An introduction to the GAMLSS packages in R

Fernanda De Bastiani[1], Cristian Villegas [2]
Thanks to Mikis D. Stasinopoulos and Robert A. Rigby

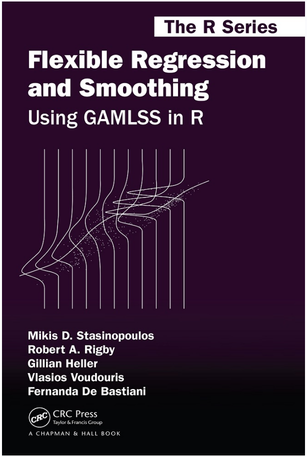
[1] Universidade Federal de Pernambuco
[2] ESALQ/USP

65ª RBRAS e 19º SEAGRO
Piracicaba (online), São Paulo, Brazil, September 2021

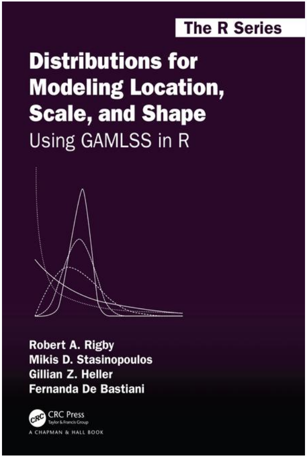
About the course

1. First Part
    - An introduction to GAMLSS and its statistical modeling philosophy. The different type of distributions within `gamlss.dist` distributions. A brief presentation of estimation, additive terms, model selection and diagnostics.
2. Second Part
    - Practical. An introduction to the `R` implementation of GAMLSS. Some `gamlss` packages functions for model fitting, modeling selection techniques and diagnostics. How to implement a new distribution within the `gamlss` packages framework.

Book 1: about GAMLSS and R packages



Book 2: about distributions in GAMLSS



Reality is not always ‘normal’ and for larger data sets the assumption of normality very seldom are satisfied when the model is checked.

Practitioners who use the standard linear regression model soon find that the normality and constant variance of the errors terms and linearity of the relationship between response variable and the explanatory variables very seldom hold.

The 1980s Munich rent data

`R` : the monthly net rent for flats in the city of Munich.

`Fl` : the floor space area in square meters

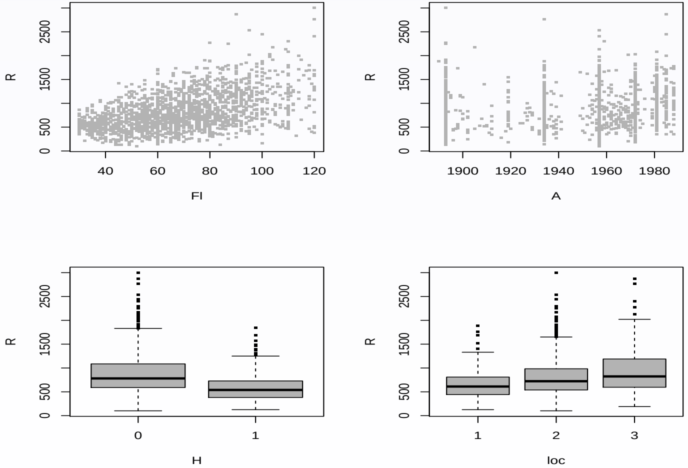
`A` : year of construction

`loc` : whether the location is below, 1, average, 2, or above average 3

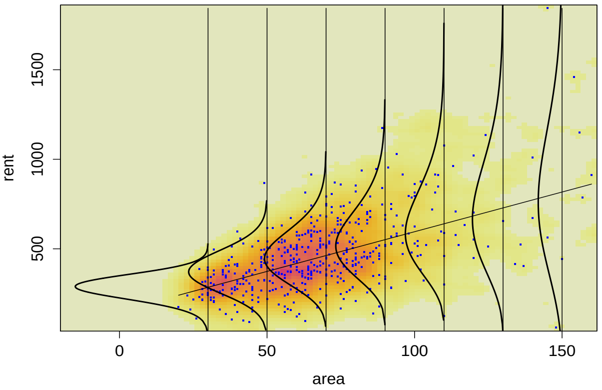
`H` : two level factor indicating whether there is central heating, (0), or not, (1).

Source: Munich rental guide 1993

The Munich rent data



The Munich rent data: area



Complexity of the relationship between rent and the explanatory variables. The dependence of the median of the response variable rent on floor space and age of construction is nonlinear, and nonparametric smoothing functions may be needed.

Non-homogeneity of variance of rent. Indication of non-homogeneity of the variance of rent. The variance of rent may depend on its mean and/or explanatory variables.

Skewness in the distribution of rent. Indication of positive skewness in the distribution of rent.

