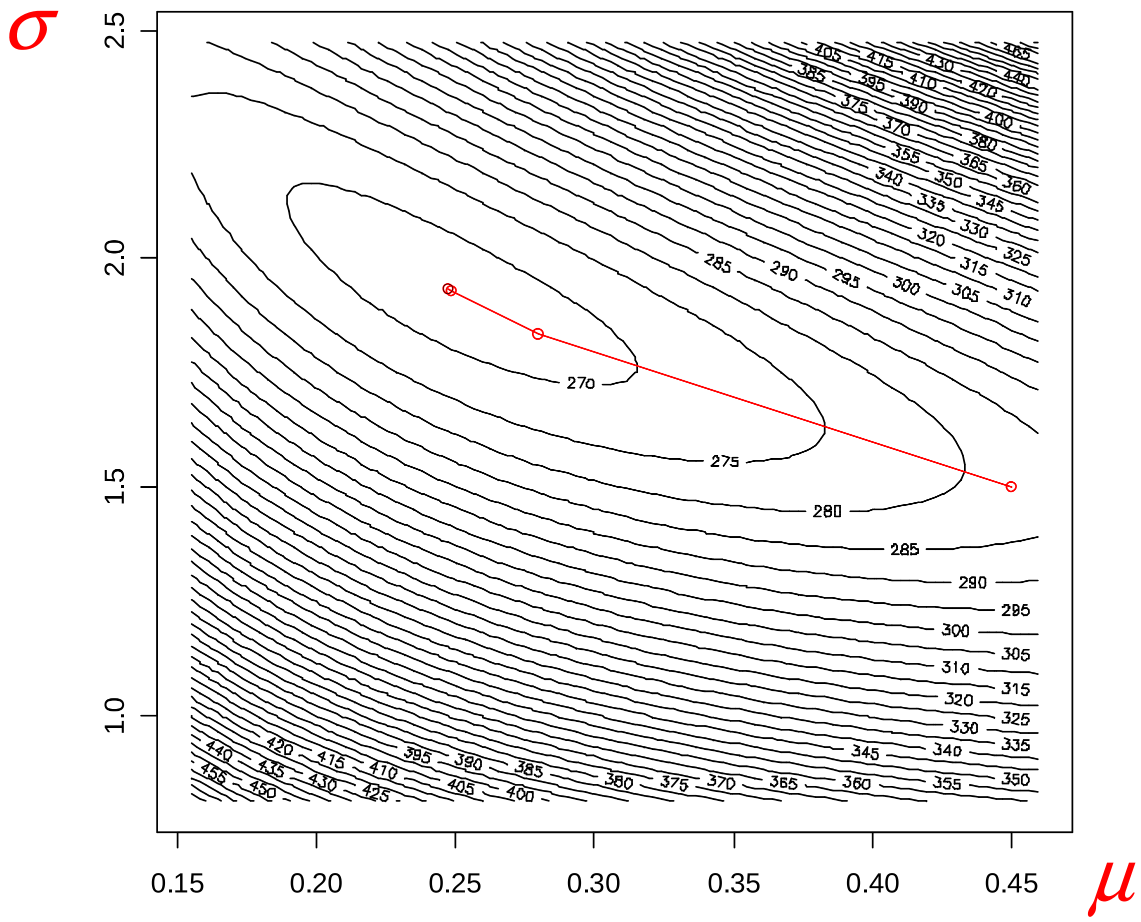
2. CG: a generalization of Cole and Green (1992)

3. mixed: a mixture of RS+CG
   (i.e. *j* iterations of RS, followed by *k* iterations of CG)

Weibull example of RS algorithm



Weibull example of CG algorithm



Advantages of algorithms

1) flexible modular fitting procedure

2) easy implementation of new distributions

3) easy implementation of new additive terms

4) simple starting values for (μ,σ,ν,τ) easily found

5) stable and reliable algorithms

6) very fast fitting (for fixed hyperparameters)

gamlss for statistical modelling

1.8 (Normalized quantile) residuals

The (normalized quantile) residuals from the fitted model are :

$\hat{r} = \Phi^{-1}[u]$	if Y is continuous, where $u = \hat{F}(y)$
$\hat{r} = \Phi^{-1}[u]$	if Y is discrete, where $u \sim U[\hat{F}(y-1), \hat{F}(y)]$

and Φ^{-1} is the inverse cdf of a N(0,1) variable

and $\hat{F}(y)$ is the fitted model cdf of Y

The true residuals r from the model have a N(0,1) distribution,

whatever the original correct distribution of Y.

1.9 Conclusion

GAMLSS allows flexible modelling of both:

i) the distribution of Y, including models for skewness and kurtosis

ii) the dependence of the distribution parameters, e.g. μ, σ, ν, τ, on explanatory variables and random effect additive terms.