Flexible Regression and Smoothing
The gamlss.family Distributions

Bob Rigby     Mikis Stasinopoulos

Dipartimento di Scienze Cliniche a de Comunita, Universita degli studi di Milano, Italy, January 2018
Books & Vignettes

The GAMLSS Books

One book on GAMLSS is published and two others are in preparation

- `Flexible Regression and Smoothing: Using GAMLSS in R` was published in April 2017. See here for an old draft version.

- `Distributions for Location Scale and Shape: Using GAMLSS in R`, is in preparation. See here for the November 2017 draft version of the book.

- `Generalised Additive Models for Location Scale and Shape: A Distributional Regression Approach` is in preparation. A draft version will be appear in the summer 2018.
Different types of distribution in GAMLSS

1. **Continuous distributions**: $f_Y(y|\theta)$, are usually defined on $(-\infty, +\infty)$, $(0, +\infty)$ or $(0, 1)$.

2. **Discrete distributions**: $P(Y = y|\theta)$ 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.
Different types of distribution: demos

1. demo.GA()
2. demo.PO()
3. demo.BI()
4. demo.BE()
5. demo.BEINF()
Example of continuous distribution: SEP1
Example of discrete distribution: Sichel

Sichel, SICHEL

SICHEL( μ = 5, σ = 11.04, ν = 0.98 )

Sichel, SICHEL

SICHEL( μ = 5, σ = 1.151, ν = 0 )

Sichel, SICHEL

SICHEL( μ = 5, σ = 0.9602, ν = −1 )

Sichel, SICHEL

SICHEL( μ = 5, σ = 43.9, ν = −3 )
Example of mixed distribution distributions: ZAGA

Zero adjusted GA

Zero adjusted Gamma c.d.f.
Skewness and kurtosis

- **Negative skewness**
- **Positive skewness**
- **Platy-kurtosis**
- **Lepto-kurtosis**

Figure: Showing different types of continuous distributions
Types of continuous distributions

\((-\infty, \infty)\): real line \(\mathbb{R}\)

\((0, \infty)\): positive real line \(\mathbb{R}^+\)

\((0, 1)\) real line on interval \(\mathbb{R}_{(0,1)}\) (not containing zero or one)
For these distributions, if

\[ Y \sim D(\mu, \sigma, \nu, \tau) \]

then

\[ \varepsilon = (Y - \mu)/\sigma \sim D(0, 1, \nu, \tau), \]

i.e. \( Y = \mu + \sigma \varepsilon \), so \( Y \) is a scaled and shifted version of the random variable \( \varepsilon \).

All \((-\infty, \infty)\) distributions in GAMLSS except \text{exGAUS}(\mu, \sigma, \tau)\) are \textit{location-scale} families
Location-scale family

(a) changing the location parameter

(b) changing the scale parameter

Bob Rigby, Mikis Stasinopoulos
Flexible Regression and Smoothing
2016 12 / 58
Two parameter distributions on $\mathbb{R}$

- **Gumbel** $GU(\mu, \sigma)$ left skew
- **Logistic** $LO(\mu, \sigma)$ lepto
- **Normal** $NO(\mu, \sigma)$ and $NO2(\mu, \sigma)$
- **Reverse Gumbel** $RG(\mu, \sigma)$ right skew
Continuous distributions

Distributions on $\mathbb{R}$

Two parameter on $\mathbb{R}$

(a) NO and LO distributions

(b) NO, GU, ans RG distributions

Bob Rigby, Mikis Stasinopoulos
Flexible Regression and Smoothing

2016 14 / 58
Continuous distributions

Three parameter on $\mathbb{R}$

exponential Gaussian: $\text{exGAUS}(\mu, \sigma, \nu)$ for modelling right skew data,

normal family: $\text{NOF}(\mu, \sigma, \nu)$ for modelling mean and variance relationships following the power law;

power exponential: $\text{PE}(\mu, \sigma, \nu)$ and $\text{PE2}(\mu, \sigma, \nu)$ for modelling lepto and platy kurtotic data;