The `aids` data

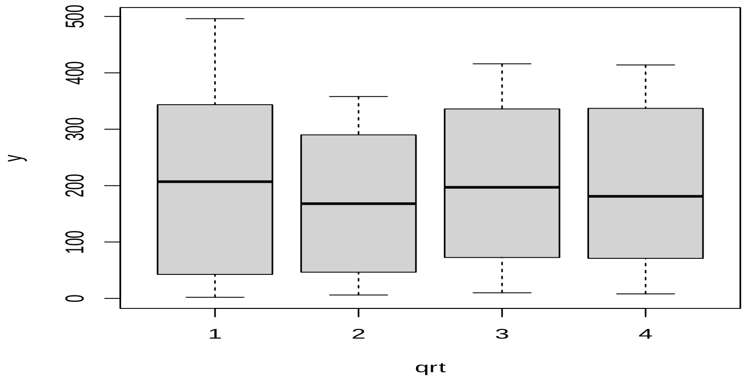
`y` : the number of quarterly aids cases in England and Wales

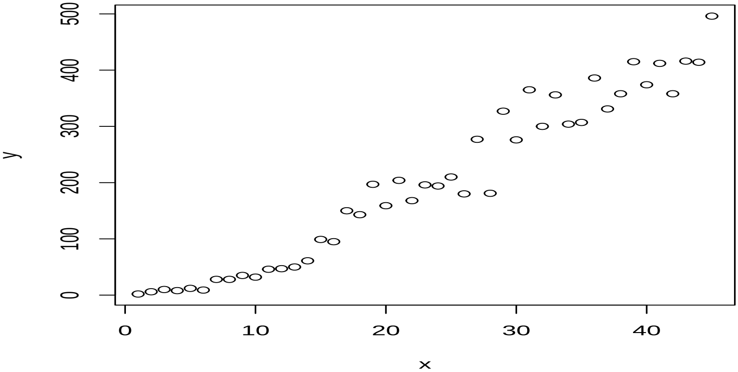
`x` : time in quarters from January 1983 to March 1994

`qrt` : a factor for the quarterly seasonal effect (1, 2, 3, 4)

Source: Public Health Laboratory Service, Communicable Disease Surveillance Centre, London

`aids` in package **gamlss.data** of dimensions 45×3


500
400
300
200
100
0
y
1
2
3
4
qrt


y
0
100
200
300
400
500
0
10
20
30
40
x

The film data

4031 film openings in US 1988-1999:

- `lborev1` : the response variable
  log of the (total revenues - the first week revenue), calculated in 1987 prices
- `lboopen` : the log the first week revenue, calculated in 1987 prices
- `lnosc` : the log of the number of screens
- `dist` : the type of distributor i.e. whether independent or major distributor

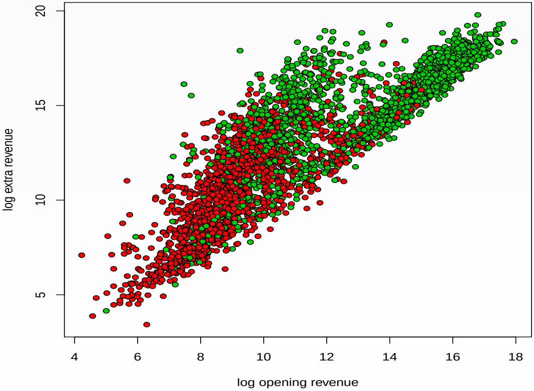
Source: Voudouris *et al.* (2010)

The film data: Question of interest

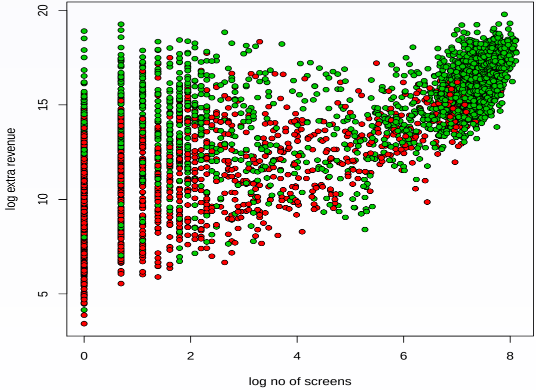
- Can we predict the extra revenue given the revenue at the first week?
- Which other explanatory variables are important?
- What is the conditional distribution of the response variable given the explanatory variables?

gamlss

The film data: log extra revenue against log first week revenue



The film data: log extra revenue against log number of screens



To learn from data (i.e. extract important relationships), especially with the large data sets generated today, we need flexible statistical frameworks to capture:

- The heavy-tailed or light-tailed characteristics of the distribution of the data.
- The skewness of the response variable, which might change as a function of the explanatory variables.
- The nonlinear or smooth relationship between the response variable and the explanatory variables.

What we need for modelling the above data?

We need

- flexible distributions for the response variable
- to be able to deal with heterogeneity in the data
- to be able to model skewness and kurtosis
- We need modelling all the parameters of the distributions
- flexible functions to model the relationship between the parameter of the distribution and the explanatory variables

gamlss

The statistical modelling philosophy

Statistical modelling is the art of building parsimonious statistical models for a better understanding of the phenomena of interest.

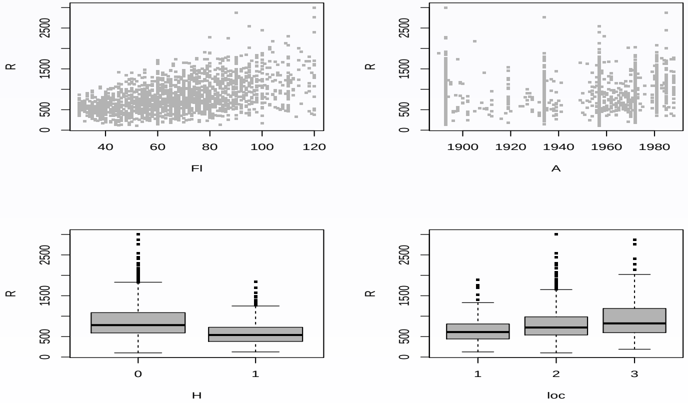
- Any model is a simplification of reality therefore no model is correct but some of them are useful
- Occam's Razor which states *'entities should not be multiplied beyond necessity'* or KISS (Keep It Simple Stupid)
- As statisticians we should try different models and choose the one appropriate for the data

Revisiting the Munich rent data

- `R` : the monthly net rent for flats in the city of Munich.
- `Fl` : the floor space area in square meters
- `A` : year of construction
- `loc` : whether the location is below, 1, average, 2, or above average 3
- `H` : two level factor indicating whether there is central heating, (0), or not, (1).

Source: Munich rental guide 1993

The Munich rent data



The Munich rent data: Linear model

Model assumptions

$$\begin{aligned} \mathbf{y} &\overset{\text{ind}}{\sim} N(\boldsymbol{\mu}, \sigma^2). \\ \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Estimation

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}{n}$$

The linear model assumptions

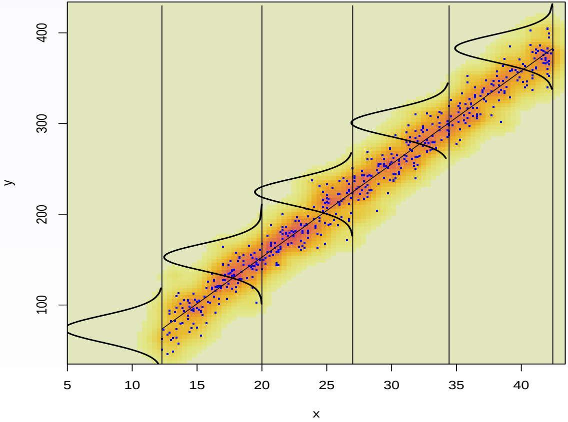


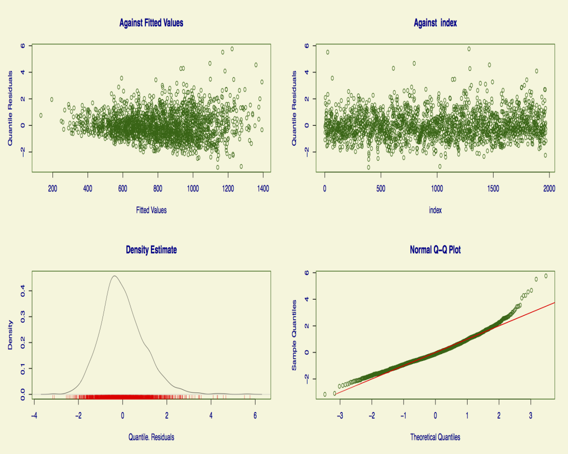
Figure: The linear regression model assumptions

gamlss

The Munich rent data: Linear model

```
r1 <-  gamlss(R ~ Fl+A+H+loc, family=NO,  data=rent)

## GAMLSS-RS iteration 1: Global Deviance = 28159
## GAMLSS-RS iteration 2: Global Deviance = 28159

l1 <- lm(R ~ Fl+A+H+loc,data=rent)
coef(r1)

##  (Intercept)          Fl           A          H1        loc2
## -2775.038803    8.839445    1.480755 -204.759562  134.052349
##         loc3
##   209.581472

coef(l1)

##  (Intercept)          Fl           A          H1        loc2
## -2775.038803    8.839445    1.480755 -204.759562  134.052349
##         loc3
##   209.581472
```

The Munich rent data: LM residuals plot



See Dunn and Smyth (1996) for a discussion of normalized (randomized) quantile residuals.

The Munich rent data: Generalized Linear Model

The model

$$\begin{aligned} \mathbf{y} &\overset{\text{ind}}{\sim} \mathit{ExpFamily}(\boldsymbol{\mu}, \phi) \\ g(\boldsymbol{\mu}) &= \mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

The exponential family

$$f_Y(y; \mu, \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}$$

The Munich rent data: GLM fit

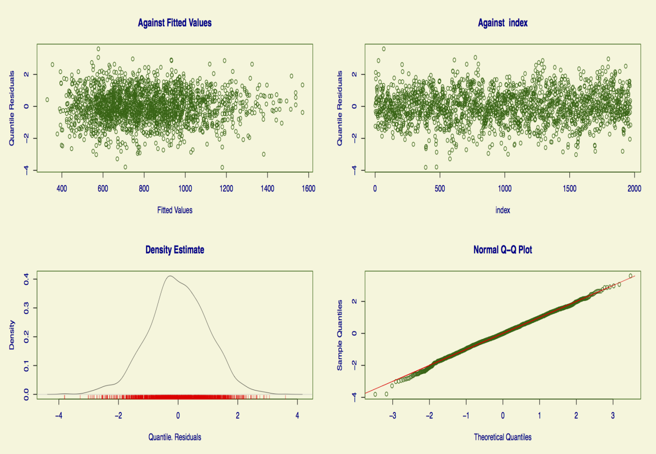
```
l2 <- glm(R ~ Fl+A+H+loc, family=Gamma(link="log"),  data=rent)
r2 <-  gamlss(R ~ Fl+A+H+loc, family=GA,  data=rent)

## GAMLSS-RS iteration 1: Global Deviance = 27764.59
## GAMLSS-RS iteration 2: Global Deviance = 27764.59
```