$t$ family: $\text{TF}(\mu, \sigma, \nu)$ and $\text{TF2}(\mu, \sigma, \nu)$ for modelling lepro kurtotic data;

skew normal: $\text{SN1}(\mu, \sigma, \nu)$ and $\text{SK2}(\mu, \sigma, \nu)$ for modelling skewness in data.
Continuous distributions

Distributions on $\mathbb{R}$

Three parameter on $\mathbb{R}$: skew normal type 1

Bob Rigby, Mikis Stasinopoulos
Flexible Regression and Smoothing

2016 16 / 58
Three parameter on $\mathbb{R}$: PE and TF distributions

(a) power exponential

(b) the t-family
Four parameter on $\mathbb{R}$

exponential generalised beta type 2: $\text{EGB2}(\mu, \sigma, \nu, \tau)$ skewness and leptokurtosis;

generalised $t$: $\text{GT}(\mu, \sigma, \nu, \tau)$ kurtosis;

Johnson’s SU: $\text{JSU}(\mu, \sigma, \nu, \tau)$ and $\text{JSUo}(\mu, \sigma, \nu, \tau)$ skewness and leptokurtosis;

normal-exponential-$t$: $\text{NET}(\mu, \sigma, \nu, \tau)$, robustly location and scale;

skew exponential power: $\text{SEP1}(\mu, \sigma, \nu, \tau)$, $\text{SEP2}(\mu, \sigma, \nu, \tau)$, $\text{SEP3}(\mu, \sigma, \nu, \tau)$ and $\text{SEP4}(\mu, \sigma, \nu, \tau)$ skewness and lepto-platy;

sinh-arcsinh: $\text{SHASH}(\mu, \sigma, \nu, \tau)$, $\text{SHASHo}(\mu, \sigma, \nu, \tau)$ and $\text{SHASHo2}(\mu, \sigma, \nu, \tau)$ skewness and lepto-platy;

skew $t$: $\text{ST1}(\mu, \sigma, \nu, \tau)$, $\text{ST2}(\mu, \sigma, \nu, \tau)$, $\text{ST3}(\mu, \sigma, \nu, \tau)$, $\text{ST4}(\mu, \sigma, \nu, \tau)$, $\text{ST5}(\mu, \sigma, \nu, \tau)$ and $\text{SST}(\mu, \sigma, \nu, \tau)$ skewness and leptokurtosis.
Example of four parameter on $\mathbb{R}$: SEP1
The NET distribution
Summary of GAMLSS family distributions defined on $(-\infty, +\infty)$

<table>
<thead>
<tr>
<th>Distributions</th>
<th>family</th>
<th>no par.</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. Gaussian</td>
<td>exGAUS</td>
<td>3</td>
<td>positive</td>
<td>-</td>
</tr>
<tr>
<td>Exp.G. beta 2</td>
<td>EGB2</td>
<td>4</td>
<td>both</td>
<td>lepto</td>
</tr>
<tr>
<td>Gen. $t$</td>
<td>GT</td>
<td>4</td>
<td>(symmetric)</td>
<td>lepto</td>
</tr>
<tr>
<td>Gumbel</td>
<td>GU</td>
<td>2</td>
<td>(negative)</td>
<td>-</td>
</tr>
<tr>
<td>Johnson’s SU</td>
<td>JSU, JSUo</td>
<td>4</td>
<td>both</td>
<td>lepto</td>
</tr>
<tr>
<td>Logistic</td>
<td>LO</td>
<td>2</td>
<td>(symmetric)</td>
<td>(lepto)</td>
</tr>
<tr>
<td>Normal-Expon.-$t$</td>
<td>NET</td>
<td>2,2</td>
<td>(symmetric)</td>
<td>lepto</td>
</tr>
<tr>
<td>Normal</td>
<td>NO-NO2</td>
<td>2</td>
<td>(symmetric)</td>
<td></td>
</tr>
<tr>
<td>Normal Family</td>
<td>NOF</td>
<td>3</td>
<td>(symmetric)</td>
<td>(meso)</td>
</tr>
</tbody>
</table>
Summary of GAMLSS family distributions defined on \((-\infty, +\infty)\)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Code</th>
<th>Parameters</th>
<th>Symmetry</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Expon.</td>
<td>PE-PE2</td>
<td>3</td>
<td>(symmetric)</td>
<td>both</td>
</tr>
<tr>
<td>Reverse Gumbel</td>
<td>RG</td>
<td>2</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>Sinh Arcsinh</td>
<td>SHASH,</td>
<td>4</td>
<td>both</td>
<td>both</td>
</tr>
<tr>
<td>Skew Exp. Power</td>
<td>SEP1-SEP4</td>
<td>4</td>
<td>both</td>
<td>both</td>
</tr>
<tr>
<td>Skew (t)</td>
<td>ST1-ST5, SST</td>
<td>4</td>
<td>both</td>
<td>lepto</td>
</tr>
<tr>
<td>(t) Family</td>
<td>TF</td>
<td>3</td>
<td>(symmetric)</td>
<td>lepto</td>
</tr>
</tbody>
</table>
Continuous distribution on $\mathbb{R}^+$: scale family