The Munich rent data: GLM and GAIC

```
r22 <-  gamlss(R ~ Fl+A+H+loc, family=IG,  data=rent)

## GAMLSS-RS iteration 1: Global Deviance = 27991.56
## GAMLSS-RS iteration 2: Global Deviance = 27991.56

GAIC(r1, r2, r22) # AIC

##     df      AIC
## r2   7 27778.59
## r22  7 28005.56
## r1   7 28173.00

GAIC(r1, r2, r22, k=log(length(rent$R))) # SBC or BIC

##     df      AIC
## r2   7 27817.69
## r22  7 28044.66
## r1   7 28212.10
```

gamlss

The Munich rent data: GLM residuals plot



The Munich rent data: Generalized Additive Model

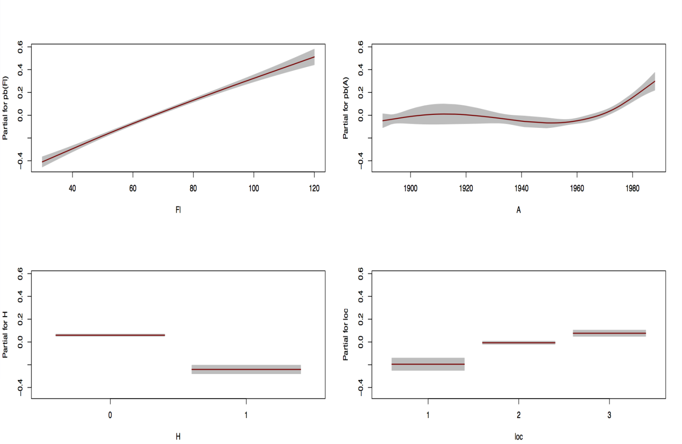
The model

$$\begin{aligned} \mathbf{y} &\overset{\text{ind}}{\sim} \mathit{ExpFamily}(\boldsymbol{\mu}, \phi) \\ g\left(\boldsymbol{\mu}\right) &= \mathbf{X}\boldsymbol{\beta} + s_1(\mathbf{x}_1) + \ldots + s_J(\mathbf{x}_J) \end{aligned}$$

The Munich rent data: GAM commands

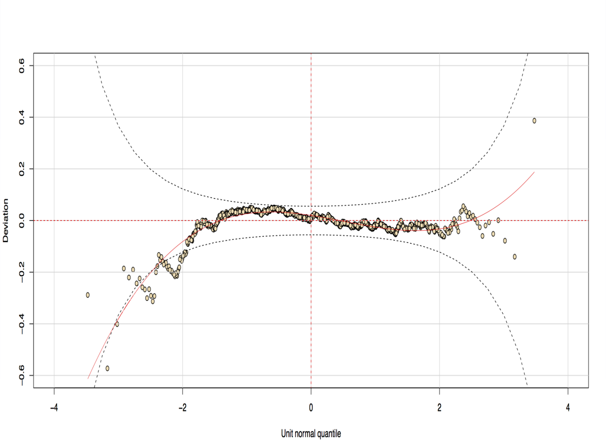
```
r3 <-  gamlss(R ~ pb(Fl)+pb(A)+H+loc, family=GA,  data=rent)

## GAMLSS-RS iteration 1: Global Deviance = 27683.22
## GAMLSS-RS iteration 2: Global Deviance = 27683.22
## GAMLSS-RS iteration 3: Global Deviance = 27683.22

AIC(r2,r3)

##          df      AIC
## r3 11.21547 27705.65
## r2  7.00000 27778.59
```

The Munich rent data: GAM term plot



gamlss

The Munich rent data: GAM worm plot



The Munich rent data: Mean and dispersion additive models

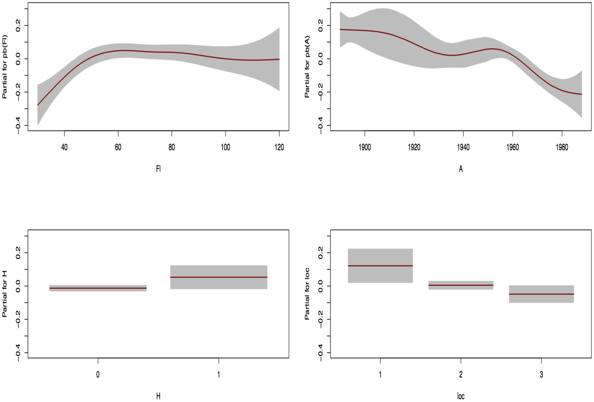
The model

$$
\begin{aligned}
\mathbf{y} &\overset{\text{ind}}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}) \\
g_1(\boldsymbol{\mu}) &= \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(\mathbf{x}_{11}) + \ldots + s_{1J_1}(\mathbf{x}_{1J_1}) \\
g_2(\boldsymbol{\sigma}) &= \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(\mathbf{x}_{21}) + \ldots + s_{2J_2}(\mathbf{x}_{2J_2})
\end{aligned}
$$

The Munich rent data: MADAM commands

```
r4 <-  gamlss(R ~ pb(Fl)+pb(A)+H+loc, sigma.fo=~pb(Fl)+pb(A)+H+loc, family=GA,
              data=rent)

## GAMLSS-RS iteration 1: Global Deviance = 27572.14
## GAMLSS-RS iteration 2: Global Deviance = 27570.29
## GAMLSS-RS iteration 3: Global Deviance = 27570.28
## GAMLSS-RS iteration 4: Global Deviance = 27570.28

r5 <-  gamlss(R ~ pb(Fl)+pb(A)+H+loc, sigma.fo=~pb(Fl)+pb(A)+H+loc, family=IG,
              data=rent)

## GAMLSS-RS iteration 1: Global Deviance = 27675.74
## GAMLSS-RS iteration 2: Global Deviance = 27672.97
## GAMLSS-RS iteration 3: Global Deviance = 27673
## GAMLSS-RS iteration 4: Global Deviance = 27673.01
## GAMLSS-RS iteration 5: Global Deviance = 27673.01
## GAMLSS-RS iteration 6: Global Deviance = 27673.02

AIC(r3, r4, r5)

##          df      AIC
## r4 22.25035 27614.78
## r3 11.21547 27705.65
## r5 21.82318 27716.66
```

The Munich rent data: MADAM term plot for σ

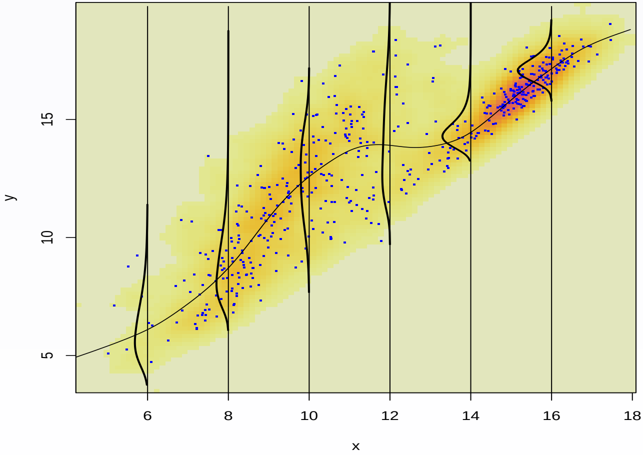


The Munich rent data: GAMLSS

The model

$$\begin{aligned}
\mathbf{y} &\overset{\text{ind}}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\
g_1(\boldsymbol{\mu}) &= \mathbf{X}_1 \boldsymbol{\beta}_1 + s_{11}(\mathbf{x}_{11}) + \ldots + s_{1J_1}(\mathbf{x}_{1J_1}) \\
g_2(\boldsymbol{\sigma}) &= \mathbf{X}_2 \boldsymbol{\beta}_2 + s_{21}(\mathbf{x}_{21}) + \ldots + s_{2J_2}(\mathbf{x}_{2J_2}) \\
g_3(\boldsymbol{\nu}) &= \mathbf{X}_3 \boldsymbol{\beta}_3 + s_{31}(\mathbf{x}_{31}) + \ldots + s_{3J_3}(\mathbf{x}_{3J_3}) \\
g_4(\boldsymbol{\tau}) &= \mathbf{X}_4 \boldsymbol{\beta}_4 + s_{41}(\mathbf{x}_{41}) + \ldots + s_{4J_4}(\mathbf{x}_{4J_4})
\end{aligned}$$

The Munich rent data: GAMLSS assumptions



The Munich rent data: GAMLSS commands

```
r6 <-  gamlss(R ~ pb(Fl)+pb(A)+H+loc, sigma.fo=~pb(Fl)+pb(A)+H+loc,
              nu.fo=~1, family=BCCGo,  data=rent)
```

```
r7 <-  gamlss(R ~ pb(Fl)+pb(A)+H+loc,sigma.fo=~pb(Fl)+pb(A)+H+loc,
              nu.fo=~pb(Fl)+pb(A)+H+loc, family=BCCGo,  data=rent)
```

```
AIC(r4, r6, r7)

##           df      AIC
## r7 28.41391 27608.15
## r6 22.48092 27611.02
## r4 22.25035 27614.78
```

BCCG: Box-Cox Cole and Green distribution

Historical development

Important events in the creation of the GAMLSS models

- Linear model (Gauss, 1809)
- Generalized Linear Models (Nelder and Wedderburn, 1972)
- Generalized Additive Models (Hastie and Tibshirani, 1990)
- Generalized Additive Models for Location Scale and Shape (GAMLSS) (Rigby and Stasinopoulos, 2005).

GAMLSS, and its associated R software

A flexible statistical framework for learning from data and enhancing data analytics.

The GAMLSS statistical framework enables flexible regression and smoothing models to be fitted to the data.

The GAMLSS model assumes that the distribution of response variable has any parametric distribution which might be heavy or light-tailed, and positively or negatively skewed. In addition, all the parameters of the distribution (location (e.g. mean), scale (e.g. variance) and shape (skewness and kurtosis)) can be modelled as linear, nonlinear or smooth functions of explanatory variables).

What is GAMLSS?

GAMLSS : are distributional based semi-parametric regression type models.