If a random variable is distributed as

$$Y \sim D(\mu, \sigma, \nu, \tau)$$

and

$$\varepsilon = \left( \frac{Y}{\mu} \right) \sim D(1, \sigma, \nu, \tau)$$

then $Y$ has a scale family and the random variable

$$Y = \mu \varepsilon$$

is a scaled version of the random variable $\varepsilon$. $\mu$ does not effect the shape.
Continuous distribution on $\mathbb{R}^+$: Weibull pdf

\[ f(y) \]

- $\mu = 1$
- $\sigma = 0.5$
- $\sigma = 1$
- $\sigma = 6$

- $\mu = 2$
- $\sigma = 0.5$
- $\sigma = 1$
- $\sigma = 6$
Continuous distribution on $\mathbb{R}^+$: one and two parameters

- Exponential: $\text{EXP}(\mu, \sigma)$
- Gamma: $\text{GA}(\mu, \sigma)$ member of the exponential family;
- Inverse gamma: $\text{IGAMMA}(\mu, \sigma)$;
- Inverse Gaussian: $\text{IG}(\mu, \sigma)$ a member of the exponential family;
- Log-normal: $\text{LOGNO}(\mu, \sigma)$ and $\text{LOGNO2}(\mu, \sigma)$.
  - Pareto: $\text{PARETO}(\mu, \sigma)$, $\text{PARETO2o}(\mu, \sigma)$ and $\text{GP}(\mu, \sigma)$ for heavy tail;
  - Weibull: $\text{WEI}(\mu, \sigma)$, $\text{WEI2}(\mu, \sigma)$ and $\text{WEI3}(\mu, \sigma)$ used in survival analysis.
Continuous distributions

Continuous distribution on $\mathbb{R}^+$: Weibull survival and hazard functions

(a) survival function

(b) hazard function

Bob Rigby, Mikis Stasinopoulos
Flexible Regression and Smoothing
2016 26 / 58
Continuous distributions on $\mathbb{R}^+$: three parameters

**Box-Cox Cole and Green** $\text{BCCG}(\mu, \sigma, \nu)$ known also as the LMS method in centile estimation.

**Generalised gamma** $\text{GG}(\mu, \sigma, \nu)$

**Generalised inverse Gaussian** $\text{GIG}(\mu, \sigma, \nu)$

**Log-normal family** $\text{LNO}(\mu, \sigma, \nu)$ based on the standard Box-Cox transformation

\[
Z = \begin{cases} 
\frac{(Y^\nu - 1)}{\nu} & \text{if } \nu \neq 0 \\
\log(Y) & \text{if } \nu = 0 
\end{cases} \quad (1)
\]
Continuous distributions on $\mathbb{R}^+$

Continuous distribution on $\mathbb{R}^+$: four parameters

**Box-Cox power exponential** $\text{BCPE}(\mu, \sigma, \nu, \tau)$ and $\text{BCPEo}(\mu, \sigma, \nu, \tau)$ skewness and platy-lepto;