- regression type: we have many explanatory variables $\mathbf{X}$ and one response variable $\mathbf{y}$ and we believe that $\mathbf{X} \rightarrow \mathbf{y}$
- distributional: a full parametric distribution assumption is made for the response variable
- semi-parametric: All the parameters of the distribution can be modelled as function of the explanatory variables. Those functions can be parametric or non-parametric smoothing functions
- GAMLSS philosophy: try different models

GAMLSS is a generalisation of GLM and GAM models.

Generalized Additive Model for Location Scale and Shape

Generalized additive model for location scale and shape Rigby and Stasinopoulos (2005)

$$\mathbf{y} \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$$

$$\begin{aligned}
g_\mu(\boldsymbol{\mu}) &= \mathbf{X}_\mu \boldsymbol{\beta}_\mu + s_{1,\mu}(\mathbf{x}_{1,\mu}) + \ldots + s_{k,\mu}(\mathbf{x}_{k,\mu}) \\
g_\sigma(\boldsymbol{\sigma}) &= \mathbf{X}_\sigma \boldsymbol{\beta}_\sigma + s_{1,\sigma}(\mathbf{x}_{1,\sigma}) + \ldots + s_{k,\sigma}(\mathbf{x}_{k,\sigma}) \\
g_\nu(\boldsymbol{\nu}) &= \mathbf{X}_\nu \boldsymbol{\beta}_\nu + s_{1,\nu}(\mathbf{x}_{1,\nu}) + \ldots + s_{k,\nu}(\mathbf{x}_{k,\nu}) \\
g_\tau(\boldsymbol{\tau}) &= \mathbf{X}_\tau \boldsymbol{\beta}_\tau + s_{1,\tau}(\mathbf{x}_{1,\tau}) + \ldots + s_{k,\tau}(\mathbf{x}_{k,\tau})
\end{aligned}$$

where $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ can be any distribution and where $h_j(\mathbf{x}_j)$ are smooth functions of the X's.

Random effects form

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

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

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

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

where $\boldsymbol{\gamma}_{jk} \sim N_{q_{jk}}(\mathbf{0}, \lambda_{jk}^{-1}\mathbf{G}_{jk}^{-1})$.

- Fitting the parametric model requires only estimates for the 'betas' $\boldsymbol{\beta}$.

- Fitting the random effects GAMLSS model requires estimates for the 'betas' $\boldsymbol{\beta}$, the 'gammas' γ, and also the 'lambdas' $\boldsymbol{\lambda}$, where:

$$\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{11}^\top, \ldots, \boldsymbol{\lambda}_{1J_1}^\top, \boldsymbol{\lambda}_{21}^\top, \ldots, \boldsymbol{\lambda}_{4J_4}^\top)^\top \ .$$

- Within **gamlss**, the parametric GAMLSS model is fitted by maximum likelihood estimation with respect to $\boldsymbol{\beta}$, whereas

- the more general random effects model is fitted by maximum penalized likelihood estimation estimation with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for fixed $\boldsymbol{\lambda}$.

For more details, see Chapter 3 of Stasinopoulos et al (2017).

Estimation

The log-likelihood function for the GAMLSS model under the assumption that observations of the response variable are independent is given by

$$\ell(\theta) = \sum_{i=1}^{n} \log f_Y(y_i|\mu_i, \sigma_i, \nu_i, \tau_i),$$

where $f_Y(.)$ represents the probability (density) function of the response variable. The respective penalized log-likelihood function is given by

$$\ell_p(\theta) = \ell(\theta) - \frac{1}{2}\sum_{k=1}^{4}\sum_{j=1}^{J_k} \lambda_{jk}\gamma_{jk}^{\top}\mathbf{G}_{jk}\gamma_{jk}.$$

The two basic algorithms for fitting the parametric model with respect to $\boldsymbol{\beta}$, and the nonparametric model with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for fixed $\boldsymbol{\lambda}$, implementted in **gamlss** in **R** are:

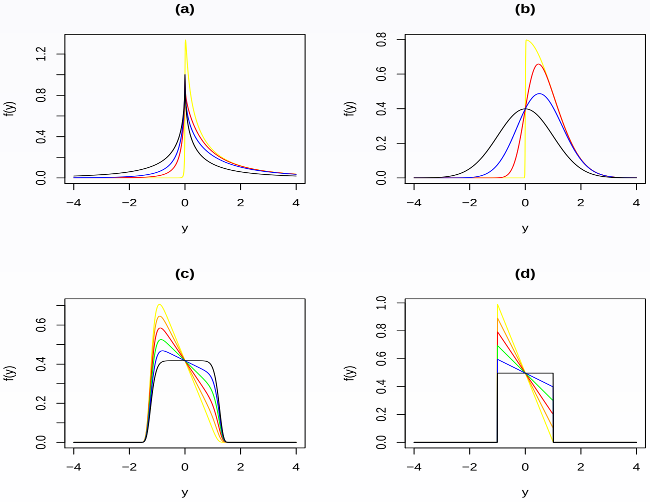
- **RS** and
- **CG**.

Both use an iteratively reweighted (penalized) least squares algorithm (Rigby and Stasinopoulos, 2013 and Stasinopoulos et al, 2017).

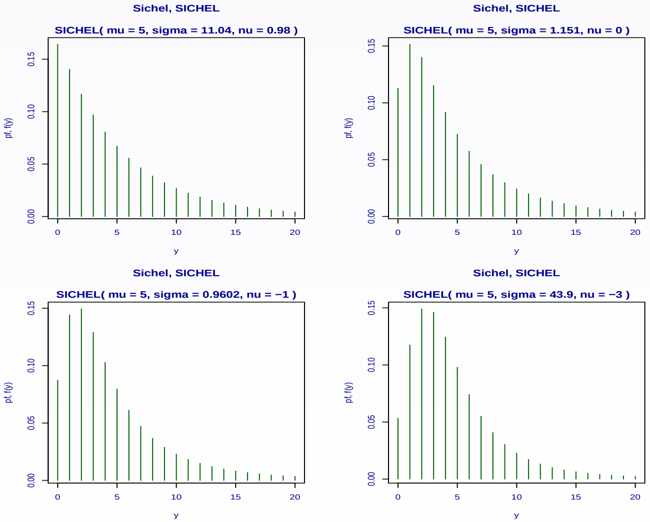
Different types of distribution within `galmss.dist`

1. *continuous distributions*: $f_Y(y|\boldsymbol{\theta})$, are usually defined on $(-\infty, +\infty)$, $(0, +\infty)$ or $(0, 1)$.
2. *discrete distributions*: $P(Y = y|\boldsymbol{\theta})$ are defined on $y = 0, 1, 2, \ldots, n$, where n is a known finite value or n is infinite, i.e. usually discrete (count) values.
3. *mixed distributions*: (finite mixture distributions) are mixtures of continuous and discrete distributions, i.e. continuous distributions where the range of Y has been expanded to include some discrete values with non-zero probabilities.

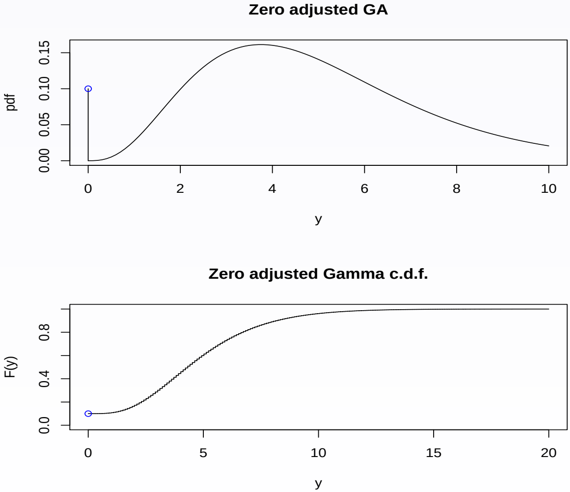
Example of continuous distribution: SEP1(μ, σ, ν, τ)



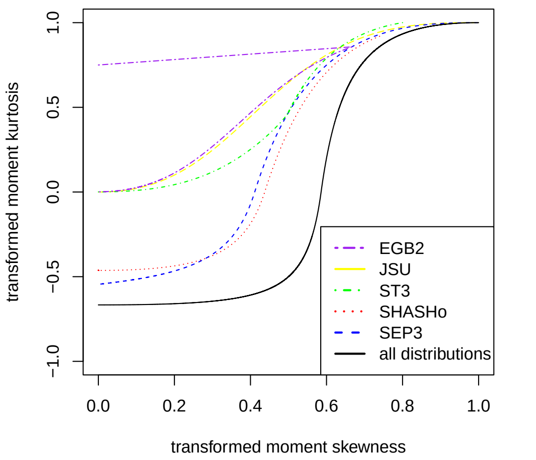
Example of discrete distribution: SICHEL(μ, σ, ν)



Example of mixed distribution distributions: ZAGA(μ, σ, ν)



Moment skewness and kurtosis comparison



Generating Distributions

There are over 100 explicit distributions in GAMLSS.

Further distributions can be generated

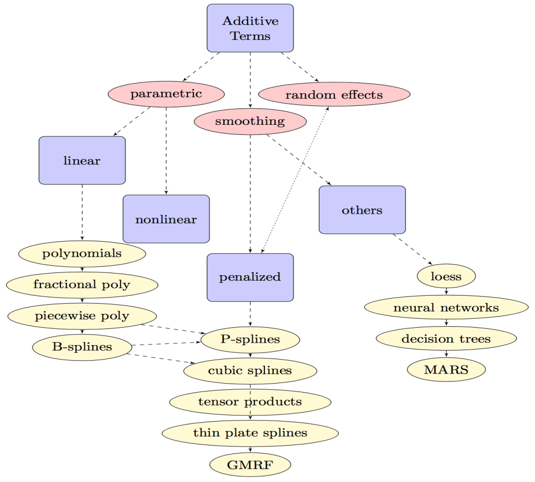
Generating Distributions

- take a continuous distribution defined on $(-\infty, \infty)$ and create a log version with range $(0, \infty)$
- take a continuous distribution defined on $(-\infty, \infty)$ and create a logit version with range $(0, 1)$
- take any continuous or discrete distribution and truncate its range. This can be “left”, “right”, or “both” truncation
- take any continuous distribution defined on $(0, \infty)$ and by interval censoring create a discrete count distribution defined on $\{0, 1, 2, \ldots\}$