**Box-Cox t** $\text{BCT}(\mu, \sigma, \nu, \tau)$ and $\text{BCTo}(\mu, \sigma, \nu, \tau)$ skewness and leptokurtosis;

**generalised beta type 2** $\text{GB2}(\mu, \sigma, \nu, \tau)$ skewness and platy-lepto.
Continuous distribution on \( \mathbb{R}^+ \): four parameters, BCT
Continuous GAMLSS family distributions defined on $(0, +\infty)$

<table>
<thead>
<tr>
<th>Distributions</th>
<th>family</th>
<th>no par.</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCCG</td>
<td>BCCG</td>
<td>3</td>
<td>both</td>
<td></td>
</tr>
<tr>
<td>BCPE</td>
<td>BCPE</td>
<td>4</td>
<td>both</td>
<td>both</td>
</tr>
<tr>
<td>BCT</td>
<td>BCT</td>
<td>4</td>
<td>both</td>
<td>lepto</td>
</tr>
<tr>
<td>Exponential</td>
<td>EXP</td>
<td>1</td>
<td>(positive)</td>
<td>-</td>
</tr>
<tr>
<td>Gamma</td>
<td>GA</td>
<td>2</td>
<td>(positive)</td>
<td>-</td>
</tr>
<tr>
<td>Gen. Beta type 2</td>
<td>GB2</td>
<td>4</td>
<td>both</td>
<td>both</td>
</tr>
<tr>
<td>Gen. Gamma</td>
<td>GG-GG2</td>
<td>3</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>Gen. Inv. Gaussian</td>
<td>GIG</td>
<td>3</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>Inv. Gaussian</td>
<td>IG</td>
<td>2</td>
<td>(positive)</td>
<td>-</td>
</tr>
<tr>
<td>Log Normal</td>
<td>LOGNO</td>
<td>2</td>
<td>(positive)</td>
<td>-</td>
</tr>
<tr>
<td>Log Normal family</td>
<td>LNO</td>
<td>2,(1)</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>Reverse Gen. Extreme</td>
<td>RGE</td>
<td>3</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>WEI-WEI3</td>
<td>2</td>
<td>(positive)</td>
<td>-</td>
</tr>
</tbody>
</table>
Transformation from \((-\infty, \infty)\) to \((0, +\infty)\)

- Any distribution for \(Z\) on \((-\infty, \infty)\) can be transformed to a corresponding distribution for \(Y = \exp(Z)\) on \((0, +\infty)\)
- For example: from \(t\) distribution to log \(t\) distribution
- `gen.Family("TF", type="log")`
Positive response: log T distribution

(a) $d\log T(x, \mu = -5, \sigma = 1, \nu = 10)$

- $\mu = -5$
- $\mu = -1$
- $\mu = 0$
- $\mu = 1$
- $\mu = 2$

(b) $d\log T(x, \mu = 0, \sigma = 0.5, \nu = 1000)$

- $\sigma = 0.5$
- $\sigma = 1$
- $\sigma = 2$
- $\sigma = 5$

(c) $d\log T(x, \mu = 0, \sigma = 0.5, \nu = 10)$

- $\nu = 1000$
- $\nu = 10$
- $\nu = 2$
- $\nu = 1$
Continuous distributions

Distributions on $\mathbb{R}^+$

Positive response: log T distribution

Histogram of $Y$

Bob Rigby, Mikis Stasinopoulos
Flexible Regression and Smoothing

2016 33 / 58
Continuous distributions

Distributions on $\mathbb{R}^+$

Truncating from $(-\infty, \infty)$ to $(0, +\infty)$

- Any distribution for $Z$ on $(-\infty, \infty)$ can be truncated to a corresponding distribution for $Y$ on $(0, +\infty)$
- For example: from $\text{SST}(\mu, \sigma, \nu, \tau)$ distribution to $\text{flipped-SST}(\mu, \sigma, \nu, \tau)$ distribution
- `gen.trun(0,"SST",type="left")`
Continuous distributions