Generating Distributions

- take a continuous distribution defined on $(-\infty, \infty)$ or $(0, \infty)$ and create by left, right or interval censoring a generalized Tobit model
- take any continuous distribution defined on $(0, \infty)$ and zero-adjust it to create a mixed distribution on $[0, \infty)$
- take any continuous distribution defined on $(0, 1)$ and zero- and/or one-inflate it to create a mixed distribution on $[0, 1)$, $(0, 1]$ or $[0, 1]$
- mix different `gamlss.family` distributions to create a new finite mixture distribution see Chapter 7 of Stasinopoulos *et al.* (2017)

Additive Terms



The R packages

gamlss	the original package
gamlss.dist	all `gamlss.family` distributions
gamlss.data	different sets of data
gamlss.add	for extra additive terms
gamlss.cens	for censored (left, right or interval) response variables
gamlss.demo	demos for distributions and smoothing
gamlss.inf	models for inflated 0 to 1 or for adjusted 0 to ∞
gamlss.nl	non-linear term fitting
gamlss.tr	generating truncated distributions
gamlss.mx	finite mixtures distributions and random effects
gamlss.spatial	for spatial models [GMRF models]
gamlss.util	for extra utilities
gamlss.foreach	for parallel computing

GAMLSS components

Let $\mathcal{M} = \{\mathcal{D}, \mathcal{G}, \mathcal{T}, \boldsymbol{\lambda}\}$ represent the GAMLSS model

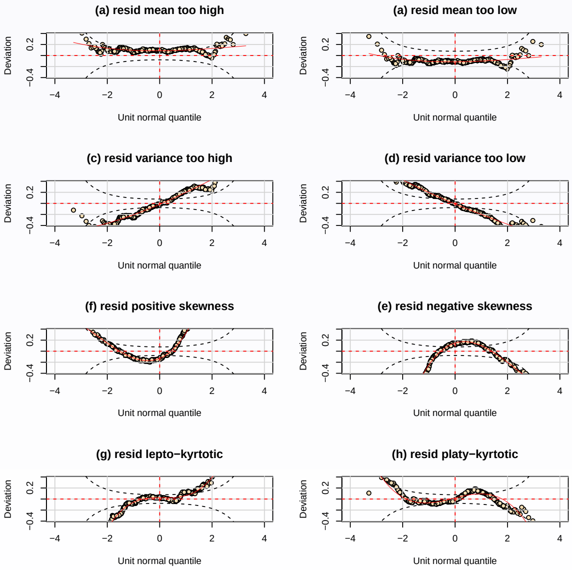
- $\mathcal{D}$: distribution
- $\mathcal{G}$: the link function for distributional parameters
- $\mathcal{T}$: predictor terms for ($\boldsymbol{\eta}$' s) i.e. $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \sum_j s_j(\mathbf{x}_j)$
- $\boldsymbol{\lambda}$: the hyperparameters

Problems

- which distribution
- which the link function for distributional parameters
- which x-variables for μ
- which x-variables for σ
- which x-variables for ν
- which x-variables for τ
- choosing the smoothing hyper parameters for terms in μ, σ, ν and τ
- selection between different (GAMLSS or not) models

The different function for model selection

Comp.	All data	K-fold CV	Val./Test
$\mathcal{D}$	GAIC() wp()	gamlssCV(), CV()	gamlssVGD(), VGD() getTGV() TGD()
$\mathcal{G}$	deviance() *	gamlssCV()	as above
$\mathcal{T}$	drop1(), add1(), add1ALL(), drop1ALL(), stepGAIC() stepGAICAll.A() stepGAICAll.B()	gamlssCV() CV()	drop1TGD() add1TGD() stepTGD()
Λ global	findhyper()	optim()*	optim()*


(a) resid mean too high
(a) resid mean too low
(c) resid variance too high
(d) resid variance too low
(f) resid positive skewness
(e) resid negative skewness
(g) resid lepto−kyrtotic
(h) resid platy−kyrtotic
Deviation
0.2
−0.4
−4
−2
0
2
4
Unit normal quantile

Conclusions

- GAMLSS is a very flexible statistical model (but rather complex)
- It is a unified framework for univariate distributional regression models
- Allows any parametric distribution for the response variable Y
- Models all the parameters of the distribution of Y
- Allows a variety of penalised additive terms in the models for the distribution parameters

Conclusions (continue)

- The fitted algorithm is modular, where different components can be added easily
- it can easily introduced to students since it relies on known concepts
- For continuous response variables deals properly with skewness and kurtosis (heavy tails)
- For discrete type response variables deals with overdispersion (or underdispersion), with excess (or lack) of zeros and with heavy tails

To be continued...

Thank you

for more information see

www.gamlss.com

References

- Dunn, P.K. and Smyth, G.K. (19996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236-244.
- Rigby, R.A. and Stasinopoulos, D.M. (2005) Generalized Additive Models for Location, Scale and Shape (with discussion). *Appl. Statist.*, **54**, 507-554.
- Rigby, R.A., Stasinopoulos, D.M., Heller, G. Z. and De Bastiani, F. (2019) *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. Chapman and Hall/CRC.
- Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V. and De Bastiani, F. (2017) *Flexible regression and smoothing: Using GAMLSS in R*. Chapman and Hall/CRC.

gamlss