Distributions on $\mathbb{R}^+$

Truncating from $(-\infty, \infty)$ to $(0, +\infty)$

(a) $dSSTr(x, \mu = 0.1, \sigma = 0.5, \nu = 1, \tau = 10)$

mu = .1
mu = 1
mu = 2

(b) $dSSTr(x, \mu = 2, \sigma = 0.5, \nu = 1, \tau = 10)$

sigma = .5
sigma = 1
sigma = 2

(c) $dSSTr(x, \mu = 2, \sigma = 0.5, \nu = 0.1, \tau = 10)$

nu = 0.1
nu = 1
nu = 2

(d) $dSSTr(x, \mu = 2, \sigma = 0.5, \nu = 1, \tau = 3)$

tau = 3
tau = 5
tau = 100
Continuous GAMLSS family distributions defined on $(0, 1)$

<table>
<thead>
<tr>
<th>Distributions</th>
<th>family</th>
<th>no par.</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>BE</td>
<td>2</td>
<td>(both)</td>
<td>-</td>
</tr>
<tr>
<td>Beta original</td>
<td>BEo</td>
<td>2</td>
<td>(both)</td>
<td>-</td>
</tr>
<tr>
<td>Logit normal</td>
<td>LOGITNO</td>
<td>2</td>
<td>(both)</td>
<td>-</td>
</tr>
<tr>
<td>Generalized beta type 1</td>
<td>GB1</td>
<td>4</td>
<td>(both)</td>
<td>(both)</td>
</tr>
</tbody>
</table>
Transformation from $(-\infty, \infty)$ to $(0, 1)$

- Any distribution for $Z$ on $(-\infty, \infty)$ can be transformed to a corresponding distribution for $Y = \frac{1}{1+e^{-Z}}$ on $(0, 1)$
- For example: from $t$ distribution to logit $t$ distribution
- `gen.Family("TF", "logit")`
Continuous distributions

Distributions on $\mathbb{R}_{(0,1)}$

logit T distribution

![Graphs showing different logit T distributions with varying parameters: (a) different values of $\mu$, (b) different values of $\sigma$, and (c) different values of $\nu$.](image-url)
Continuous distributions

Distributions on \( \mathbb{R}_{(0,1)} \)

**logit T distribution**

![Graphs of \( \logit T(x, \mu = 0) \), \( \logit T(x, \mu = 0) \), and Histogram of Y](image)
Truncating from \((-\infty, \infty)\) to \((0, 1)\)

- Any distribution for \(Z\) on \((-\infty, \infty)\) can be truncated to a corresponding distribution for \(Y\) on \((0, 1)\).
- For example: from \(SST(\mu, \sigma, \nu, \tau)\) distribution to \(\text{trun-SST}(\mu, \sigma, \nu, \tau)\) distribution.
- \text{gen.trun(c(0,1),"SST",type="both")}
Truncated SST on $(0, 1)$
Summary of methods of generating distributions

1. univariate transformation from a single random variable (LOGNO, BCCG, BCPE, BCT, EGB2, SHASHo, inverse log and logic transformations . . .)
2. transformation from two or more random variables (TF, ST2, exGAUS)
3. truncation distributions
4. a (continuous or finite) mixture of distributions (TF, GT, GB2, EGB2)
5. Azzalini type methods (SN1, SEP1, SEP2, ST1, ST2)
6. splicing distributions (SN2, SEP3, SEP4, ST3, ST4, NET)
7. stopped sums
8. systems of distributions
Continuous distributions

Comparison of properties of distributions

1. Explicit pdf cdf and inverse cdf
2. Explicit centiles and centile based measures (e.g. median)
3. Explicit moment based measures (e.g. mean)
4. Explicit mode(s)
5. Continuity of the pdf and its derivatives with respect to $y$
6. Continuity of the pdf with respect to $\mu$, $\sigma$, $\nu$ and $\tau$
7. Flexibility in modelling skewness and kurtosis
8. Range of $Y$
Theoretical comparison of distributions

- OK we have a lot of distributions in GAMLSS. Do we need any more?
- Do we need all of them or some of them are redundant?

Is there any way to compare them theoretically?

- We can compare the tail behaviour in of the distributions
- Since most of them are location-scale we can compare their flexibility in terms of skewness and kurtosis
Comparing tails: real line
Comparing tails: positive real line
Types of tails

There are three main forms for $\log f_Y(y)$ for a tail of $Y$

- **Type I:** $-k_2 \left( \log |y| \right)^{k_1}$,
- **Type II:** $-k_4 |y|^{k_3}$,
- **Type III:** $-k_6 e^{k_5 |y|}$,

decreasing $k$’s results in a heavier tail,
### Types of tails: classification (real line)

<table>
<thead>
<tr>
<th>Value of $k_1-k_6$</th>
<th>Distribution name</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1 = 1$</td>
<td>Cauchy</td>
<td>$CA(\mu, \sigma)$</td>
</tr>
<tr>
<td></td>
<td>Generalized $t$</td>
<td>$GT(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Skew $t$ type 3</td>
<td>$ST3(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Skew $t$ type 4</td>
<td>$ST4(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Stable</td>
<td>$SB(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>$TF(\mu, \sigma, \nu)$</td>
</tr>
<tr>
<td>$k_1 = 2$</td>
<td>Johnson’s SU</td>
<td>$JSU(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Johnson’s SU original</td>
<td>$JSUo(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td>$0 &lt; k_3 &lt; \infty$</td>
<td>Power exponential</td>
<td>$PE(\mu, \sigma, \nu)$</td>
</tr>
<tr>
<td></td>
<td>Power exponential type 2</td>
<td>$PE2(\mu, \sigma, \nu)$</td>
</tr>
<tr>
<td></td>
<td>Sinh-arcsinh original</td>
<td>$SHASHo(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Sinh-arcsinh</td>
<td>$SHASH(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Skew exponential power type 3</td>
<td>$SEP3(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Skew exponential power type 4</td>
<td>$SEP4(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td>$k_3 = 1$</td>
<td>Exponential generalized beta type 2</td>
<td>$EGB2(\mu, \sigma, \nu, \tau)$</td>
</tr>
<tr>
<td></td>
<td>Gumbel</td>
<td>$GU(\mu, \sigma)$</td>
</tr>
<tr>
<td></td>
<td>Laplace</td>
<td>$LA(\mu, \sigma)$</td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>$LG(\mu, \sigma)$</td>
</tr>
<tr>
<td></td>
<td>Reverse Gumbel</td>
<td>$RG(\mu, \sigma)$</td>
</tr>
<tr>
<td>$k_3 = 2$</td>
<td>Normal</td>
<td>$NO(\mu, \sigma)$</td>
</tr>
<tr>
<td>$0 &lt; k_5 &lt; \infty$</td>
<td>Gumbel</td>
<td>$GU(\mu, \sigma)$</td>
</tr>
<tr>
<td></td>
<td>Reverse Gumbel</td>
<td>$RG(\mu, \sigma)$</td>
</tr>
</tbody>
</table>
# Types of tails: classification (positive real line)

<table>
<thead>
<tr>
<th>Value of ( k_1-k_6 )</th>
<th>Distribution name</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_1 = 1 )</td>
<td>Box-Cox Cole-Green</td>
<td>( \text{BCCG}(\mu, \sigma, \nu) )</td>
</tr>
<tr>
<td></td>
<td>Box-Cox power exponential</td>
<td>( \text{BCPE}(\mu, \sigma, \nu, \tau) )</td>
</tr>
<tr>
<td></td>
<td>Box-Cox t</td>
<td>( \text{BCT}(\mu, \sigma, \nu, \tau) )</td>
</tr>
<tr>
<td></td>
<td>Generalized beta type 2</td>
<td>( \text{GB2}(\mu, \sigma, \nu, \tau) )</td>
</tr>
<tr>
<td></td>
<td>Generalized gamma</td>
<td>( \text{GG}(\mu, \sigma, \nu) )</td>
</tr>
<tr>
<td></td>
<td>Inverse gamma</td>
<td>( \text{IGA}(\mu, \sigma) )</td>
</tr>
<tr>
<td></td>
<td>( \log t )</td>
<td>( \text{LOGT}(\mu, \sigma, \nu) )</td>
</tr>
<tr>
<td></td>
<td>Pareto Type 2</td>
<td>( \text{PA2O}(\mu, \sigma) )</td>
</tr>
<tr>
<td>( k_1 = 2 )</td>
<td>Box-Cox Cole-Green</td>
<td>( \text{BCCG}(\mu, \sigma, \nu) )</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>( \text{LOGNO}(\mu, \sigma) )</td>
</tr>
<tr>
<td></td>
<td>Log Weibull</td>
<td>( \text{LOGWEI}(\mu, \sigma) )</td>
</tr>
</tbody>
</table>

| \( 1 \leq k_1 < \infty \) | Box-Cox power exponential | \( \text{BCPE}(\mu, \sigma, \nu, \tau) \) |

| \( 0 < k_3 < \infty \) | Box-Cox Cole-Green | \( \text{BCCG}(\mu, \sigma, \nu) \) |
|                        | Box-Cox power exponential | \( \text{BCPE}(\mu, \sigma, \nu, \tau) \) |
|                        | Generalized gamma     | \( \text{GG}(\mu, \sigma, \nu) \) |
|                        | Weibull               | \( \text{WEI}(\mu, \sigma) \) |

| \( k_3 = 1 \)         | Exponential           | \( \text{EX}(\mu) \) |
|                        | Gamma                | \( \text{GA}(\mu, \sigma) \) |
|                        | Generalized inverse Gaussian | \( \text{GIG}(\mu, \sigma, \nu) \) |
|                        | Inverse Gaussian     | \( \text{IG}(\mu, \sigma) \) |
How to fit distributions in R

- `optim()` or `mle()` requires initial parameter values
- `gamlssML()` performs mle using a variation of the `mle()` function
- `gamlss()` performs mle using RS or CG or mixed algorithms,
- `histDist()` performs mle using RS or CG or mixed algorithms, for a univariate sample only, and plots a histogram of the data with the fitted distribution.
- `fitDist()` fits a set of distributions and chooses the one with the smallest GAIC.
Strength of glass fibres data

Data summary:

- **R data file**: glass in package `gamlss.data` of dimensions $63 \times 1$
- **source**: Smith and Naylor (1987)
- **strength**: the strength of glass fibres (the unit of measurement are not given).
- **purpose**: to demonstrate the fitting of a parametric distribution to the data.
- **conclusion**: a SEP4 distribution fits adequately
Strength of glass fibres data: SBC
Strength of glass fibres data: creating truncated distribution

data(glass)
library(gamlss.tr)
gen.trun(par = 0, family = TF)
A truncated family of distributions from TF has been generated and saved under the names: dTFtr pTFtr qTFtr rTFtr TFtr
The type of truncation is left and the truncation parameter is 0
> m1<-fitDist(strength, data=glass, k=2, extra="TFtr")
   # AIC
> m2<-fitDist(strength, data=glass, k=log(length(strength)),
               extra="TFtr") # SBC
> m1$fit[1:8]
       SEP4   SEP3  SHASHo   EGB2    JSU  BCPEo   ST2  BCTo 
> m2$fit[1:8]
       SEP4   SEP3  SHASHo   EGB2    GU    WEI3   JSU    PE 
    36.26044 36.55444 36.59054 36.95945 38.19839 38.69995 38.98634 39.00792
Strength of glass fibres data: Results

```r
histDist(glass$strength, SEP4, nbins = 13,
        main = "SEP4 distribution",
        method = mixed(20, 50))
```
Strength of glass fibres data: SBC

SEP4 distribution

Bob Rigby, Mikis Stasinopoulos
Flexible Regression and Smoothing
2016 56 / 58
Strength of glass fibres data: summary

Call: gamlss(formula = y~1, family = FA)
Mu Coefficients:
(Intercept)
  1.581
Sigma Coefficients:
(Intercept)
  -1.437
Nu Coefficients:
(Intercept)
  -0.1280
Tau Coefficients:
(Intercept)
  0.3183

Degrees of Freedom for the fit: 4 Residual Deg. of Freedom  59
Global Deviance:  19.8111
  AIC:  27.8111
  SBC:  36.3836
END

for more information see

www.gamlss.org